CEO Behavior and Firm Performance*

Oriana Bandiera

London School of Economics

Stephen Hansen

Universitat Pompeu Fabra

Andrea Prat

Columbia University

Raffaella Sadun

Harvard University

May 5, 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) a survey that records each activity the CEOs undertake in a random work-week and (ii) a machine learning algorithm that projects these high dimensional data onto one CEO behavior index. A simple firm-CEO matching model yields the null hypothesis that, in absence of matching frictions, CEO behavior is uncorrelated with firm performance. Combining the CEO behavior index with firm level accounting data we reject this null. We find a large and significant correlation between CEO behavior and firm performance, which appears gradually over time after the CEO is appointed and is stronger in poorer regions. Structural estimates of the share of mismatched firm-CEO pairs reveal that eliminating matching frictions would have a large effect on productivity.

1

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? Do different CEOs behave differently? And do differences in behavior matter for firm performance?

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 simple firm-CEO matching model.

Our survey methodology is inspired by the classic study of CEO behavior by Mintzberg [1973], who shadowed five CEOs over the course of one week. We scale-up this methodology by recording 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.

The survey is implemented by a team of forty enumerators who phoned the CEOs or their PAs every day. For each activity, the survey collected information on five features: its nature (e.g. meeting, site visits, public event, etc.), planning horizon, number of participants involved, number of different functions and the type of participants (i.e firm employees vs external people) and their function (e.g. finance, marketing, clients, suppliers, etc.). Overall, we collected data on 42,233 events of different lengths, covering an average of 50 working hours per CEO.[1] Each of these events is characterized by one of the 4,253 combinations of the five features described above, or activities.

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 CEO behaviors as probability vectors over the activity feature set described above. We then

---

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

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 heterogeneous 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 feature combinations 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, less planning and one-on-one meetings with outsiders alone. In contrast, high values are associated with CEOs participating in larger meetings that involve functions at high managerial levels (typically C-suite functions), and that are planned in advance.

While the diary data reveal that different CEOs behave differently on all dimensions, there is no theoretical reason to expect either type of behavior to lead to heterogeneous performance across firms, 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. Indeed, in our data we find that on average CEOs with high values of the behavior index are more likely to be found in larger firms, in which the demand for structured and coordinative activities is presumably more intense.

However, performance differentials related to CEO behavior may still arise in the presence of sufficiently large matching frictions, i.e. if CEOs with different behavioral patterns are not optimally matched with the specific needs of the firms they run. To illustrate this point, we develop a model of firm-CEO matching 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. We allow for two types of frictions in the market for CEOs. First, the screening technology is imperfect and 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.

The model specifies 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 the correlation between CEO behavior and firm performance will be zero -

there would not be enough variation in CEO type within broadly similar firms identify any effect. This is the null hypothesis we test in the data. 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 would be able to observe some variation in CEO type within the same broad type of firms. In particular, the mismatched firms would be be associated with worse firm performance, and a positive correlation would emerge between the presence of a CEO of the scarce type and firm performance.

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 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. In light of the model, these results imply that matching 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 - the coordinative type - is relatively scarce in the population.

This interpretation relies on the 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. In other words, conditional on size, capital and industry, the potential 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. To allay this concern we use a within firm estimator that exploits accounting data for the period before the current CEO was appointed. If indeed the previous estimates of the behavioral index captured firm's unobservables these will be absorbed by the fixed effects 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 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 before across firms.

Next, we exploit the cross-regional variation in regional GDP to proxy for differences in the severity of matching frictions across regions and test the prediction that the correlation between behavior and performance should be lower when matching frictions are less severe.

In line with the predictions of the matching model, we find that the correlation between the CEO behavior and firm productivity to be weaker in richer regions (we cannot reject the null that the correlation equals zero in the richest regions in our sample). These findings provide further support to the hypothesis that matching frictions in the market for CEOs drive the correlation between CEO behavior and firm performance found in the data. Moreover, the findings also cast doubt on the alternative hypothesis that one type of behavior is always better for all firms. If it were so, we would find that behavior to be positively correlated with firm performance regardless of the severity of the frictions (i.e. across all countries and regions in our sample).

The final part of the analysis brings the model to the data to back out the share of mismatched firms-CEOs pairs and calibrate the parameters that cause the mismatch, and the extent to which matching frictions may be able to account for productivity differences across countries.

To the best of our knowledge, this is the only study that collects time use data to measure CEO behavior in large samples, and uses it to study its link to 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. 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. 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. 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 as they aim to explain variation in style and values rather than the link with performance. Finally, this paper is complementary to a

---

[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.

growing literature documenting the role of basic 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 activity influences firm performance.

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 matching model, which is then used to inform the empirical analysis in section 4. Section 5 investigates the extent to which matching frictions vary across regions, while section 6 calibrates the model to quantify the share of mismatches and their consequences for productivity differentials across countries. Section 7 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, a data set that contains firm level accounting data for more than 30 million firms around the world. In line with other studies [Bloom et al., 2016], the sample is restricted to manufacturing to be able to more reliably compare performance across firms. 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.[4] The final number

---

[4]The reasons for non eligibility included recent bankruptcy or the company's not being in manufacturing. 310 of the 1217 could not be contacted before the project ended.

of eligible firms was thus 6,527 in 32 two-digit SIC industries. We randomly assigned these to different enumerators to call to seek the CEOs' participation, and we managed to interview the CEOs of 1,114 of them[5] - a 17% response rate. 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. 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).

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 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.[6]

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.[7] On day one of this week (typically a Monday), the enumerator called in the morning and gathered detailed information on all the events planned in the CEO diary for the day. The enumerator then called again in the evening, to gather information on the actual events occurred during the day (including those that were not originally included in the planned agenda), and the events planned for the following day. On subsequent days, the

---

[5]1,131 CEOs agreed to participate but 17 dropped out before the end of the data collection week for personal reasons.

[6]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 (e.g. The Economic Times of India, Bloomberg, etc.), information provided on the company website and supplemental phone interviews. We define a firm to be owned by an entity if this controls more than 25.01% of the shares; if no single entity owns at least 25.01% of the share the firm is labeled as "Dispersed shareholder".

[7]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 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.

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).

enumerator called in the evening, again to collect data on the actual events happened during the day, and the planned schedule for the next day.[8]

The survey collects information on all the events lasting longer than 15 minutes in the order they occurred during the day. Figure A.1 shows a screen-shot of the survey tool.[9] For each event we collect information on the following features: (1) type (e.g. meeting, public event, etc.); (2) duration (15m, 30m, etc.); (3) planning (planned or unplanned); (4) number of participants (one, more than one); (5) functions of participants, divided between employees of the firms or "insiders" (finance, marketing, etc.) and "outsiders" (clients, banks, etc.).

Overall we collect data on 42,233 events of different duration, equivalent to 225,721 15-minute blocks. The average CEO thus has 202 15-minute events, adding up to 50 hours per week.

2.2.2 Feature description and combinations

In 57,216 times blocks (25.3% of total time), CEOs are either working alone or sending emails; in 21,895 (9.7%) they are engaged in personal or family time; and in 18,950 (8.4%) they are traveling. In the remaining 127,660 time blocks (56.6% of total time), CEOs spend time with at least one other person. In the baseline analysis we only consider these latter interactive events because they are the ones for which we can measure with precision the vector of specific attributes (e.g. planning, number of participants) which, as described below, we use to derive classifications of CEO behavior. Since this approach may potentially eliminate useful information, we also include robustness checks showing that the main results are not sensitive to this choice.

Table 2, panel A shows summary statistics describing different aspects of these interactive events. Thus, within the "type" feature, the most frequent entry is "meeting", which accounts for 74.1% of total interactive time. Table 2, panel B shows the time the average CEO spends with different functions. Furthermore, 64% of time is spent in events lasting more than 1 hour, and 75% of time in activities that are planned in advance, while 62% of events include more than one other participant beyond the CEO. 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

---

[8]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.

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

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 events, the 25th and 75th percentiles are 64% and 91%, respectively. The corresponding percentiles for time spent with production is 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 we define as an *activity*. More specifically, we define each block of time 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.[10] Examples of such combinations are:

1. Meeting; Duration of 1 hour or more; Planned; Two or more participants; With production

2. Meeting; Duration of 30 minutes max; Unplanned; One participant; With marketing

3. Meeting; Duration of 1 hour or more; Unplanned; Two or more participants; With marketing and production

4. Public Event; Duration of 1 hour or more; Planned; Two or more participants; With clients, suppliers and competitors

The most frequent, associated with 3,620 15-minute time blocks, is example (1) above.

2.3   Projecting high dimensional data onto a CEO behavior index

The data on CEO behavior is by its own nature unstructured and has a large number of dimensions. Each event is described by an array of attributes and, as mentioned above, we observe 4,253 distinct activities, or combinations of attributes.

In the absence of theories of CEO behavior that can suggest ways to aggregate the data, we provide structure in a non-arbitrary way through a dimensionality reduction approach, the Latent Dirichlet Allocation or LDA [Blei et al., 2003], which is part of the broader family

---

[10]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.

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.

of unsupervised learning techniques. LDA is one of the most widely used topic modeling techniques in natural language processing. Broadly speaking, it is a "generative" statistical model that allows sets of observations to be explained by "hidden" topics, which in turn explain why some parts of the data are similar. In our 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 (akin to managerial styles), and that the creation of each activity is attributable to one of these pure behaviors.

