PRELIMINARY:
Managers and Productivity in Retail

Robert D. Metcalfe, Alexandre B. Sollaci, Chad Syverson*

July 27, 2022

Abstract

Research has established that management explains a notable portion of productivity differences across producers. A more open question is whether it is managers or management practices that matter. We shed light on this question by analyzing store-level data from two large retail companies in the UK and US. In our setting, managers move between stores but management practices are set by firm policy and largely fixed, allowing us to hone in on managers' personal roles in determining store performance. We find (i) managers affect and explain a large share of the variance of store-level productivity; (ii) managers who move do so on average from inherently less productive to more productive stores; (iii) manager quality is correlated with the ratio of full-time to part-time workers, but the direction of the correlation differs between firms; (iv) good managers in times of steady economic growth are also good managers during turbulent times; and (v) exogenous productivity shocks improve the quality of initially low quality managers, suggesting managers can learn.

*Metcalfe: University of Southern California & NBER; Sollaci: International Monetary Fund; Syverson: University of Chicago Booth School of Business & NBER

Acknowledgments: We thank participants at the Empirical Management Conference, Tim Armstrong, and Matt Kahn for comments. We also appreciate the willingness of the two retailers to share their data for research. We received no remuneration from them for this work. The views expressed in this paper are those of the author(s) and do not necessarily represent the views of the IMF, its Executive Board, or IMF management.

1 Introduction

A growing literature has established that management influences business performance. This work has spanned settings that vary both by industry/sector and the development level of the broader economy the business operates within. Examples of relevant studies include, among others, papers by Ichniowski et al. (1997); Ichniowski and Shaw (1999); Bertrand and Schoar (2003); Ichniowski and Shaw (2003); Bloom and Van Reenen (2007); Bloom et al. (2012, 2013, 2014); Kaplan and Sorensen (2017); Bruhn et al. (2018); Bloom et al. (2019); Giorcelli (2019); Janke et al. (2019); Bandiera et al. (2020); Gosnell et al. (2020).

Management's effects on performance can operate through a combination of not mutually exclusive channels: the manager herself, or management practices (Syverson, 2011). This is an important distinction. In principle, management practices can be part of the intangible capital of an organization and as such transferable over time and space. Personal manager performance effects, on the other hand, are inherently tied to the manager. This distinction has implications regarding both firms' optimal strategies for improving performance as well as how the rents yielded by superior management are distributed.

Much of the literature has focused on the managerial practices channel. This is especially true of the experimental literature, as practices are more amenable to manipulation than are managers' identities and innate abilities. Some work has looked at managers' personal effects (such as innate ability and personality styles), though this has focused mostly on corporate CEOs (Bertrand and Schoar, 2003; Fee et al., 2013; Bandiera et al., 2015, 2020; Baltrunaite et al., 2021; Huber et al., 2021). One potential confound in the corporate CEO effects literature is that it relies on CEOs moving across firms for identification, which opens the possibility that firm-level unobservables may shape the selection of certain managers or their influence on performance.

In this paper, we shed additional light on individual managers' influence on business performance. Unlike the CEO effects literature, our setting allows us to control for firm-level influences on performance like governance, corporate culture, branding, and firm policies regarding management practices (like operations, monitoring, information, targets, and incentives). This is because our data include information on the performance and management of thousands of stores owned by two retail firms, one in the US and one in the UK. Our data are rich enough to not just measure how much individual managers matter, but also to allow us to characterize the mechanisms through which their influence flows into measured

performance. This includes, for example, managers' effects on labor efficiency and energy efficiency. We do all of this through operations data rather than surveys.

Further still, our estimation approach allows for heterogeneity not just in managers but in the stores themselves, separate from any manager effects. This allows us to capture how managers might influence stores of different types. It also lets us see whether and to what extent the firm's allocation of managers across stores harnesses any complementarities between store and manager types. In this way, our setting and approach lets us zoom out beyond measuring manager heterogeneity and ask more broadly whether the allocation of those heterogeneous managers is in any way optimal.

Estimating individual manager and store effects on productivity poses methodological challenges. First, not all manager and store fixed effects can be separately estimated (Abowd et al., 1999). Identification of both effects relies on managers moving between stores so that their effect on productivity can be measured relative to a common baseline. Second, even when those effects can be separately identified, there is seldom a large number of observations for each individual. This means the estimated fixed effects contain measurement error. As shown by Andrews et al. (2008), this introduces bias into the estimated variance and covariances of the fixed effects. To overcome those biases, we implement the correction procedures described in Andrews et al. (2008) and Gaure (2014) to the covariance of our fixed effects. Across the two companies, we can estimate 3,584 manager fixed effects separately from 2,773 store fixed effects.

In addition to those procedures, we use two different methodologies to directly address the measurement error in our estimates and recover "true" effects. The first is an Empirical Bayes adjustment where, given prior distributions for measurement error and true manager/store effects, we construct a posterior update for those distributions using the estimated data. The mean of this posterior is the adjusted fixed effect (Chandra et al., 2016). Our second approach abandons the estimation of individual fixed effects and instead focus on groups of similar units (Bonhomme et al., 2019). This process drastically reduces the number of estimated parameters, thus reducing measurement error, but it has the setback of being dependent on the choice of groups.

In either case, these adjustments allow us to learn about the joint distribution of manager and store fixed effects. Its shape and range in particular tell us a lot about the extent to which managers influence stores' productivities. This is another element that distinguishes our approach from previous studies, which had been more concerned with particular cases

of measurement error bias, such as attenuation when including estimated fixed effects as regressors (e.g., (Bertrand and Schoar, 2003)).

We obtain four major findings from monthly operations and revenue data for every store of the two large retail companies. First, managers matter for productivity. While store fixed effects are somewhat more diffuse and explain a greater fraction of productivity variation, managers still explain an important share. In contrast to our expectation, the covariance between manager and store effects tends to be negative in our data. Taken at face value, this means that either there are no substantial complementarities between managers and stores (Becker, 1973), or that firms are not taking advantage of them in their manager allocation decisions. However, we show that productivity-based selection on manager-store pairings may impart a negative bias on this covariance, perhaps clouding true complementarities. This result has broader implications for empirical tests of complementarities in settings where potentially complementary inputs are subject to performance thresholds.

Second, there are changes at the manager and store level when new managers replace former managers at stores. In one company, old managers' earnings fall in the months prior to their replacement. On the other company in our dataset, it is their stores' productivity and sales that falls before their departure, but this reverses once the new manager arrives. New store managers are newer to the firm: they average almost 17 months shorter tenure in one company, 6.5 months in the other. Despite those differences, incoming managers are not systematically better than outgoing managers by our metric: the distribution of changes in managerial quality (fixed effect) is centered around zero and its average is not statistically distinguishable from zero at the 10% level. In contrast, managers who move between different stores in the same retailer tend to move to better (higher fixed effect) stores.

Third, we characterize which observable features in our data are correlated with manager and store quality. We find a statistically significant relationship between manager quality and the ratio of full-time to part-time workers in their stores, although the sign of the effect is different between the two companies. We also find a positive correlation between manager quality and energy efficiency, suggesting that good managers are like TFP in that they raise the productivity of multiple inputs. Finally, and as mentioned above, there is strong negative correlation between manager and store quality. That said, most observable features do not have any statistically significant effect on manager quality, including the manager's tenure, gender, distance to competitors, and even wages. In addition, a large part of the variation in manager quality still goes unexplained, which suggests that other, less

measurable, characteristics associated with the manager (e.g., leadership style and charisma) could play an important role.