To be more concrete, suppose all CEOs have F possible ways of organizing each unit of their time, which we define for short activities, and let x_f be a particular activity. Let $X \equiv \{x_1, \ldots, x_F\}$ be the set of activities. A management *pure behavior* k is a probability distribution $\boldsymbol{\beta}_k$ over X that is common to all CEOs. That is, every CEO who adopts management behavior k draws elements from the activity feature set according to the same distribution $\boldsymbol{\beta}_k$. The fth element of $\boldsymbol{\beta}_k$ $\left(\beta_k^f\right)$ gives the probability of generating x_f when adopting behavior k. All behaviors are potentially associated with all elements of X ($\boldsymbol{\beta}_k \gg 0$ is compatible with the definition of behavior), but some can be associated with some behaviors more than others ($\beta_k^f \neq \beta_{k'}^f$ in general when $k \neq k'$).

In our baseline specification, we focus on the simplest possible case in which there exist only two possible pure behaviors: β^0 and β^1. This choice - while most likely an underestimation 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. Furthermore, using such coarse characterization of behavior effectively stacks the deck against finding a correlation between CEO behavior and firm performance, thus providing a conservative estimate of the actual relationship existing in the data.[11] The *behavior* of CEO i is given by a combination of the two pure behaviors according to weight $\theta_i \in [0,1]$. So, the probability that CEO i assigns to activity x_f is:

$$\beta_i(x_f) = (1 - \theta_i)\beta^0(x_f) + \theta_i\beta^1(x_f).$$

---

[11]One of the robustness checks we perform is to extend it to three pure behaviors, which delivers results that are broadly comparable to the two pure behaviors case (see Appendix). An alternative approach would be to apply statistical criteria to choose K such as cross-validation or marginal likelihood methods (see Taddy, 2012 for further discussion). Preliminary analysis indicates that, for our data, the resulting K is larger than 50. While such a model might predict feature combinations better than our baseline, interpreting its results would be very challenging. In a natural language context, Chang et al. [2009] also show a tension between predictability and interpretability in the choice of K, with larger values favoring the former and smaller values the latter.

Figure 1: Two Pure Behaviors
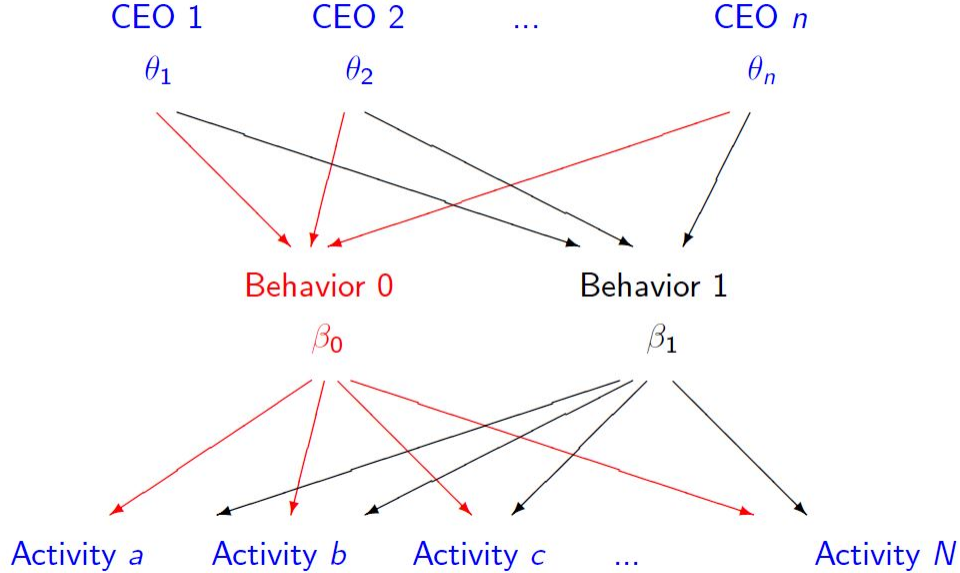
We refer to the weight θ_i as the *behavioral index* of CEO i.[12] The model corresponds to assuming that, every time t CEO i chooses an activity, he first draws one of the two pure behaviors according to θ_i, and he then draws an activity according to the pure behavior (β^0 or β^1) that he has drawn, as illustrated in Figure 1.[13]

Importantly, this 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.

For each of the $N = 1,114$ CEOs, we observe T_i distinct units of managerial time, each with an associated $y_{i,t} \in X$. Denoting with z_{it} the random variable encoding which behavior is chosen by each CEO i at each time, the probability of observing $y_{i,t}$ given the parameters $\boldsymbol{\beta} \equiv (\beta^0, \beta^1)$ and $\boldsymbol{\theta} \equiv (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_N)$ is:

$$\Pr[y_{i,t} \mid \boldsymbol{\beta}, \boldsymbol{\theta}] = \sum_{z_{i,t}} \Pr[y_{i,t}, z_{i,t} \mid \boldsymbol{\beta}, \boldsymbol{\theta}] = \sum_{z_{i,t}} \Pr\left[y_{i,t} \mid \boldsymbol{\beta}_{z_{i,t}}\right] \Pr[z_{i,t} \mid \boldsymbol{\theta}_i]. \tag{1}$$

By independence, the probability of all the observed data is $\prod_i \prod_t \Pr[y_{i,t} \mid \boldsymbol{\beta}, \boldsymbol{\theta}]$.

---

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

[13]Note that LDA allows the same activity to have strictly positive probability in both pure behaviors. A clustering algorithm would instead force it to belong to only one of the behaviors.

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 the economics of management literature is to understand overall patterns of time use for each CEO 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.[14]

2.3.1 Inference algorithm

While in principle one can attempt to estimate $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ via maximum likelihood, in practice this problem is intractable due to the fact that the number of parameters to be estimated grows linearly with the number of CEOs included in the sample. LDA overcomes these challenges by using Bayesian inference, and placing Dirichlet priors on each of the β's and the θ's.

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 hyperparameters, 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 the activity set. 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.

Exact posterior inference for LDA is intractable due to the high dimensionality of the model. In a model with K behaviors there are K^T possible realizations of the latent variables where T is the total number of time units in the data: in other words, each block of time in

---

[14]The independence assumption taken literally also implies that units of time within the same activity are independent. We explain in section 2.3.1 why we treat the unit of analysis as a time block rather than an activity.

the data can take any of K values. Enumerating all these events to compute the posterior distribution is computationally infeasible. 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.[15] We provide the details of the MCMC estimation in Appendix.

2.3.2 LDA vs. other dimensionality reduction techniques

One key advantage of LDA over simpler dimensionality-reduction techniques like principal components analysis (PCA) or k-means clustering is that - being a generative model - it provides a complete probabilistic description of time use patterns linked to statistical parameters. In this sense, LDA is akin to structural estimation in econometrics, and allows for a transparent interpretation of the estimated parameters, as it will be clearer in the next section. In contrast, PCA performs an eigenvalue decomposition of the variance-covariance matrix, while k-means solves for centroids with the smallest squared distance from the observations. Neither procedure estimates the parameters of a statistical model, which can make interpreting their output difficult.[16]

## 2.4 Estimation Results

2.4.1 Behaviors

The first two objects of interest are the pure behaviors β^0 and β^1. A first question of interest is the extent to which the algorithm identifies behavioral differences in the data. To answer it, we construct Figure 2. First, we reorder the elements of the activity feature set according to their estimated probability in $\widehat{\boldsymbol{\beta}}_0$. Second, we plot the estimated probabilities of each element of X in both behaviors. There is a clear overall pattern in which the combinations most associated with behavior 0 have low probability in behavior 1 and vice versa. In other words, behaviors are indeed sharply characterized.

Since the elements of X are combinations of features, interpreting the raw estimated probabilities associated to behaviors is rather difficult. Instead we compute marginal distributions over separate, individual features. For example, from the 654 elements of $\widehat{\boldsymbol{\beta}}_0$ and

---

[15]We use a collapsed Gibbs sampling procedure that integrates out the $\boldsymbol{\beta}_k$ and $\boldsymbol{\theta}_i$ terms from the posterior distribution, and samples just the latent assignment variables $z_{i,t}$ described above. For a more technical discussion, see Heinrich [2009] or the appendix of Hansen et al. [2014].

[16]We also replicate our exercise with k-means clustering and obtain similar results. PCA is unfeasible because it requires continuous variables, while some of our dimensions are intrinsically discrete.

Figure 2: Probability of Feature Combinations
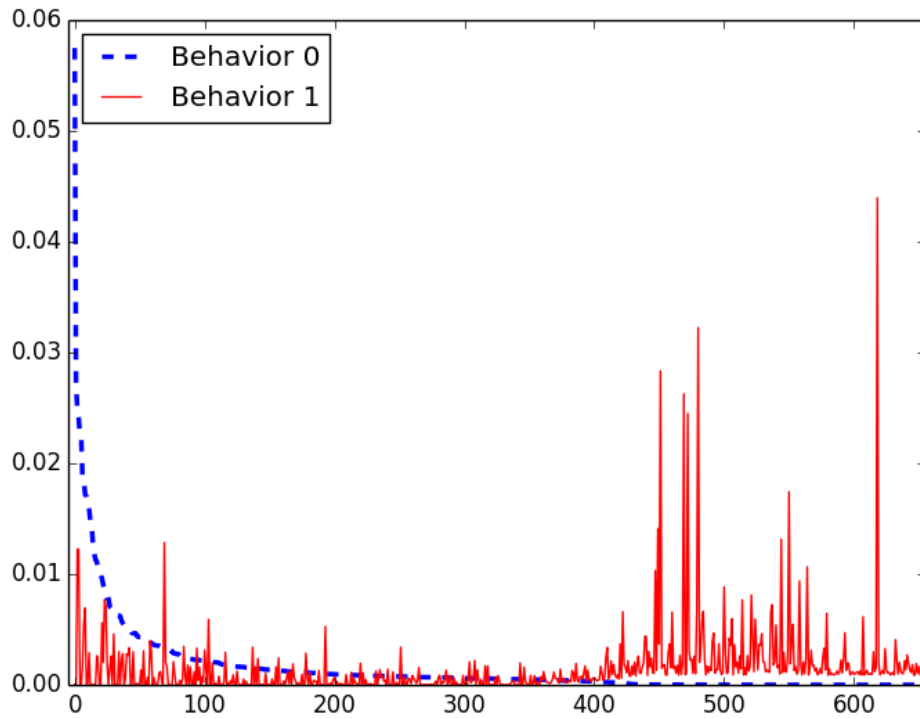
Notes: This plots the probability of different elements of the activity feature set in behaviors 0 and 1. The 654 elements of X are ordered left to right according to their probability in behavior 0.

$\widehat{\boldsymbol{\beta}}_1$ one can compute a two-element marginal distribution over the "planned" feature in each behavior. Figure 3 displays the ratios of all the marginal distributions that we compute.[17] A value of 1 for the ratio indicates that both behaviors placed 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 visiting the manufacturing shop floor - which is ten times more likely in behavior 0. Another notable difference is for business meals, which behavior 1 is over twice as likely to generate. Less prominent differences exist for phone calls, which are 34% more likely in behavior 1 and meetings, which are 7% more likely in behavior 0.

For meeting duration, behavior 0 is clearly more associated with shorter activities, with 30-minute durations 54% more likely and 1-hr durations 36% more likely. In contrast, behavior 1 is 17% more likely to generate activities that last more than one hour. 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).

The remaining differences we explore are time spent with functions. While both behaviors spend time in activities with only inside functions in equal amounts, behavior 1 is twice as likely to spend time with both inside and outside functions together, and behavior 0 is twice as likely to spend time with only outside functions. Very stark differences emerge in time spent with specific inside functions. 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, behavior 0 is over twice as likely to engage in activities with production. Smaller differences exist for finance (50% more likely in behavior 0) and marketing (10% more likely in behavior 1) functions. In terms of outside functions, behavior 0 is over three times as likely to spend time with suppliers and 25% more likely to spend time with clients, while behavior 1 is almost eight times more likely to attend trade associations.

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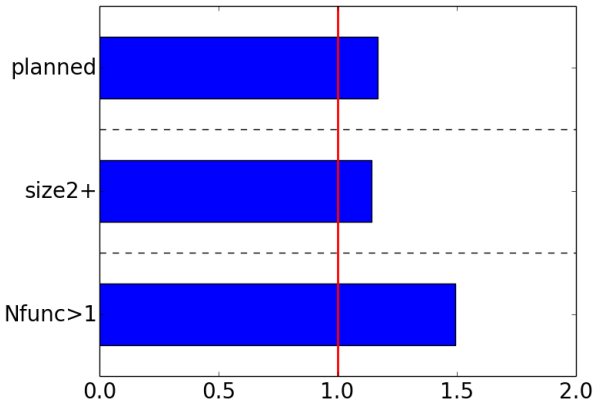
---

[17]We only report feature categories for which at least one of the two estimated behaviors has more than 0.05 probability in its marginal distribution.

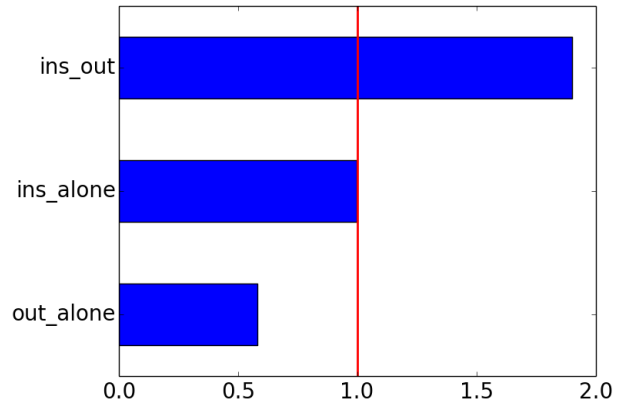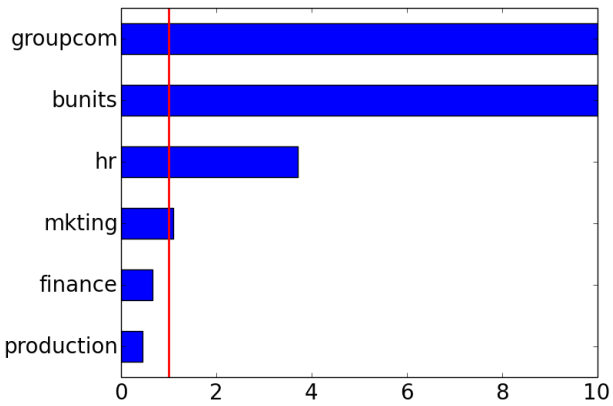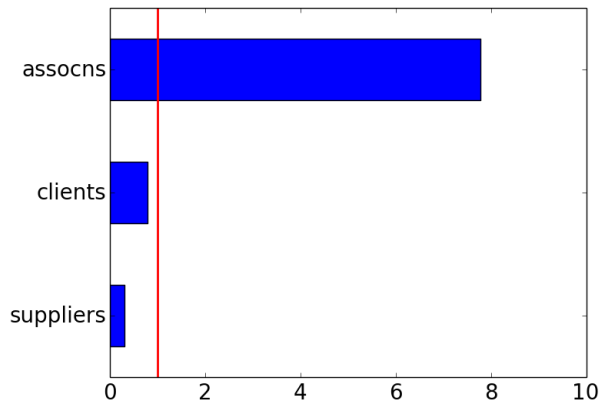Figure 3: Ratios of Marginal Distributions
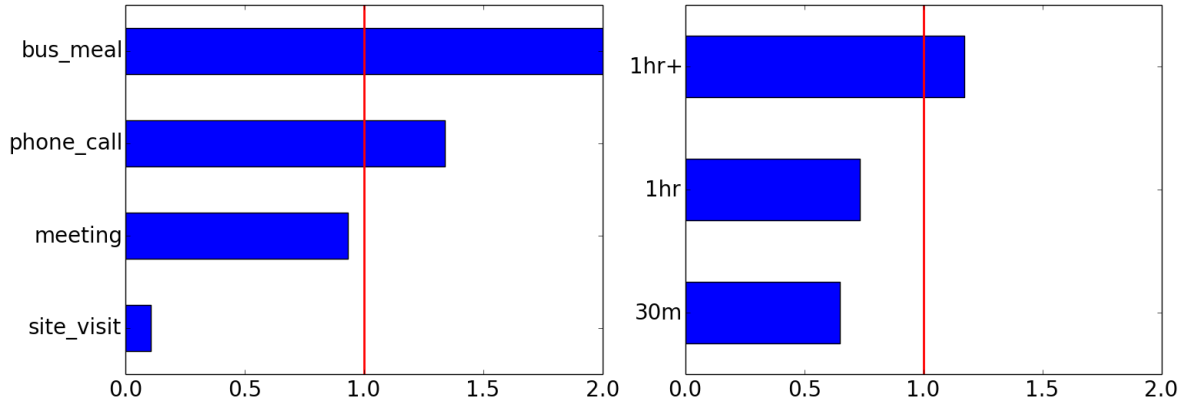
(a) Activity Type

(b) Duration

(c) Planning; Size; Number of Functions

**(d)** Inside vs Outside Time

(e) Inside Functions

(f) Outside Functions

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.

Figure 4: CEO behavior and Index Distribution

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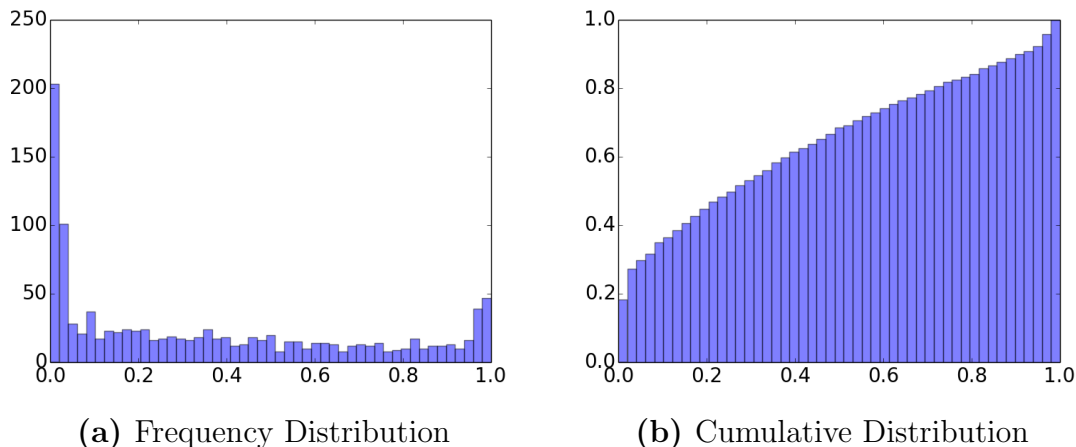
Notes: The left-hand side plot displays the number of CEOs with behavioral indices in each of 50 bins that divide the space [0,1] evenly. The right-hand side plot displays the cumulative percentage of CEOs with behavioral indices lying in these bins.