Higher quality stores tend to be in locations with higher average incomes, even among the network of stores that share the same managers (i.e., the stores for which their fixed-effects are comparable). More productive stores also pay higher wages, which could explain the average increase in store quality when managers switch between stores in the same company. The remaining drivers of store-level productivity are somewhat similar to managers'. The ratio of full-time to part-time workers is correlated with the store's quality, but the sign changes between companies. Better stores are also slightly more energy efficient, and the correlation with manager quality is negative. Distance to competitors, store size (floor area), or manager gender do not have any effect on store quality.

Fourth, the end of our sample includes one year of the COVID-19 pandemic period. This allows us to measure manager performance during both stable and turbulent times. While research has investigated how companies have responded to disruptions Decker et al. (2018); Aghion et al. (2021) and during the COVID-19 pandemic in particular Bell et al. (2021); Bloom et al. (2020); Andrews et al. (2021); Bloom et al. (2021); Barry et al. (2022), no previous work has examined managers' behavior and performance in response to such shocks. We find that managers with high fixed effects (estimated using only data during stable periods) also performed better than average during COVID.

Moreover, using detailed store-level data on exogenous disruptions before and during COVID-19, we find disruptions do have a strong negative impact on productivity. Higher-type managers are no better on average than lower-type managers at limiting the productivity effects of COVID-19 disruptions. However, managers of any type exposed to more disruptions pre-COVID did better at mitigating COVID-related productivity disruptions, though this was especially so for low types. A possible interpretation is that "worse" managers can learn more from experience with previous shocks, perhaps because they have more room to grow in the first place.

Our research here speaks to existing work on ways managers improve the productivity of high-performing workers (Lazear et al., 2015; Best et al., 2017; Adhvaryu et al., 2019; Frederiksen et al., 2020; Hoffman and Tadelis, 2021; Limodio, 2021). It also has ties to the matching literature in labor economics (Eeckhout and Kircher, 2018) and more specifically to the negative assortative matching result in (Adhvaryu et al., 2020), who examine how

managers choose workers to increase productivity.[1] The two closest papers are Lazear et al. (2015) and Fenizia (2022). The former shows that supervisors of computer-based test graders matter: where the difference in the 10th to 90th percentile supervisor is equivalent on output as adding one worker in a team of 9. The latter, using a movers design, shows that within the Italian social security department, manager fixed effects explain 9% of the total variation claims productivity at the office level. Our result on the impact of managers on productivity is much larger than both of these results. We show the difference in the 10th to 90th percentile manager is equivalent on productivity as adding one worker in a team of 4, and managers explain 25-35% of the overall productivity variation. There could be many reasons for these differences from our two multi-billion dollar companies to the settings they use, namely these are private retail units, where managers might matter more to the profitability of the industry.

We extend the two-way fixed effects sorting models of labor across firms (Abowd et al., 1999; Card et al., 2013; Lopes de Melo, 2018) by examining how the manager impacts on productivity and not just whether their wages change in the new firm/location. This is important because wages do not always map well onto productivity and there are a range of identification and measurement issues using wage data (Eeckhout and Kircher, 2011; Hagedorn et al., 2017). Such measurement and identification issues do not exist when revenues and costs are observed within each unit that the manager manages.

The remainder of this paper is structured as follows. Section 2 highlights the data from the two companies, and section 3 then estimates the manager and store fixed effects. Section 4 goes into depth on explaining the causes and effects of managerial changes within the firms, and section 5 calculates what explains manager and store productivity. Section 6 then uses the COVID-19 shocks to understand how manager effects mediate responses to the pandemic, and section 7 concludes.

2 Data

Our dataset comprises monthly operations data at the store level for two large retail companies. Data for Company A, an American company, is available between April of 2018 and December of 2020 (we divide this sample into pre- and post-COVID-19 periods with

[1] It is also related to choosing social movement (union) leaders on preferences for wage changes (Boudreau et al., 2021).

the threshold being February 2020; COVID-19 seriously disrupted the company's regular operations). Data for Company B, a British company, is available between April of 2014 and May of 2017.

In both cases, the data includes thousands of stores located across practically the whole territory of their respective countries. For each store, we observe monthly sales, full- and part-time employment, store size (floor area), location (city and state/region), other factor inputs, and a set of additional store characteristics. Importantly, we also observe the identity of the manager in each store, which allows us to track managers as they move across establishments over time.[2]

Stores change managers relatively frequently, offering us reasonable leverage to identify manager effects. In both companies, more than half of stores change their manager at least once during our sample. Of those, about 60% change managers only once, 30% change managers twice, and the remainder more frequently than that. Similarly, 14 or 23% of managers (depending on the company) in our sample work in two different stores, and a small fraction move through three or more. Between 20 and 30% of each company's manager changes are internal; that is, involve moving a manager between different stores of the same company.

We do not observe the reasons why managers left their stores (this issue is better explored in section 4), and therefore cannot rule out the possibility that manager departures are related to store characteristics. Nevertheless, we do see a significant overlap between the distributions of outcomes across "mover" and "non-mover" stores (see table B.1). In addition, figure A.1 shows that average outcomes within these two groups of stores follow very similar paths over time, despite differences in levels (see appendix E for a more detailed discussion on pre-trends).

3 Manager and Store Effects

In this section we detail how we decompose store productivity into manager and store fixed effects. We discuss the identification of those effects and the steps we take to mitigate estimation error.

[2]We drop the handful of stores in each company that have multiple managers within a given month.

3.1 Measuring Productivity

In this paper, we measure store-level productivity in each period as revenue (sales) per full-time equivalent employee:

$$prod_{s,t} = \frac{sales_{s,t}}{employment_{s,t}}.$$

There are two main benefits for adopting this definition. First, revenue per employee can be directly constructed in the data and is intuitively related to less observable primitives driving stores' performance levels. Second, it is a widely used metric and likely to reflect the performance indicators that managers and companies target. In this sense, our definition of productivity may well be more aligned with managers' incentives than underlying demand or supply primitives.

Like all productivity measures, this metric is a residual. Any variation in revenues unaccounted for by labor inputs or other controls will be in our productivity measure. This might include variations in stores' physical capital, locations, or nearby competitors. In section C.1 of the appendix, we compare sales per employee with other potential measures of productivity, such as sales per store area (which proxies for physical capital), and revenue-based TFP (which controls for all observable inputs). We find all measures are highly correlated and appear to capture the same underlying store characteristics.

3.2 Decomposing Productivity

We decompose each store's productivity into three components: a store-specific component, a manager component, and a time (month) component. Formally, we estimate the fixed effect model

$$\log(prod_{s,t}) = \mu_s + \mu_{m(s,t)} + \mu_t + \varepsilon_{s,t} \tag{1}$$

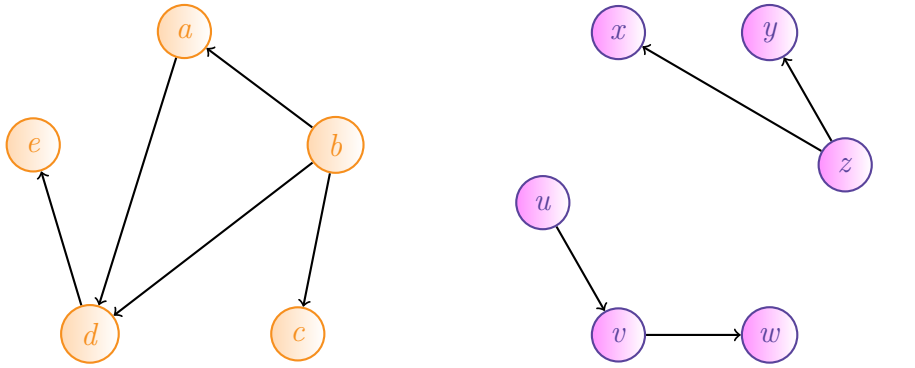
where μ_s is the store fixed effect, $\mu_{m(s,t)}$ the fixed effect of the manager associated with store s and time t, μ_t a time fixed effect, and $\varepsilon_{s,t}$ the residual. The store fixed effect in this model captures any time-invariant influences on store productivity, be it physical size, location, type of establishment, local competition, or other possibilities. The time fixed effect absorbs any company-wide productivity shocks.