2.4.2 The CEO Behavior Index

The two 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 behavior 1. Figure 4 plots both the frequency and cumulative distributions of θ_i in our sample.

Many CEOs are estimated to be mainly associated with one 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 bulk of CEOs draw their time use from both behaviors. This again highlights the value of using a mixed-membership 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 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.[18] 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, i.e. whether there is perfect assortative matching between CEOs. The next section presents a simple matching model to guide the empirical investigation of this specific question.

3 CEO Behavior and Firm Performance: Theory

In this section we develop a model to guide the empirical analysis of the correlation between CEO behavior and firm performance. In particular, the model illustrates a simple mechanism through which mismatch between CEOs and firms may appear in equilibrium, and how the mismatch would affect the cross sectional correlation between CEO behavior and firm performance.

This minimalistic CEO-firm matching model is based on two assumptions. First, 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. Moreover, 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 makes our story different from existing theories of manager-firm matching, 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]).

The model shows the conditions under which there may be mismatch in equilibrium. In this case, some prospective CEOs who belong to the more abundant type will "pass" as CEOs of the scarce type. After they are hired, they will behave in a way that is suboptimal for their firms. The firms they run will have lower productivity. Because of this mismatch, abundant-

---

[18]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.

type CEOs will underperform on average scarce-type CEOs. Moreover, as the mismatch is based on screening errors, once one conditions on CEO type, observable attributes of CEOs have no predictive value on firm performance.

The model has a natural dynamic extension where the effect of CEO behavior on firm performance is gradual. The dynamic extension of the model builds off the assumption that it takes time for a newly hired CEO to affect the performance of the firm. This leads to predictions on the shape of the performance residual of firms as a function of CEO tenure and CEO type, which we test in our data.

3.1   Model set-up

There are two possible CEO behaviors: $x = 0$ and $x = 1$. Once a CEO is hired, he decides how he is going to manage the firm that hired him. CEOs come in two types. Type 0 prefers behavior 0 to behavior 1. Namely, he incurs a cost of 0 if he selects behavior 0 and cost of c, which we normalize to one, 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 (he finds one behavior more enjoyable) or his skill set (he finds one behavior less costly to implement).

Firms too have types. A type-0 firm is more productive if the CEO chooses $x = 0$. A type-1 firm is more productive if the CEO chooses 1. Namely, a type-j firm's output is $R = y_j$ if $x \neq j$, and $R = y_j + 1$ if $x = j$. Note that the baseline output of a firm y_j depends on the firm's type j.

All firms offer the same linear compensation scheme

$$w(R) = \bar{w} + \beta(R - y_j),$$

where $\bar{w}$ is a is a fixed part, and $\beta \geq 0$ 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).[19]

The total utility of the CEO is equal to compensation less behavior cost. After a CEO is hired, he chooses his behavior. If the CEO is hired by a firm with the same type, he will

---

[19]We assume that CEO compensation is not directly dependent on CEO behavior or CEO type. If it were, we would be in a frictionless environment.

obviously choose the behavior that is preferred by both parties. The interesting case is when the CEO type and the firm type differ. If $\beta > 1$, the CEO will adapt to the firm's desired behavior, produce an output of $y_j + 1$, and receive a total payoff of $\bar{w} + \beta - 1$. If instead $\beta < 1$, the CEO will choose his preferred behavior, produce low output $R = y_j$ and receive a payoff $\bar{w}$. We think of β as a measure of governance. A higher β makes CEO behavior more aligned with the firm's interests.

Now that we know what happens once a match is formed, let us turn our attention to the matching process. There are 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 scarce type is actually more numerous than the abundant type.

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 firm of type 0 or a firm of type 1. 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.[20] If the application is of the wrong type, deception is detected with probability $\rho \in [0, 1]$, where $\rho = 1$ denotes perfect screening and $\rho = 0$ represents no screening.[21]

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).[22]

We can show:

Proposition 1 *Assume that the screening process is sufficiently unreliable, governance is sufficiently poor, and one CEO type is sufficiently abundant.[23] Then, in equilibrium:*

---

[20]We can allow firms to mis-represent their type. In equilibrium, they will report their type truthfully.

[21]We assume that would-be-CEOs know their own type before they apply to firms. It is easy to see that our mismatch result would hold a fortiori if prospective applicants had limited or no knowledge of their own type.

[22]We assume that k is sufficiently low that a firm would not hire the first applicant independently of her type.

[23]Formally, this is given by the conditions: $\beta < 1$ and

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

- *All scarce-type CEOs are correctly matched;*

- *Some abundant-type CEOs are mismatched;*

- *The average productivity residual of firms run by abundant-type CEOs is lower than that of firms run by scarce-type CEOs.*

Proof. We verify that the situation described in the proposition corresponds to a Bayesian equilibrium. First note, that if $\beta > 1$, 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 $\beta < 1$.

In that case, when a CEO is hired, her utility is $\bar{w} + \beta$ 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, normalize $\bar{w} + \beta$. Hence the utility of a correctly matched CEO is one and the utility of a mismatched CEO is

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

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

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

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 $m((1 - \gamma) z + \gamma)$, the probability that an application is considered is

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

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

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

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:

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

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

24

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:

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

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

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 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.

3.1.1 Implications for firm performance

Proposition 1 has a key implication for firm performance. The application strategy of CEOs means that all scarce-type CEOs are matched to the right firm and will produce high performance, while some abundant-type CEOs are matched to the wrong firm and will produce low performance. Therefore, when matching frictions exist, the average performance of firms led by abundant-type CEOs - conditional on the firm being of type j - is lower.

Note that three conditions are required for the Proposition to hold: imperfect screening, imperfect governance, and a degree of unbalance between firm types and CEO types. Let us review all three possible violations. 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 ($\beta > 1$), 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 applied to firms of their type). If at least one of the three conditions fail, then in equilibrium we should observe no correlation between CEO type and firm performance because - conditional on the firm being of type j - all CEOs would behave in the same way.

It is important to note that Proposition 1 holds conditional on firm's type specific base-

line performance level y_j. In other words, it assumes that all type-j firms have identical performance potential, except for the type of CEO that leads them. The empirical challenge imposed by this assumption is that - even controlling for a host of firm observables to proxy for firm type - there might still be unobservable factors driving firm performance and, at the same time, influencing CEO behavior. In this case, the cross-sectional correlation between CEO behavior and firm performance would simply capture the importance of these firm unobservables, rather than the existence of matching frictions. We return to this identification problem in Section 4 below.

3.1.2 Additional remarks

Some remarks are in order. First, under Proposition 1, 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.[24]

Second, one can consider the extreme case where there are no 0-type firms: $\phi = 1$. This is no longer a matching problem. No firm wants the abundant-type CEO. Everybody applied to type-1 firms.

Corollary 1 *In the extreme case where there are no 0-type firms, all abundant-type CEOs apply to the wrong firm type. A share $\gamma (1 - \rho)$ of firms is run by the wrong type of CEO. All employed abundant-type CEO underperform.*

Third, 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, one can easily see 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 type-1 firms used the presence of an MBA degree to screen applicants, then only abundant-type CEOs with an MBA will apply, but 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.

---

[24]If side transfers were feasible, this would also be a Pareto-improvement as a type-1 CEO matched with a 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.

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.[25]

3.2   Dynamic Version

We now explore the dynamic implication of our CEO-firm matching 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 1 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

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

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?

---

[25]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.

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 $1 - \pi$.[26]

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

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

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 2 *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 5depicts 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.
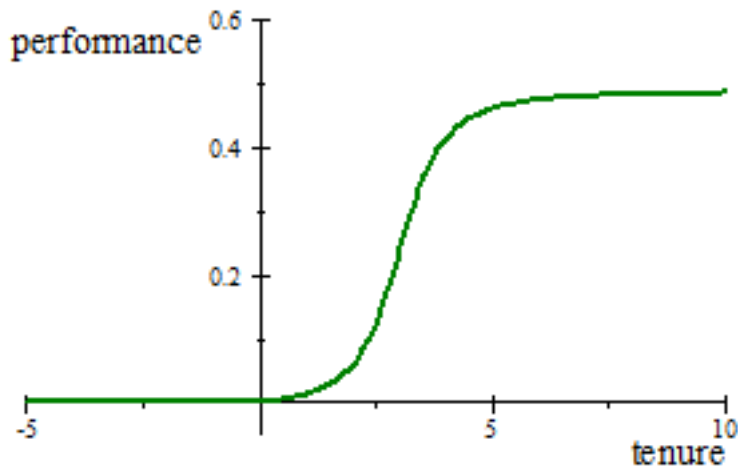
---

[26]This probability is given in equilibrium by

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

where

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

Figure 5: Average performance of firms run by type-1 CEOs

4 CEO Behavior and Firm Performance: Evidence

4.1 Cross-sectional estimates.

Guided by the model, we now test the null hypothesis of zero correlation between CEO behavior and firm performance.[27] Our main measure of performance is the value of sales (in constant 2010 dollars) controlling for the number of employees - a measure of labor productivity. This is the performance measure that is available for the largest number of firms and countries. Conditional on data availability, we also test the relationship between the CEO index and sales controlling also for capital and materials (so closer to a TFP specification), and profits per employee smaller samples.

We test the null that the correlation between the CEO behavior index and firm performance is zero. To give equal weight to each CEO regardless of his tenure, we collapse the data at the firm level using up to 5 most recent years pre-dating the survey. Our baseline specification is a production function of the form:

$$y_{ifts} = \alpha\theta_i + \beta^E e_{ft} + \beta^K k_{ft} + \beta^M m_{ft} + \mathbf{Z}_i\gamma + \zeta_t + \eta_s + \varepsilon_{ifts} \tag{2}$$

---

[27]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 outside the 5 year window pre-dating the survey. The data 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.

29

where y_{ifts} is the performance of firm f, led by CEO i, in year[28] t and sector s, θ_i is the 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. $\mathbf{Z}_c$ is a vector of CEO characteristics (MBA dummy and log(1+years as CEO)), ζ_t and η_s are year and SIC2 sector fixed effects, respectively. We include country by year dummies throughout, as well as a set of noise controls.[29]

Column 1, Table 3 shows the estimates of Equation (2) 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 at the 1% level (coefficient 0.374, standard error 0.088). Column (2) adds the log of capital, which is positive and statistically significant. While including the capital variable restricts the sample to 618 firms, the coefficient on the CEO behavior index is unaffected (coefficient 0.373, standard error 0.113). In column (3) we estimate the correlation in the even smaller sample (448 firms) for which we have at least one year of capital and materials to look at a specification closer to TFP. We find that inputs are statistically significant and of expected magnitudes and that their inclusion does not change substantially the magnitude and the significance of the CEO behavior index (coefficient 0.286, standard error 0.130). Column (4) tackles the concern that data for private firms is noisier by restricting the sample to firms that, in addition to having data on capital and materials, are listed on stock market. 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 whether the correlation captures other observable firm and CEO characteristics, rather than behavior per se. To do so, we add as controls a set of firm and CEO variables (a dummy to denote firms with a formal COO position, an MBA dummy and log of CEO tenure) which we collected as part of the Executive Time Use Survey. Including these variables hardly changes the magnitude of the CEO behavior index (coefficient 0.271,

---

[28]Since the data is aggregated into a single average, year dummies are set as the 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.

[29]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.

Table 3: CEO Behavior and Firm Performance

Dependent Variable	(1)	(2)	(3)	(4)	(5)	(6)
			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.

standard error 0.135), and the variables themselves are not significant at standard levels. This last finding is consistent with Corollary 2 of our model. If firms are using an observable CEO trait to select among candidates, then in equilibrium that trait cannot predict the probability that the CEO they hire is mismatched (if it did, it would mean the firm has not used that information optimally).

To benchmark the magnitude of the coefficient of the CEO behavior index, consider the results shown in Column (2), where we control for capital and employment. The coefficient 0.373 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, 2015) and about 15% of the effect of a one standard deviation increase in capital (taking the coefficient of 0.398 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.[30]

In light of the model, the results in Table 3 are consistent with the idea that frictions are sufficiently large to create some mismatches between firms and CEOs, and that the (unobserved) CEO type associated with higher values of the CEO behavior index is relatively scarce in the population. This interpretation relies on the identifying assumption that, conditional on observable firm characteristics such as factor inputs and sector of activity, 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

---

[30]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. 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.

just reflect unobservable firm traits that lead to low productivity. We study the empirical validity of this assumption in the next section.

4.2   Fixed effects estimates

Proposition 1 makes clear that the mismatch between CEO behavior and firm type affects firm performance conditional the firm-type specific baseline productivity level. However, firm type is not directly observable. This leaves open the possibility that firm type affects both firm performance and CEO behavior, thus creating a correlation between the two even in the absence of matching frictions.

In this section we use data that predates the appointment of the current CEO and estimate the correlation between firm performance and CEO behavior with firm fixed effects to absorb all time invariant firm characteristics. To be precise we estimate how the correlation between the productivity of a firm and the behavior index of the current CEO changes before and after he was appointed. If indeed our earlier estimates were driven by unobservable time invariant firm traits, these will be absorbed by the fixed effects and the correlation between CEO behavior before the CEO was appointed will be the same as the correlation after appointment. In contrast, if our earlier estimates were driven by the effect of CEO behavior on performance we expect this to materialize only after his appointment.

While this strategy accounts for all unobservable firm traits that are time invariant, time varying factors might still confound the interpretation of the estimates. In particular, the fact that the correlation between CEO behavior and performance materializes after the appointment might be due to the fact that firms that eventually appoint a high index CEO and those that eventually appoint a low index CEO were on different trends, so that a high value of the index captures the continuation of this trend rather than the effect of CEO behavior on performance.

To provide evidence on the validity 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:

$$y_{ist} = \lambda_i + \gamma_t + \eta_s + \sum_{-7 \leq \tau_{it} \leq 7, \tau_{it} \neq 0.} \kappa_{\tau_{it}} + \sum_{-7 \leq \tau_{it} \leq 7, \tau_{it} \neq 0} \kappa_{\tau_{it}} \tilde{\theta}_i + \beta^E e_{it} + \varepsilon_{is\tau_i} \qquad (3)$$

where $(\lambda_i, \gamma_t, \eta_s)$ are firm, year, and sector fixed effects, $\kappa_{\tau_{it}}$ are event time dummies, and the event time indicator τ_{it} is normalized so that for each firm $\tau_{it} = 0$ in the year the current CEO is appointed, so, for instance, $\tau_{it} = 1$ if year t is the first year after firm i appointed the

33

current CEO. The first summation term contains event dummies before the current CEO was appointed while the second contains event time dummies after the current CEO was appointed. Each event time dummy measures the difference in conditional log productivity between that year and the year of CEO appointment.

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 $\kappa_{\tau_{it}}$. To illustrate the results graphically we use a discretized version of the CEO behavior index so that $\tilde{\theta}_i = 1$ if $\theta_i \geq 0.5$, although the main results hold when we use the continuous CEO behavior index.

If our cross-sectional estimates were due to firm time invariant unobservables these will be captured by the firm fixed effects λ_i and the appointment of the CEO with index $\tilde{\theta}_i$ will not affect productivity, namely the coefficients on the interactions between year dummies and the index ($\sum_{-7 \leq \tau_i \leq 7, \tau_i \neq 0} \kappa_{\tau_i} \tilde{\theta}_i$) will be zero before and after the appointment. Alternatively, if our cross-sectional estimates were capturing differential trends between firms that eventually appoint a high index CEO and those that eventually appoint a low index CEO, the coefficients on the interactions between year dummies and the index ($\sum_{-7 \leq \tau_i \leq 7, \tau_i \neq 0} \kappa_{\tau_i} \tilde{\theta}_i$) will be different from zero before and after the appointment. In contrast, if it is the CEO who affects productivity, the coefficients on the interactions will be zero before and positive after the appointment.

The results of this analysis are shown 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.[31] This sample includes 3,433 observations, of which 762 relative to years pre-dating the CEO appointment, and the rest relative to years in which the CEO was in office.[32] Column (1) shows the estimated coefficients on the event time dummies for firms whose current CEO has $\tilde{\theta}_i = 0$. Productivity appears to be significantly higher in the years immediately preceding the CEO appointment (up until $\tau = -2$) relative to the appointment year, and significantly lower in the year following the appointment (from $\tau = 2$). Column (2) shows that before the appointment of the current CEO, the productivity pattern is very similar for firms that eventually appoint $\tilde{\theta}_i = 1$. As shown in column (3) the differences

---

[31]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 3, column (1) and just using the years in which the CEO is in office is 0.427 (standard error .103).

[32]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 $\tau > 7$.

between the two types of firms in the pre-appointment period are all insignificant, suggesting that productivity trends were similar in the two groups before the current appointment.

However, for firms with $\tilde{\theta}_i = 1$ the productivity decline stops after the CEO appointment, and it is reverted 3 years after the appointment. The difference in the point estimates of the tenure dummies after the CEO appointment between the two types of firms, that is the coefficients on the interactions between year dummies and the index ($\sum_{-7 \leq \tau_{it} \leq 7, \tau_{it} \neq 0} \kappa_{\tau_{it}} \tilde{\theta}_i$), are significant after year 3. These results are also shown visually in Figure 6. Panel A illustrates the point estimates on the event time dummies for the two types of firms, while Panel B shows the difference between the dummies within each event time, as well as their confidence interval. The shape observed in these figures is consistent with the shape predicted in Proposition 2. The effect of CEO behavior on performance is s-shaped over the years: there is no detectable effect for the first three years followed by a fast increase (and possibly plateauing later on). The gradual emergence of the productivity differentials across the two types is consistent with the idea that the effect of CEOs on firm performance takes time to materialize as illustrated in the dynamic extension of our baseline model. 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.[33]

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 CEO tenure (4 years and + after appointment). Column (3) shows that this result is robust to including additional controls for capital and material inputs.[34] Finally, column (4) shows that the mean change in productivity experienced by $\tilde{\theta}_i = 1$ firms is not accounted for by

---

[33]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.

[34]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.