Note that the log-linear regression specification (1) imposes assumptions on the com-

plementarity between the manager and store components of productivity. Specifically, it implies that the level of productivity is supermodular in the store and manager fixed effects (assuming their elasticities are positive). Section C.2 of the appendix digs into this issue, showing that more general specifications produce results very similar to our baseline.

Estimation of this fixed effects structure brings a series of challenges. First, not all store and manager fixed effects can be separately identified. This point can be better understood if we think of μ_s and $\mu_{m(s,t)}$ as the coefficients on a complete set of dummies for stores and managers. If a store has had the same manager for the duration of our sample, there is no way to separately identify each fixed effect, since the manager and store dummies are perfectly collinear. As a result, we rely on managers that move across stores for identification.

Figure 1: Examples of Store Networks



(a) One connected set (b) Two connected sets

The network structure of stores also matters for identification. Figure 1 shows two examples of store networks. Each circle represents a store and a link between them means that these stores have shared a manager (for example, a manager that worked at store a moved to store d). Network (a) (containing stores a through e) has one single *connected set*, meaning that all stores in the network have, directly or indirectly, shared at least one manager. Because of that, we can estimate the fixed effects of managers and stores relative to a common baseline (the excluded dummy) and therefore meaningfully compare the quality of managers and stores withing that network.

In contrast, network (b) (stores u through z) has two separate connected sets, $\{u,v,w\}$ and $\{x,y,z\}$. Consider, for example, a firm that has establishments in both the East Coast and the West Coast of the US. Suppose managers are happy to move to different stores along

the same coast but unwilling to move across the country. This would create the pattern in network (b), where stores in each coast would be connected to each other, but no managers connect stores across coasts.[3]

Because there are two separate connected sets in network (b), the estimation of fixed effects requires the normalization of two separate baselines—one omitted dummy per connected set. And because the omitted dummy is arbitrary, the fixed effect of store x, for example, cannot be compared to the fixed effect of store w, as they are measured relative to different, unobserved baselines. This means that we can only compare magnitudes of fixed effects inside the same connected set (see also Abowd et al., 1999; Fenizia, 2020).

In our data, we have about 890 connected sets in Company A and just over 1,200 connected sets for Company B. However, the vast majority of these are trivial, including just a single store and manager (i.e., managers that do not move). The four largest connected sets in Company A have between 13 and 18 managers and 7 to 10 stores. They are much bigger in Company B. The largest set has 202 managers and 130 stores. Set sizes to drop quickly, however, with the next 3 largest connected sets respectively containing 100, 66, and 55 managers and 67, 44, and 32 stores. In many of our results below, we focus only on the largest connected sets of each company as those allow for more precise estimation. Figure A.4 in the appendix summarizes the structure of the connected sets in our data.

A second estimation challenge in model (1) is that the number of parameters can be very large. This poses a computational problem, as regular OLS estimation methods become slow and memory-intensive. Fortunately, a number of algorithms have been developed for this purpose. We use the `lfe` package in R (Gaure, 2013).[4]

The third major challenge in our framework involves estimation error, as our fixed effects are typically estimated with relatively few observations. Estimation error causes issues in the analysis of the fixed effects. For example, even if measurement error has mean zero and is independent from the "true" effects, it can still introduce bias into the variance of our estimates, as well as to the covariance between store and manager fixed effects, sometimes referred to as limited mobility bias (Andrews et al., 2008). We address this below.

[3]This is somewhat like what we observe in the data for Company A: the overwhelming majority of connected sets includes only stores in adjoining states.

[4]For consistency throughout the estimation process, we estimate fixed effects using the fixed-point iteration algorithm, implemented in Stata by Correia (2017).

3.2.1 Variance Decomposition

The relationship between the store- and manager-specific components of productivity is a major revelation of the productivity decomposition. We can study this relationship by looking at the variance of both sides of (1):

$$\text{Var}(\log(prod_{s,t}) - \mu_t) = \text{Var}(\mu_s) + \text{Var}(\mu_{m(s,t)}) + 2 \times \text{Cov}(\mu_s, \mu_{m(s,t)}) + \text{Var}(\varepsilon_{s,t}). \quad (2)$$

The left-hand side of this equation has the variance of logged productivity adjusted for time effects, while the right-hand side splits this variance into the variances of the store and manager fixed effects, their covariance, and the variance of the residual (assumed to be exogenous). Two things are of primary interest in this equation: (1) how much of the variance of productivity can be attributed to each component, and (2) the sign of the covariance between the manager and store fixed effects.

Recall from our discussion of estimation issues that this decomposition only has meaning inside connected sets, as manager and store effects are normalized to different levels across sets. In addition, if our estimates of μ_s and μ_m contain measurement error, their variances will be upwardly biased. Further, although this might not be immediately clear, their covariance will be biased down (Andrews et al., 2008).[5] We correct for this bias using the method described in Gaure (2014).

Table 1 shows the ratio of the store fixed effect (FE) variance, manager FE variance, and their covariance to the variance in productivity (after time effects are removed) within each of the four biggest connected sets (CS). There are two key takeaways from this table. First, in every case the store FE variance is larger than the manager FE variance, and thus accounts for a greater share of the observed productivity variation. For example, in the biggest connected set for Company A, the variance in store FEs accounts for 72% of total productivity variation, while the variance in manager FEs accounts for only about 5%. A lot of the dispersion in store productivity levels is therefore attributable to stable features of stores themselves, like location (including local demographics, competition from nearby retailers, etc). That said, plenty of productivity variation remains even after store effects are accounted for, suggesting an important role for managers to influence outcomes.

[5] Intuitively, the bias in the covariance comes from the fact that when the measurement error in the manager fixed effect is positive, the measurement error in the store fixed effect will tend to be negative, and vice versa, as the two components sum to the same total. This conclusion is specific to our setting, however, where there are no other covariates in the model.

Table 1: Decomposition of the Variance of Productivity

CS Rank	Obs	Company A $\frac{\text{Var}(\mu_s)}{\text{Var}(\log(prod_{s,t})-\mu_t)}$	$\frac{\text{Var}(\mu_m)}{\text{Var}(\log(prod_{s,t})-\mu_t)}$	$\frac{\text{Cov}(\mu_s,\mu_m)}{\text{Var}(\log(prod_{s,t})-\mu_t)}$
1	215	0.72	0.05	-0.07
2	191	1.80	1.01	-1.14
3	171	0.21	0.12	0.15
4	142	0.48	0.07	-0.02
CS Rank	Obs	Company B $\frac{\text{Var}(\mu_s)}{\text{Var}(\log(prod_{s,t})-\mu_t)}$	$\frac{\text{Var}(\mu_m)}{\text{Var}(\log(prod_{s,t})-\mu_t)}$	$\frac{\text{Cov}(\mu_s,\mu_m)}{\text{Var}(\log(prod_{s,t})-\mu_t)}$
1	3,710	1.64	1.49	-1.22
2	1,979	0.73	0.22	-0.14
3	1,179	1.03	0.77	-0.62
4	1,008	0.39	0.30	-0.03

Columns 3-5 in this table show the ratio of the variance in the store FE (or manager FE, or their covariance) and the variance in total productivity according to equation (2). Results are computed within the four largest connected sets for each company.