Table 4: Productivity by year of CEO Tenure

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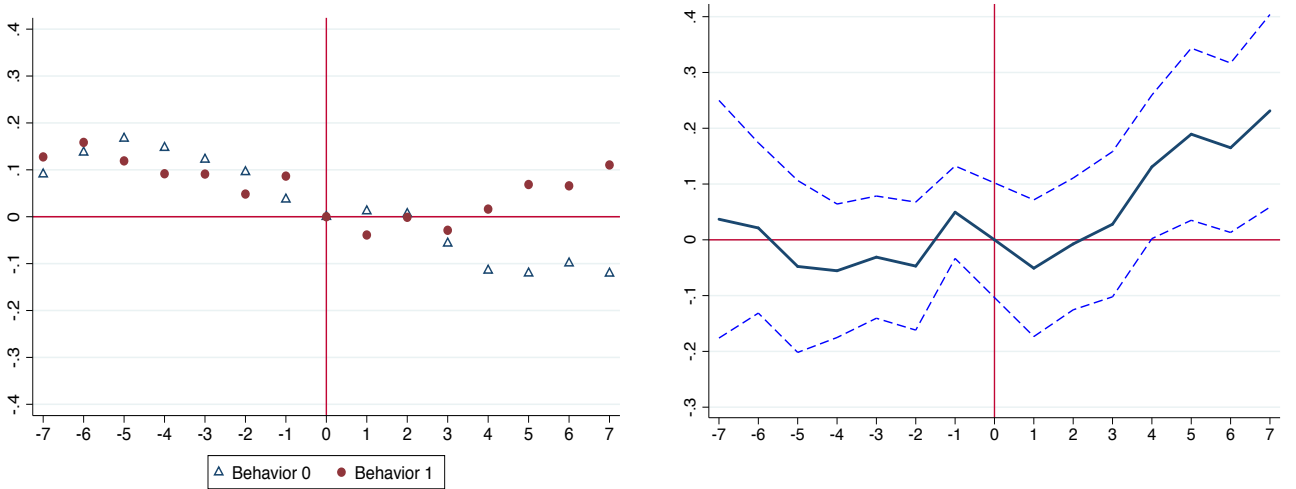
Notes: These figures accompany Table 4. Panel A plots the point estimates of the coefficients reported in Table 4, columns (1) and (2). Panel B plots the difference between the two points in each year and its confidence interval.

differential time trends between $\tilde{\theta}_i = 1$ and $\tilde{\theta}_i = 0$ firms.

Overall, the results in Tables 4 and 5 suggest that the cross-sectional estimates shows in Table 3 are not entirely driven by differences in firms' unobserved time invariant traits, or differential productivity trends before appointment. This rules out a common type of time varying firm unobservables, namely that firms appoint a $\tilde{\theta}_i = 0$ CEO in response to a productivity decline.

Therefore, the patterns we observe seem inconsistent with two classes of alternative assumptions: (i) the observed behavior-performance relationship is due to stable unobserved characteristics of the firm that affect both performance and behavior; (ii) the observed behavior-performance relationship is due to unobserved firm characteristics that change before the CEO is hired. 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 four years after the appointment date. This requires the ability to forecast that long ahead and to act on that forecast. For instance, boards may be able to predict four years ahead of time that some (firm-specific) factors will affect performance and use this information to replace (four years ahead of the performance change) the existing CEO, who has the right θ given current circumstances, with another CEO with a θ that will become optimal in four years time. While our data cannot rule out this possibility, two observations

37

make this board strategy implausible in practice. First, this strategy requires boards to 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. Second, this strategy implies that boards are willing to bear losses due to the wrong match 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.

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.[35] To our knowledge, this is the first time that data on middle level management practices and information on CEO behavior is systematically analyzed.[36]

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.[37]   Table 6 Column (1) shows that higher values of the CEO behavior index are

---

[35]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.

[36]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.

[37]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

Table 5: Tenure Regressions

Dependent Variable	(1)	(2)	(3)	(4)
		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.

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).

We then turn to analyzing both the CEO behavior index and the management variables in the context of the production function of Equation (2). 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.

4.3.2   Alternative ways of building the CEO Behavior Index

TBC

4.3.3   Alternative estimation methods of the production function

TBC

5   CEO-Firm Match along the Development Path

In this section we exploit regional variation in development across and within countries to provide further evidence on whether the correlation between CEO behavior and firm

---

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.

Table 6: Management Regressions

Dependent Variable	(1)	(2)	(3)	(4)	(5)	(6)
	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.

performance may be driven by matching frictions in the market for CEOs. The analysis relies on the assumption that frictions are more severe in poorer regions. The reasons underpinning this assumption are manyfold. At the hiring stage, screening might be worse in low income regions because the market for CEOs is less thick and professional headhunting services less common. After hiring, governance might be worse because contract enforcement is less effective in low income regions and courts are slower. To operationalize this idea, we use regional GDP to proxy for the severity of matching frictions and test whether: a) the quality of the match is higher and; b) the correlation between behavior and firm performance is lower when matching frictions are less severe.

We use within country regional variation in development, using the data on regional income per capita in current purchasing-power-parity (PPP) dollars developed by Gennaioli et al. [2013]. The sample firms are located in 121 regions within 6 countries that are at very different stages of the development path. Income per capita in the poorest region in the sample (Uttar Pradesh, India) is \$1,300; in the richest region it is \$143,000 (DC, USA). The median within country range in income per capita is \$23,942 and the median within country standard deviation is \$5,695. The median number of regions within each country is 16.

We start the analysis by showing in Figure 7 a box plot of the CEO behavior index across countries. Clearly, the median value of the CEO behavioral index is higher in richer countries, and significantly lower in Brazil and India. The graph also shows that there is ample within country variation in each country, although the distribution is more compressed in India.

Figure 8 shows the values of the CEO behavior index across different terciles of the within country distribution of regional income per capita (i.e. we first normalize regional income per capita by its country mean, and then look at relatively poorer or richer regions within country). The left bars in the graph refer to raw averages, while right bars report employment weighted averages. Even within countries, richer regions tend to have a higher value of the CEO behavior index. Furthermore, differences between poor and richer regions are larger when we consider employment weighted averages, which is consistent with the idea that, in richer regions, CEOs with high values of the behavior index are more likely to be found in large firms relative to poor regions.

Table 7 exploits within-country, cross-regional variations in income per capita to test whether the quality of the match is higher when frictions are less severe. We estimate:

$$\theta_{ifsr} = \alpha + \beta e_f + \gamma e_f * Y_r + \vartheta a_s + \mathbf{Z}_i \delta + \varepsilon_{ifsr} \tag{4}$$

Figure 7: CEO behavior index across countries
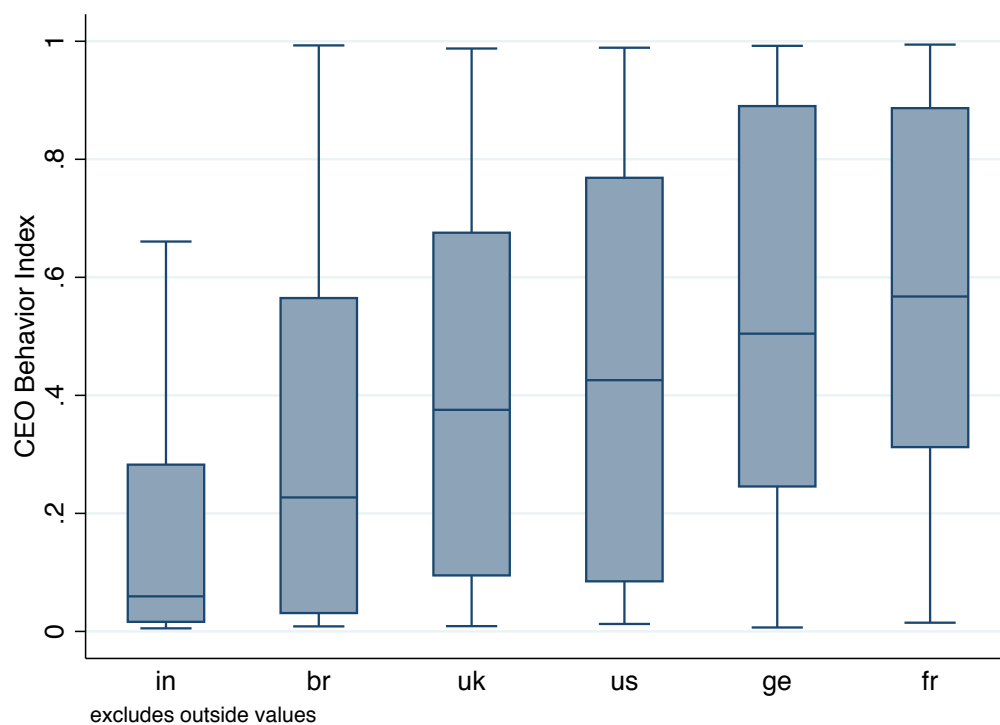
Notes: The graph shows the box plot of the CEO behavior index by country of CEO location. For each country, the upper line denotes upper adjacent value, the upper side of the box denotes the 75th percentile, the mid line the median, the lower side the 25th percentile and the lower line the lower adjacent value. Number of observations: India=358; Brazil=280; UK=87; US=149; Germany=125; France=115.