The second takeaway is that the covariance between manager and store effects is almost always negative. Given that our structure implies persistent store and manager attributes are complements (log productivity is additive in the fixed effects, implying a multiplicative effect in levels and hence a positive cross-derivative of their marginal products), this is somewhat puzzling, as matching models would predict assortative matching between heterogeneous complementary inputs(e.g., Becker, 1973). We estimate that Company A and Company B could increase their sales revenue by 2.2 and 6.1 percent (USD 2.2 million and GBP 13.2 million) each month, respectively, by reallocating their managers across stores in accordance with positive assortative matching (see appendix D for details). This negative covariance is consistent across both companies in our sample and is also encountered by Fenizia (2020) when researching the matches between Italian government offices and managers, suggesting that other mechanisms might be responsible for what we observe.

Alternative Explanations. One possible explanation for the negative correlation between the qualities of stores and managers is that companies are simply unaware that they are not assortatively matched, as neither type is directly observable. This also suggests that companies would benefit from measuring and learning more about managerial quality.

Another possibility is that our model is misspecified and despite our estimates, the "true" production function does exhibit complementarities between manager and store quality. This would mean positive assortative matching is not in fact the optimal, pointing to the possibility that output is maximized by assigning good managers to bad stores. Or, even if this does not maximize output, it would reduce productivity variations across locations. Companies would find this desirable if they had risk-sharing priorities (e.g., Adhvaryu et al., 2020), or if there are dynamic gains from workers learning from the best managers.

A third explanation for the negative correlation we observe rests on a combination of the difficulty of separately observing both manager and store quality, and the selection of which matches are stable. Suppose, for example, that a firm is unable to differentiate between the impact of a manager and store on productivity, but still needs to decide how to allocate managers to stores. One plausible rule for doing so is to keep all matches such that $\mu_s + \mu_m \geq k$, where μ_s is the store quality (fixed effect), μ_m is the manager's quality, and k is an arbitrary constant. We argue that the adoption of such a rule could decrease the observed correlation between the quality of managers and stores.

To illustrate this point, let us analyze the implications of the matching process. In stable matches, we have $\mu_m \geq k - \mu_s$. As a result, for a given store quality m_s, the expected value of a manager's quality in a stable match is $\mathbb{E}[\mu_m|\text{stable match}] = \mathbb{E}[\mu_m|\mu_m > k - m_s]$ – which decreases with m_s! This drives down the correlation between the quality of managers and stores in a stable match. Interestingly, the same is true for unstable matches, since imposing $\mu_s + \mu_m < k$ eliminates any matches where both the store and the manager have high quality (but still allows for a high quality store to be paired with a low quality manager, and vice versa).

In addition to that, we must consider the limitations of our estimation strategy, which only allows us to separately identify the quality of managers that move between stores at least once. Hence, all managers for which we are able to estimate a fixed effect have been in an unstable match at least once (but might have found stable matches at some later point in time). To tackle this issue, we simulate the allocation of managers to stores in a two-round process. In the first round, managers and stores are randomly matched. By construction, the correlation between their quality is zero. In the second round, after observing the quality of the store they are matched with, managers can choose to leave their current match and be randomly assigned another store (that was vacated by another manager).

Following the rule above, we assume that managers remain in their store if $\mu_m + \mu_s \geq k$,

and move otherwise. For one million managers and stores, we draw $\mu_j \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$ for $j \in \{m, s\}$, set $k = 0$, and assume that no manager leaves the company altogether. We construct a list of manager-store pairs and their respective qualities analogous to what we estimate from our dataset, consisting of both the first and second matches of all "mover" managers while excluding managers-stores that have a stable match in the first round (whose qualities would not be separately identified). We find that, even though managers and stores are randomly assigned to each other in each round, the correlation between the quality of the "observed" pairs of managers and stores is close to -0.23, significantly below 0, due to the selection of stable matches.

We note that the selection issue we point out here is different from limited mobility bias and, to the best of our knowledge, has not been discussed elsewhere. Specifically, the limited mobility bias described by Andrews et al. (2008) can be understood as an incidental parameter issue, where a large number of firm-specific parameters is identified by a limited number of "movers" that go through that firm (see also Bonhomme et al., 2020). In particular, the bias can be reduced by simply increasing the number of workers/managers that go through that firm/store. This is not the case in the setting we describe above. Increasing the sample size changes nothing about the selection of managers to stores. It is also different from endogeneity bias, which arises when managers choose to move to a store because of unobservable characteristics of that store. As shown in our example, the *observed* correlation between manager and store quality is negative even when the store assigned to each manager is chosen randomly.

3.3 Correcting Measurement Error

The correction procedure adopted in the previous section accounts for the bias induced by measurement error on the variance-covariance matrix of the fixed effects, but does not recover the effects themselves. In this section, we focus on methods to recover the "true" fixed effects by reducing the measurement error of our estimates. In what follows, we briefly describe the Empirical Bayes adjustment technique used to recover fixed effects from noisy data (see Chandra et al., 2016). Section C.3 of the appendix introduces an alternative method that reduces measurement error by abandoning the estimation of individual fixed effects and focusing on groups of similar managers and stores (Bonhomme et al., 2019).

To implement the Empirical Bayes adjustment we assume that, because of measurement

(or estimation) error, the estimated fixed effects are random variables that are normally distributed around their true value. Formally, for each manager m and store s, let $\hat{\mu}_m$ and $\hat{\mu}_s$ denote the estimated effect and μ_m^* and μ_s^* the true effect. Then

$$\hat{\mu}_m|\mu_m^*, \sigma_m \sim \mathcal{N}(\mu_m^*, \sigma_m^2) \text{ and } \hat{\mu}_s|\mu_s^*, \sigma_s \sim \mathcal{N}(\mu_s^*, \sigma_s^2)$$

independently in both cases. Note that independence in this context means that the measurement error in a manager's estimated fixed effect is independent only from the measurement error on other *managers'* fixed effects. It does not imply that the measurement error in a manager effect is independent from the error in their associated store effect.

The prior distribution of true manager and store productivity is

$$\mu_m^*|m_m, \tau_m \sim \mathcal{N}(m_m, \tau_m^2) \text{ and } \mu_s^*|m_s, \tau_s \sim \mathcal{N}(m_s, \tau_s^2),$$

independently in either case. We assume that the underlying means m_m and m_s are constants, although the method allows for them to be functions of manager and/or store characteristics as well. By conditioning on the estimated fixed effects $\hat{\mu}_m$ and $\hat{\mu}_s$, we can find the posterior distributions for μ_m^* and μ_s^*. The Empirical Bayes (EB)-adjusted fixed effects, μ_m^{EB} and μ_s^{EB}, are the means of those posterior distributions.[6] Quantitatively, we find that the EB-adjusted and unadjusted fixed effects are highly correlated. Figure C.5 compares how each of our measures of store and manager quality is related to productivity.

Our results suggest that managers can have a significant impact on the productivity of stores—just as much each store's own characteristics. This can be seen in figure 2, which plots the distributions of the EB-adjusted store and manager fixed effects inside the largest connected set for each company. In each panel of the figure, the fixed effects have been demeaned within their respective connected set, so we focus on the range of the distributions. Note that the distributions of manager and store fixed effects have about the same range in both companies, suggesting that changing a manager can have a productivity effect just as large as a hypothetical change in the store's otherwise persistent characteristics (assuming for the sake of argument that the observed relationships between store characteristics and

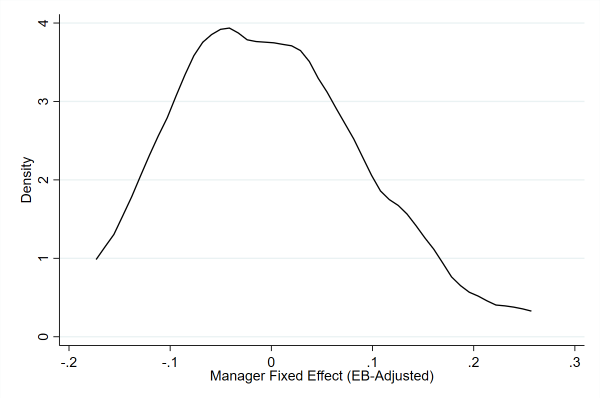
[6] We refer to online Appendix C in Chandra et al. (2016) and the references therein for more details on the methodology employed.

productivity are causal).[7] In Company A, for example, replacing a very low-quality manager by a very high-quality one could increase the store's productivity by almost 45 log-points (over 50%). In Company B, the effects would be considerably larger, though there is also more variation in the store effects in this company as well.

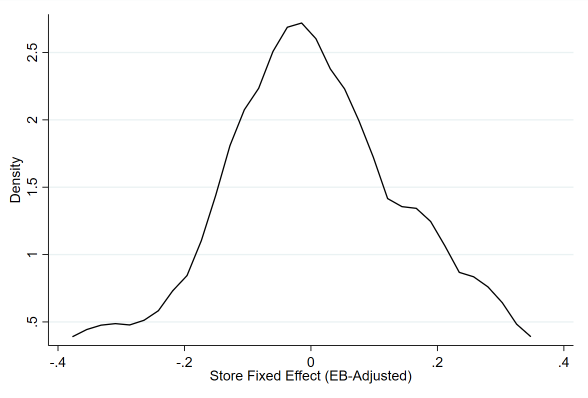
These results above suggest that managers have a large scope to influence the productivity of the stores at which they work. Figures A.2 and A.3 show that those findings are more broadly valid by plotting the EB-adjusted manager and store FE distributions in each of the four largest connected sets in our data. We end this section by adding the caveat that the comparisons here are only valid inside connected sets, and it may be the case that there is more variation among stores and managers across connected sets than within them. In other words, managers' ties to store-level productivity might be even larger than the numbers we estimate here.

[7]This observation might seem at odds with the conclusions from table 1. However, note that the variances in that table are computed across all periods in the dataset, so the fixed effect of each store and manager is effectively weighted by the number of periods they are present in the data. The distributions in figure 2 count each store/manager only once.

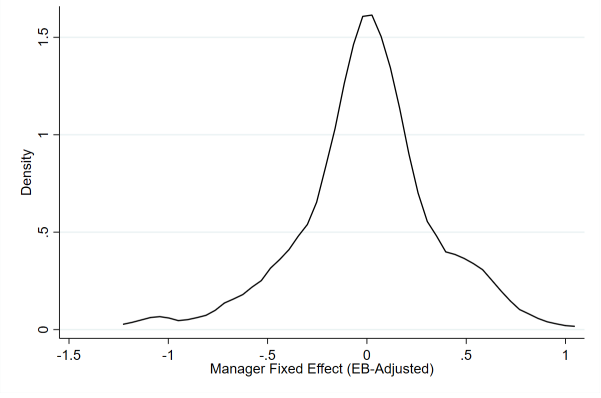
Figure 2: Distribution of Manager and Store Quality (EB-Adjusted)



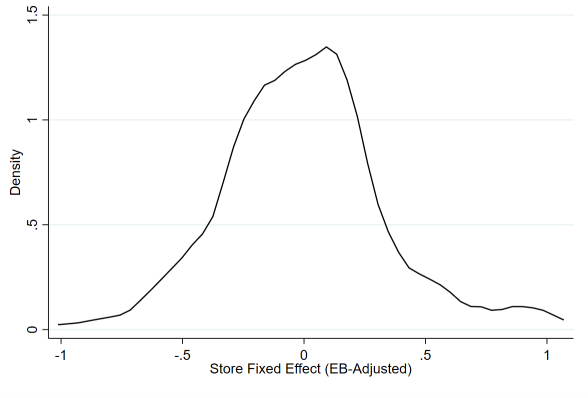
(a) Manager FE - Company A



(b) Store FE - Company A



(c) Manager FE - Company B



(d) Store FE - Company B

4 The Causes and Effects of Managerial Changes

We now address some of the potential causes that drive managers to leave a store, and what happens when a new manager is brought in. We estimate event study specifications assessing how store outcomes shift in the periods around a management change. We also ask whether stores that receive a new manager receive one who has a higher FE than the manager who left, and similarly whether managers who switch stores end up moving to more productive stores.

We estimate the following equation:

$$y_{st} = \alpha_s + \alpha_t + \mathbf{1}\{s \in \mathbb{T}\} \left[\sum_{k=-6}^{k=3} \beta_k \mathbf{1}\{K_{st} = k\} + \beta_{4+} \mathbf{1}\{K_{st} \geq 4\} \right] + \varepsilon_{st}, \tag{3}$$

where $y_{s,t}$ a store outcome; α_s and α_t are respectively store and time fixed effects; $\mathbf{1}\{s \in \mathbb{T}\}$ indicates whether store s belongs to the "treated" group $\mathbb{T}$ (i.e., changed its manager); and $K_{st} = t - E_s$ is the number of months before/after the time of event E_s (when a manager moves to/from store s). Note that $\mathbf{1}\{s \in \mathbb{T}\}$ does not have a time subscript, indicating that stores that switch managers are in the treatment group even before the switch has happened. In contrast, stores in the control group are the ones that never changed their manager. We define the treatment and control groups in such a way for two important reasons. First, it avoids the identification issues discussed in Borusyak and Jaravel (2016). Second, a change in the store manager is often an expected event. As a result, it can affect store outcomes before it actually happens, making simple before-and-after comparisons less informative.

We construct the control group for each treated store by finding its nearest neighbors—the untreated stores that most resemble the each treated unit before treatment—according to the following metric: we estimate the conditional probability of treatment (propensity score) for each store s using observations that date to at least 6 months before the change in management takes place. Our empirical model is

$$\text{logit } \mathbb{P}\left(\text{mng change}_s\right) = \gamma_1 \log(revenue_s) + \gamma_2 \log(employment_s) + \gamma_3 \log(area_s) + X_s' \Gamma_0.$$

The dependent variable indicates if store s switched managers at any point in the data. The first set of regressors (sales revenue, FTE employment, and total area) control for a store's size and productivity. The remaining regressors in X_s control for the store's location (a state fixed effect) and entry and exit dates. All the variables in the right-hand-side are constructed by averaging the data for each store over time.

We match each treated store to its nearest untreated neighbors using the odds ratio as our distance metric.[8] We then construct the outcome of each control store $y_{s,t}(control)$ as the average of the outcomes of the 3 nearest neighbors of treated store s. Once this is done, we estimate (3) by noting that it maps into a potential outcomes framework

$$\Delta y_{st} = \sum_{k=-6}^{k=3} \beta_k \mathbf{1}\{K_{st} = k\} + \beta_{4+} \mathbf{1}\{K_{st} \geq 4\} + e_{st}, \tag{3'}$$

[8] Estimation is done through the Stata module psmatch2 (see Leuven and Sianesi, 2003).

where $\Delta y_{st} = y_{s,t}(treated) - y_{s,t}(control)$. As discussed above, $y_{s,t}(treated)$ is the outcome of the treated store (observed) and $y_{s,t}(control)$ is the outcome of the constructed control for s.