Figure 8: CEO behavior index by tercile of regional income per capita
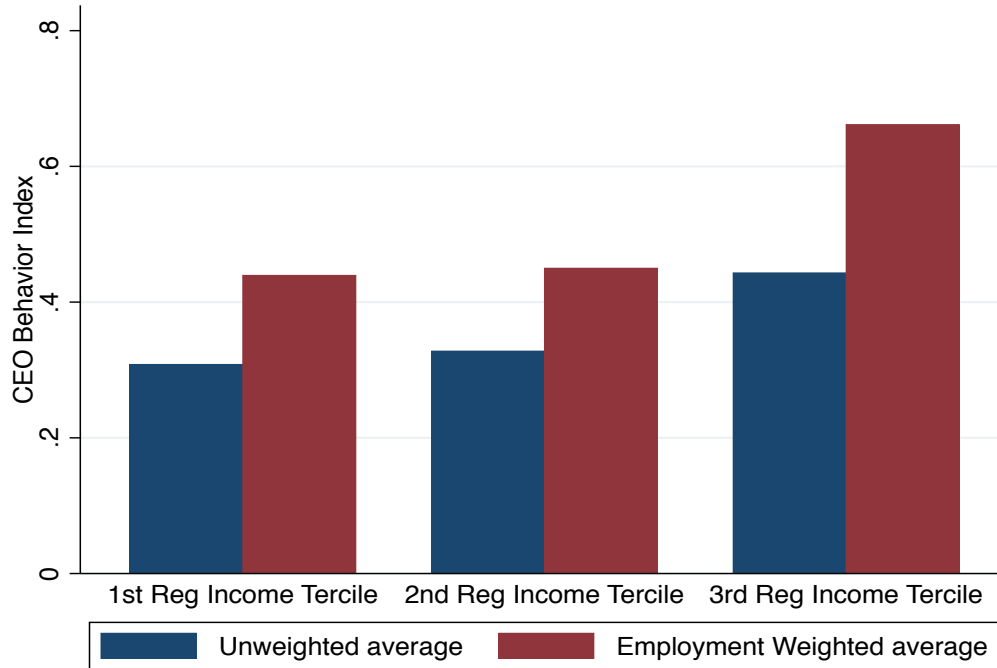
Notes: Each bar in the figure represents the average of the CEO behavior index by tercile of regional income per capita (removing country specific averages first). Within each tercile, the left bar shows raw averages, while the right bar shows employment weighted averages (employment weights computed within each country).

where θ_{ifsr} is the behavior index of CEO c in firm f, industry s and region r, e_f is log firm employment and Y_r is a measure of regional development that proxies for matching frictions. All specifications include the same set of CEO and noise controls discussed above, as well as SIC2 industry dummies. Throughout the analysis we control for country fixed effects, thereby exploiting the variation in development levels across regions within countries.

If the quality of the match improves as frictions become less severe, we expect $\gamma > 0$. Column (1) proxies for development with a dummy that equals one if the region is located in a high income country (France, Germany, UK and US). In line with the hypothesis that the quality of the match is higher when frictions are less severe, we find that the correlation between firm size and the behavioral index is significantly larger in richer countries. Column (2) uses log regional income per capita to proxy for Y_r, and finds a similar result: the strength of the correlation between the CEO behavior index and firm size increases as regional income per capita increases, indicating that in highly developed regions large firms are more likely to hire CEOs with a high behavior index. Column (3) further probes this correlation with the inclusion of a full set of regional dummies, thus exploiting *within region* variation in CEO behavior and firm size, with remarkably similar results.

Second, we test whether the correlation between CEO behavior and firm performance *decreases* with the level of regional development. Intuitively, if the match between firms and CEOs improves with development, the share of mismatched CEOs should decrease as well, and so should the difference in performance between firms led by different CEO types. To test this idea, we estimate:

$$y_{iftsr} = \alpha\theta_i + \delta\theta_i * Y_r + \beta^E e_{ft} + \mathbf{Z}_i\gamma + \zeta_t + \eta_s + \varepsilon_{iftsr} \tag{5}$$

where all variables are defined above. If the quality of the match improves as frictions become less severe, we expect $\delta < 0$. This is because as fewer CEOs are mismatched, differences in behavior are more likely to reflect optimal responses in firm needs.

Table 7, columns (4) to (6) report the estimates of Equation (5). Column (4) shows that the positive correlation between the CEO behavior index and firm performance shown in 3, column (1) is, in fact, an average of two very different magnitudes: 0.509 in low income countries and 0.068 in high income countries. The difference is precisely estimated at the 1% level. Column (5) defines Y_r as regional income per capita and again finds $\delta < 0$. Figure 9plots the estimated correlation $\alpha + \delta Y_r$ evaluated at all sample values of Y_r. The correlation between the CEO behavior index and firm performance becomes statistically indistinguishable from zero at log(regional income per capita)=10.5, which is at the 75th

Table 7: CEO-Firm match by level of development

Dependent Variable	(1)	(2)	(3)	(4)	(5)	(6)
	CEO behavior index			Log(sales)		
log(employment)	0.047***	-0.122	-0.135	0.634***	0.635***	0.630***
	(0.010)	(0.076)	(0.087)	(0.041)	(0.047)	(0.054)
log(employment) * High income country	0.034**					
	(0.016)					
Log Region income per capita		-0.043			0.090	
		(0.065)			(0.091)	
log(employment) *Log Region income per cap		0.020**	0.021**			
		(0.008)	(0.009)			
CEO behavior index				0.509***	1.707***	1.687**
				(0.121)	(0.622)	(0.725)
CEO behavior* High income country				-0.441***		
				(0.149)		
CEO behavior*Log Region income per capita					-0.147**	-0.145*
					(0.063)	(0.074)
log(capital)				0.167***	0.168***	0.175***
				(0.027)	(0.025)	(0.028)
log(materials)				0.342***	0.338***	0.342***
				(0.047)	(0.052)	(0.063)
Adjusted R-squared	0.277	0.284	0.314	0.853	0.852	0.850
N	1114	1114	1114	920	920	920
Controls:						
Region dummies	n	n	y	n	n	y
Cluster	Industry	Region	Region	Industry	Region	Region

Notes: *** (**) (*) denotes significance at the 1%, 5% and 10% level, respectively. Columns (1) to (3) include country fixed effects and noise controls. Columns (4) to (6) include country by year fixed effects 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. 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 2 digit SIC dummies. All columns weighted by the week representativeness score assigned by the CEO at the end of the interview week. "High income country" is a dummy taking value 1 for firms located in France, Germany, UK or US. "Log regional income per capita" is in current purchasing-power-parity (PPP) dollars and is drawn from Gennaioli et al (2013). Errors clustered as noted.

percentile of the regional income per capita distribution in our sample. Finally, in column (6) we include in the specification a full set of regional dummies, thus estimating $\delta < 0$ relying exclusively on within region variation. The results are robust to the inclusion of the regional dummies, although the significance level of the interaction term drops to 10%.

Taken together, the results shown in Table 7 are consistent with the idea that the correlation between CEO behavior and firm performance arises because of matching frictions. Importantly, they rule out a simpler moral hazard story where the high index behavior is always more productive (i.e. there is no demand for low index behavior), since if it were so we would find a positive correlation between firm performance and the index across all regions and countries in our sample.

Figure 9: CEO Behavior index and firm productivity by level of regional income
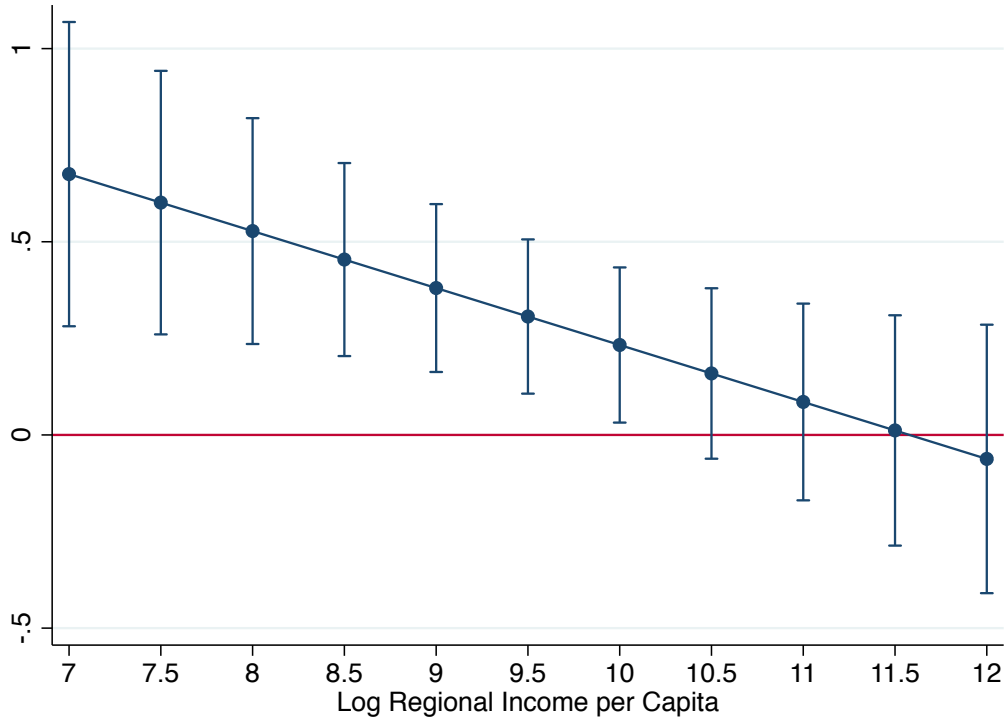
Notes: The figure plots the coefficient on the CEO behavior index for different levels of regional income per capita, as predicted by the coefficients on the CEO Behavior Index * Log regional income per capita reported on Table 7, column (5).