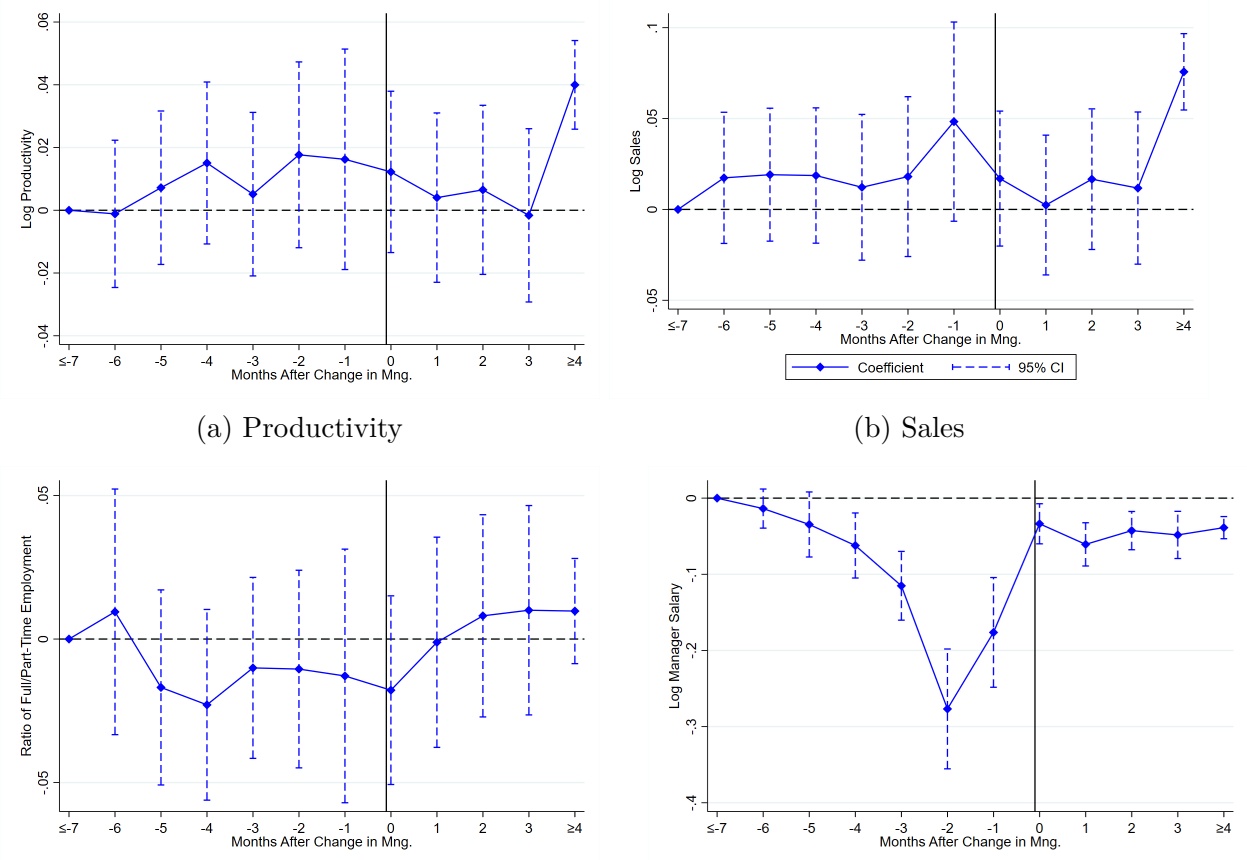
Figures 3 and 4 plot the estimated coefficients from model (3') for four outcome variables: productivity, sales, full-time to part-time employment ratio, and manager earnings (only available for Company A). Interestingly, the scenarios that lead to the separation of a manager are very different in companies A and B, which highlights the importance of "company culture" in determining the mobility (and potentially the impact) of managers. Looking at the event studies for Company A, we see no statistically significant impact on productivity, sales, or the composition of employment around the time a manager leaves a store. However, there is a massive reduction in the manager's earnings (by almost 30% at the lowest point) a couple of months before he/she leaves the store, and an immediate recovery once the new manager takes over.

This reduction could be related to reason the manager is leaving in the first place. For example, if bonuses drop due to poor performance. In appendix E, we test whether the same result holds true for managers who move to another store but stay within the company. These moves are less likely to be due to poor performance at the initial store. We find that the effects in figures 3 and 4 are still mostly present, but considerably attenuated, which is circumstantial evidence in favor of the hypothesis above. We also refer to appendix E for evidence on the absence of pre-trends during the months used for matching.

While manager salaries are not available for Company B, patterns of productivity, sales, and employment composition around a managerial change paint a completely different picture than the one seen in Company A. There is a clear and steady decline in sales in the months leading to a manager's departure, leading to a simultaneous decline in productivity. Evidence that these changes are related to the reason for the manager's departure is seen in that the trends of both productivity and sales revert shortly after the arrival of the new manager. The pattern is damped but still present for managers who stay with Company B at a different store (see appendix E).

Lastly, one pattern common to both companies is that the point estimates imply new managers accompany an increase in the ratio of full-time to part-time workers. In both cases, however, the standard errors are large enough that the effect cannot be distinguished from zero at the usual confidence levels.

Figure 3: Event Studies: Changes in Outcomes Around Manager Change
Company A



(a) Productivity

(b) Sales

(c) Full/Part-Time Emp. Ratio

(d) Manager Salary

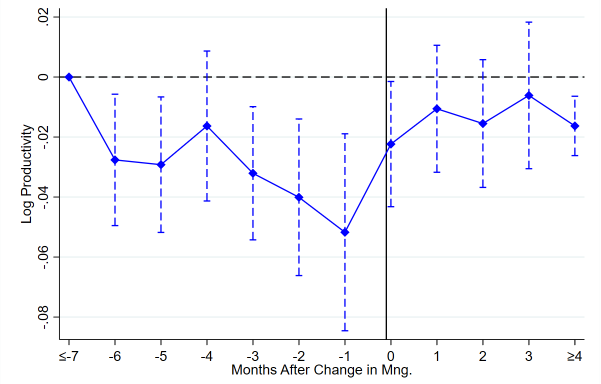
Note: Dots show the value of each coefficient β_k, while whiskers indicate the 95% confidence interval (computed using robust standard errors). The normalized period (≤ -7) indicates coefficients for periods that are precede the manager change by 7 or more months. Alternative specifications (e.g., using only 7-10 months prior to change) produce very similar results.

4.1 Are Stores Switching to Better Managers?

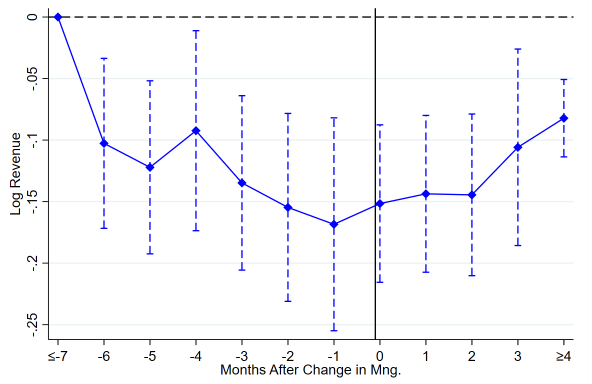
The event studies above shed some light into the reasons that might lead to a separation between manager and store. In this section we explore whether newly hired managers are, on average, different from those they replaced. We consider two manager attributes: their experience (tenure in the company) and their fixed effect.

Figure 5 plots the distribution of differences in tenure between the new and old manager after the switch happens. At both companies, the average incoming manager is less experi-

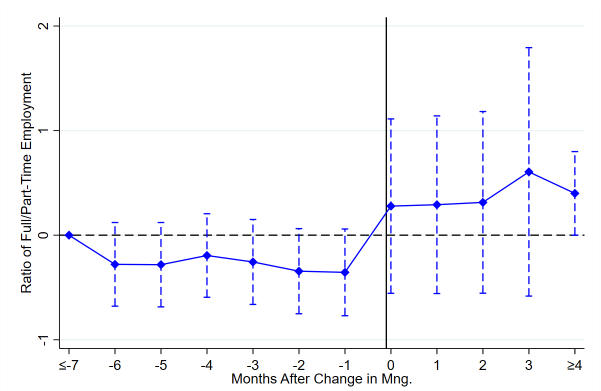
Figure 4: Event Studies: Changes in Outcomes Around Manager Change
Company B



(a) Productivity



(b) Sales



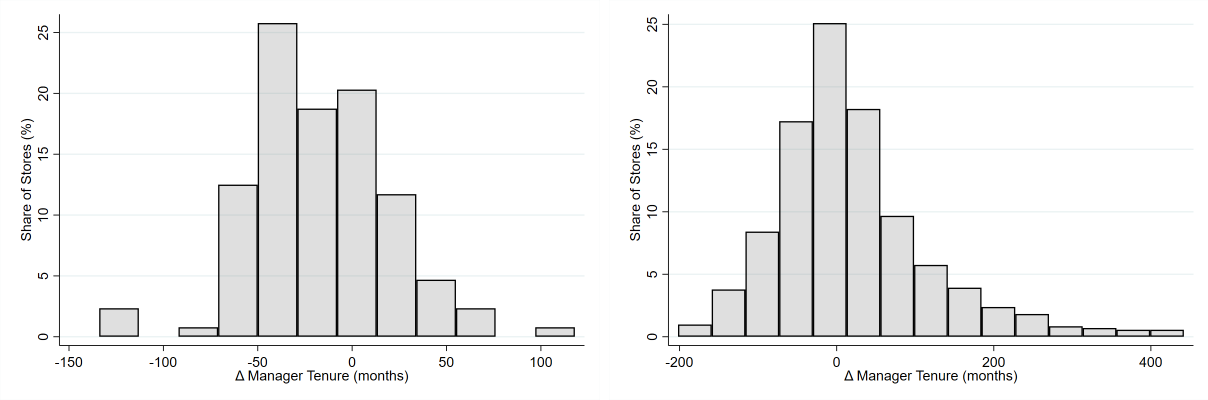
(c) Full/Part-Time Emp. Ratio

Note: Dots show the value of each coefficient β_k, while whiskers indicate the 95% confidence interval (computed using robust standard errors). The normalized period (≤ -7) indicates coefficients for periods that are precede the manager change by 7 or more months. Alternative specifications (e.g., using only 7-10 months prior to change) produce very similar results.

enced than the outgoing manager. The average difference in tenure in Company A is around 16.8 months and significantly different from zero at the 1% level, while it is about 6.5 months and not statistically significant at the usual levels in Company B. Compared to the range of the distribution of changes in tenure, those average differences are relatively small.

The distribution of the differences in incoming and outgoing managers' EB-adjusted fixed effects are shown in figure 6. Panel (a) plots the results for Company A, and panel (b) does so for Company B. Because all managers that have worked in the same store are always in the same connected set, their effects can be directly compared. In both companies, the

Figure 5: Distribution of Changes in Manager Tenure



(a) Tenure Change (Company A)

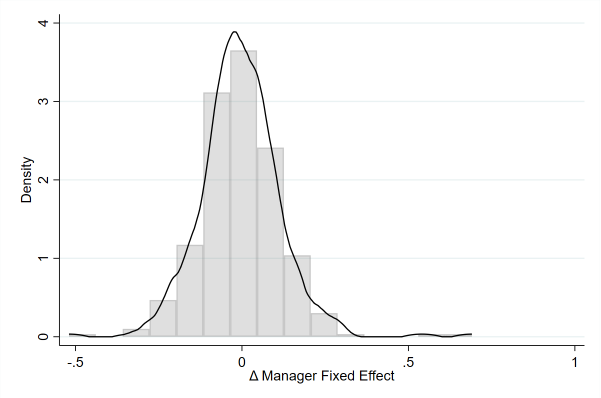
(b) Tenure Change (Company B)

distribution is approximately symmetric around zero, and mean difference is not significantly different from zero at a 10% significance level. Thus, despite the circumstances that might lead a manager to leave a store, on average, the new manager is no different from the one they replaced.

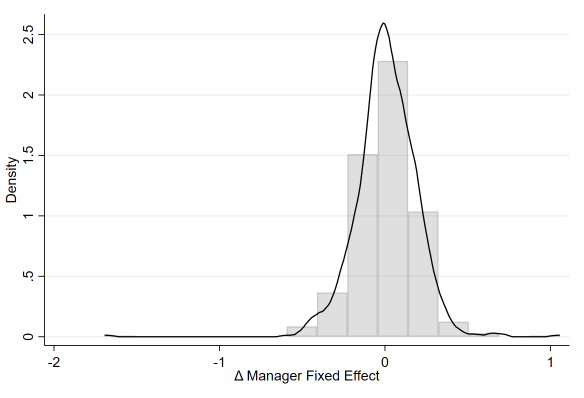
The same exercise can be done from the viewpoints of managers: after moving to a different store, how does the new store quality compare to that of the one they left? Panels (c) and (d) of figure 6 show, for Companies A and B respectively, the distribution of changes in store fixed effects when managers move. On average, the quality of the store a manager moves to increases in both companies, but this is more pronounced in Company A. Not only is the average difference larger (approximately 0.05 for Company A versus 0.02 for Company B), the entire right-side of Company A's distribution of changes in the EB-adjusted store fixed effect is bulkier.

The fact that stores do not on average receive better managers, but managers do move to better stores on average could indicate a few interesting features of the manager-store allocation process for the companies in our data. One potential explanation for this result is that manager quality is hard for the firms to measure and therefore to base allocative decisions on. In addition, managers might have more incentive to learn store quality, as it affects their store's overall productivity and as such might impact bonuses or other types of compensation. We explore those possibilities in the next section. However, by simple comparison of means, we can say that managers in Company A tend to move to stores that (a) are larger, (b) pay higher wages, and (albeit with less statistical precision) (c) are located

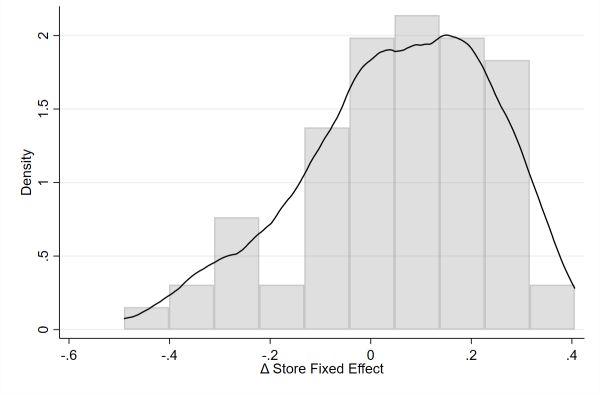
Figure 6: Distribution of Changes in Manager and Store FE



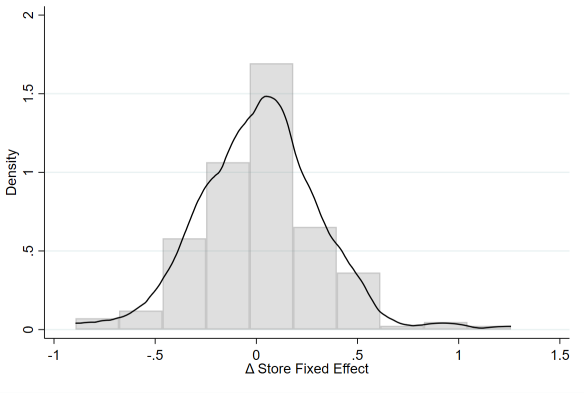
(a) Manager FE (Company A)



(b) Manager FE (Company B)



(c) Store FE (Company A)



(d) Store FE (Company B)

in more densely populated regions.