6 Calibration

TBC

7 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, and that these differences tend to co-vary with observable firm characteristics, such as firm size, organizational structure and industry characteristics. Guided by a simple firm-matching model, we also show evidence of significant matching frictions in the assignment of CEOs to firms, and that these frictions appear to be particularly severe in emerging economies.

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 leadership behaviors affect firm performance. The CEO behavior that according to our CEO-firm matching model and our data is scarcer in the population of potential 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 is a *coordinator*, who delegates operational tasks to high-level executives and spends more of his time making ensuring good communication in the top management team. Within the same interpretation, a CEO that displays the other CEO behavior emerging from the classification exercise is instead a *micromanager*, who tends to intervene directly in operational aspects, who prefers one-on-one meetings with a variety of internal and external constituents, and who puts less emphasis on long term planning.

To the best of our knowledge, the coordinator/micromanager 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.[38] Future work could utilize information about CEO behavior to inform

---

[38]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

alternative leadership models. At the same time, it would also be interesting to better explore the connection between our observed behavioral patterns and contributions in the management literature. For example, Kotter [1999] proposes that the key task of a CEO is to align the organization behind a common vision - the emphasis of our Behavior-1 CEOs on large, planned, multi-functional meetings is consistent with an alignment effort.

More generally, 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.

---

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, 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

Oriana Bandiera, Luigi Guiso, Andrea Prat, and Raffaella Sadun. What Do CEOs Do? CEP Discussion Papers dp1145, Centre for Economic Performance, LSE, May 2012. URL `https://ideas.repec.org/p/cep/cepdps/dp1145.html`. 7

Oriana Bandiera, Andrea Prat, and Raffaella Sadun. Managing the Family Firm: Evidence from CEOs at Work. Harvard Business School Working Papers 14-044, Harvard Business School, December 2013. URL `https://ideas.repec.org/p/hbs/wpaper/14-044.html`. 1

Oriana Bandiera, Luigi Guiso, Andrea Prat, and Raffaella Sadun. Matching Firms, Managers, and Incentives. *Journal of Labor Economics*, 33(3):623 – 681, 2015. URL `https://ideas.repec.org/a/ucp/jlabec/doi10.1086-679672.html`. 3

Morten Bennedsen, Kasper Meisner Nielsen, Francisco Perez-Gonzalez, and Daniel Wolfenzon. Inside the Family Firm: The Role of Families in Succession Decisions and Performance. *The Quarterly Journal of Economics*, 122(2):647–691, 2007. URL `https://ideas.repec.org/a/oup/qjecon/v122y2007i2p647-691..html`. 1

Marianne Bertrand and Antoinette Schoar. Managing with Style: The Effect of Managers on Firm Policies. *The Quarterly Journal of Economics*, 118(4):1169–1208, 2003. URL `https://ideas.repec.org/a/oup/qjecon/v118y2003i4p1169-1208..html`. 1

David M. Blei, Andrew Y. Ng, and Michael I. Jordan. Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3:993–1022, March 2003. ISSN 1532-4435. URL `http://dl.acm.org/citation.cfm?id=944919.944937`. 1, 2.3

Nicholas Bloom and John Van Reenen. Measuring and Explaining Management Practices Across Firms and Countries. *The Quarterly Journal of Economics*, 122(4):1351–1408, 2007. URL `https://ideas.repec.org/a/oup/qjecon/v122y2007i4p1351-1408..html`. 1

Nicholas Bloom, Raffaella Sadun, and John Van Reenen. Management as a Technology? mimeograph, 2016. 1, 2.1, 33, 4.3.1

Patrick Bolton, Markus K. Brunnermeier, and Laura Veldkamp. Leadership, Coordination, and Corporate Culture. *The Review of Economic Studies*, 80(2):512–537, 2013. doi: 10.1093/restud/rds041. URL `http://restud.oxfordjournals.org/content/80/2/512.abstract`. 38

J Chang, S Gerrish, J L Boyd-Graber, and D M Blei. Reading Tea Leaves: How Humans Interpret Topic Models. In *Advances in Neural Information Processing Systems*, 2009. 11

Wouter Dessein and Tano Santos. Managerial Style and Attention. mimeograph, 2016. 38

Xavier Gabaix and Augustin Landier. Why has CEO Pay Increased So Much? *The Quarterly Journal of Economics*, 123(1):49–100, 2008. URL `https://ideas.repec.org/a/oup/qjecon/v123y2008i1p49-100..html`. 3

Nicola Gennaioli, Rafael La Porta, Florencio Lopez de Silanes, and Andrei Shleifer. Human Capital and Regional Development. *The Quarterly Journal of Economics*, 128(1):105–164, 2013. URL `https://ideas.repec.org/a/oup/qjecon/v128y2013i1p105-164.html`. 5

T. L. Griffiths and M. Steyvers. Finding Scientific Topics. *Proceedings of the National Academy of Sciences*, 101(Suppl. 1):5228–5235, April 2004. 2.3.1

Marina Halac and Andrea Prat. Managerial Attention and Worker Engagement. CEPR Discussion Papers 10035, C.E.P.R. Discussion Papers, June 2014. URL `https://ideas.repec.org/p/cpr/ceprdp/10035.html`. 33

Donald C. Hambrick and Phyllis A. Mason. Upper Echelons: The Organization as a Reflection of Its Top Managers. *The Academy of Management Review*, 9(2):193–206, April 1984. 7

Stephen Hansen, Michael McMahon, and Andrea Prat. Transparency and Deliberation within the FOMC: a Computational Linguistics Approach. Discussion Paper 9994, CEPR, May 2014. 2, 15

Gregor Heinrich. Parameter Estimation for Text Analysis. Technical report, vsonix GmbH and University of Leipzig, 2009. 15

Benjamin E Hermalin. Toward an Economic Theory of Leadership: Leading by Example. *American Economic Review*, 88(5):1188–1206, December 1998. URL `https://ideas.repec.org/a/aea/aecrev/v88y1998i5p1188-1206.html`. 38

Benjamin E. Hermalin. Leading for the Long Term. *Journal of Economic Behavior & Organization*, 62(1):1–19, January 2007. URL `https://ideas.repec.org/a/eee/jeborg/v62y2007i1p1-19.html`. 38

Steven N. Kaplan and Morten Sorensen. Are CEOs Different? Characteristics of Top Managers. mimeograph, 2016. 1

Steven N. Kaplan, Mark M. Klebanov, and Morten Sorensen. Which CEO Characteristics and Abilities Matter? *The Journal of Finance*, 67(3):973–1007, 2012. ISSN 1540-6261. doi: 10.1111/j.1540-6261.2012.01739.x. URL `http://dx.doi.org/10.1111/j.1540-6261.2012.01739.x`. 1

John P. Kotter. *John Kotter on What Leaders Really Do*. Harvard Business School Press, Boston, 1999. 3, 7

Fred Luthans. Successful vs. Effective Real Managers. *Academy of Management Executive*, 2(2):127–132, 1988. 3

Ulrike Malmendier and Geoffrey Tate. CEO Overconfidence and Corporate Investment. *Journal of Finance*, 60(6):2661–2700, December 2005. URL `https://ideas.repec.org/a/bla/jfinan/v60y2005i6p2661-2700.html`. 1

Ulrike Malmendier and Geoffrey Tate. Superstar CEOs. *The Quarterly Journal of Economics*, 124(4):1593–1638, 2009. URL `https://ideas.repec.org/a/oup/qjecon/v124y2009i4p1593-1638..html`. 1

Henry Mintzberg. *The Nature of Managerial Work.* Harper & Row., New York, 1973. 1

William Mullins and Antoinette Schoar. 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, September 2013. URL `https://ideas.repec.org/p/nbr/nberwo/19395.html`. 1

Chad Syverson. What Determines Productivity? *Journal of Economic Literature*, 49(2):326–65, June 2011. URL `https://ideas.repec.org/a/aea/jeclit/v49y2011i2p326-65.html`. 1

Matthew A. Taddy. On estimation and selection for topic models. In *Proceedings of the 15th International Conference on Artificial Intelligence and Statistics (AISTATS)*. JMLR: W&CP 22, 2012. 11

Marko Tervio. The Difference That CEOs Make: An Assignment Model Approach. *American Economic Review*, 98(3):642–68, June 2008. URL `https://ideas.repec.org/a/aea/aecrev/v98y2008i3p642-68.html`. 3

Eric Van den Steen. On the origin of shared beliefs (and corporate culture). *RAND Journal of Economics*, 41(4):617–648, 2010. URL `https://ideas.repec.org/a/bla/randje/v41y2010i4p617-648.html`. 38

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.[39]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.

---

[39]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   MCMC Estimation

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:[40]

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

---

[40]We run five chains beginning from five different seeds, and select the one for analysis that has the best goodness-of-fit across the draws we take after burn in.

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.[41]

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 $X = \{\text{unplanned},\text{planned}\} \times \{\text{size1}, \text{size2+}\}$. 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 $x_f = \text{size2+planned}$); 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 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   Appendix Figures and Tables

---

[41]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.

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

(a) Random Seed

CEO	size1unplanned B0	B1	size1planned B0	B1	size2+unplanned B0	B1	size2+planned 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

CEO	size1unplanned B0	B1	size1planned B0	B1	size2+unplanned B0	B1	size2+planned 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

CEO	size1unplanned B0	B1	size1planned B0	B1	size2+unplanned B0	B1	size2+planned 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.

Figure A.1: Survey Instrument

Table A.4: CEO behavior index and firm characteristics

Dependent Variable	(1)	(2)	(3)	(4)	(5)	(6)
			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.