5 What Explains Manager and Store Productivity?

In this section, we explore whether observable characteristics and attributes in our data can explain variation in manager or store FEs. As a starting point, we assume that the quality of a manager can be described by

$$q_{m,t} = X_{s(m),t}\beta + q_{s(m),t} + \varepsilon_{m,t} \tag{4}$$

where $s(m), t$ indicates the store s where manager m worked in month t. Given our discussion above, we also assume

$$q_{m,t} = \mu_m + \epsilon_{m,t},$$

where μ_m is the manager's fixed effect in equation (1). Plugging this expression into the equation above and taking the average over all periods (since the manager fixed effect is invariant across time), we find

$$\mu_m = \bar{X}_m\beta + \frac{1}{T}\sum_{t=1}^{T}\sum_{s\in S(m)} \mathbf{1}\{s = s(m), t\}\mu_s + \tilde{\varepsilon}_m,$$

where $S(m)$ is the set of stores where manager m was employed. Note that the RHS of this equation simply reflects the average store features ($\bar{X}_m$) and quality (μ_s) across each store where manager m was employed (weighted by the number of periods that the manager spent at each store). In similar fashion, $\tilde{\varepsilon}_m = \frac{1}{T}\sum_{t=1}^{T}\varepsilon_{m,t} - \epsilon_{m,t}$ is a combined residual.

Similarly, we have that the quality of stores can be expressed as

$$\mu_s = \bar{X}_s\beta + \frac{1}{T}\sum_{t=1}^{T}\sum_{m\in M(s)} \mathbf{1}\{m = m(s), t\}\mu_m + \tilde{\varepsilon}_s.$$

In our empirical analysis, we use the EB-adjusted manager and store fixed effects as the measure of μ_m and μ_s. To account for the fact that FEs are normalized differently in each connected set, we include a full set of connected set dummies in all regressions.

Manager/store quality and consumer satisfaction Along with accounting data, we have also obtained the results from consumer satisfaction surveys at the store level for company A, starting in January 2019. Those surveys include "overall" scores given by customers

to each store, such as the likelihood to recommend the store to others and the satisfaction with the service or the overall experience, and also detailed features such as helpfulness, availability, how knowledgeable the staff/manager was, among others.[9] As with productivity, the outcome of the consumer satisfaction survey is affected by both store and manager features. Hence, we model those outcomes as

$$CS_{s(m),t} = \delta_m + \delta_s + \delta_t + u_{s(m),t},$$

where $CS_{s(m),t}$ is the outcome of the survey for store s, where manager m was employed at time t. The δ's represent fixed effects and $u_{s,t}$ is a residual.

This specification for the outcomes of the consumer satisfaction survey creates two significant hurdles. First, the survey includes several possible outcomes/questions, and it is not immediately clear which should be used to explain the variation in store/manager quality. Hence, we first calculate the first principal component of the scores for all outcomes, $CS^{(1PC)}_{s(m),t}$, and use that as a one-dimensional measure of consumer satisfaction.

Second, the estimation of δ_m and δ_s carry the same restriction discussed in previous sections. In particular, the estimated δ's would only be comparable inside connected sets. However, because the consumer survey data is available only after January 2019, the network structure of this data (i.e., the connected sets) is different from the network structure of the accounting data, which starts in April 2018. This means that fixed effects estimated from the consumer survey data might not be comparable to the fixed effects estimated using productivity. To avoid this issue, we instead regress $CS^{(1PC)}_{s(m),t}$ on a store (manager) and time fixed effect. The residual, $v_{m,t} = \delta_m + u_{s(m),t}$ (or $v_{s,t} = \delta_s + u_{s(m),t}$), is used as a measure of the manager's (store's) impact on consumer satisfaction. Once $v_{m,t}$ is computed, it can be inserted into (4) and we proceed the analysis in the same way as above.

Note that the methods we use to deal with the issues outlined above might have important consequences for our results. We discuss alternatives and their outcomes in the robustness checks subsection below.

[9]The full list of features include: how likely the customer is to recommend the store, satisfaction with the service, satisfaction with the experience, helpfulness, knowledgeable, treated like a valued customer, availability, ease of checkout, appearance of the store, selection, value for price paid, and quality. Each of those is given a number between 0 and 10, with 10 indicating the best outcome.

5.1 Results

Manager quality. Most observable features seem to be uncorrelated with managerial quality. Table 2 shows selected coefficients, with column (1) being the baseline regression, column (2) includes the average EB-adjusted fixed effect across the stores where each manager has worked, and column (3) adds the consumer satisfaction survey outcomes (available only for company A).

The first thing to note from table 2 is that there is no statistically significant relationship between a manager's quality and their salary. It is interesting to note that the coefficient on wages is negative in the baseline regression, but switches sign when the average store fixed effect is introduced. This could indicate that a large portion of a manager's compensation is due to the store where he/she works (note from table 3 that better stores pay higher wages). However, in either case the effect is too imprecise to draw conclusions.

Manager earnings are unfortunately not observed for Company B; however, we do observe but energy costs (electricity and gas). We construct a measure of energy efficiency as the ratio of sales revenue to energy costs and find that it is positively correlated to the manager fixed effect. To some extent, this result is expected, as energy is an input cost to stores, and more productive managers would take actions to reduce costs per unit of output. In practice, however, this might not be as obvious, as energy efficiency tends not be to focus of most managers. We develop this line of inquiry a bit more in appendix F, where we present evidence that more productive managers can reduce the consumption of energy itself, so that the perceived higher energy efficiency is not a consequence of lower energy prices. The evidence here supports prior work by Bloom et al. (2010); Martin et al. (2012); Boyd and Curtis (2014), who find that well-managed firms have superior environmental performance. Our results suggest that managers themselves (independent of management policies and other company-level factors) influence energy performance.

We also find that the ratio of full-time to part-time employees has an economically and statistically significant effect on the quality of managers. However, the sign of the coefficient is different across stores. This contrasting result highlights a potential trade-off in hiring decisions: part-time workers offer more flexibility on staffing decisions, while full-time workers tend to be better trained and might reduce moral hazard issues. Company-wide features, as well as manager tastes, could influence which option is more appropriate.

Again, we find a negative relation between store and manager fixed effects, corroborating

our earlier results. In fact, it is by far the largest coefficient we estimate, which leads us to conclude that a manager's match (to a store) is by and large the most important predictor of the manager's quality. However, this result is subject to the caveats discussed in section 3.2.1, which suggest that the negative correlation we observe might not be causal. Finally, we find a very small and statistically insignificant coefficient for the relationship between manager quality and consumer surveys. While puzzling, this is a robust result, which we discuss more below.

It is also worth to draw attention to the R^2 of each regression, which leaves a large portion of the variation in manager quality unexplained.[10] This suggests that other, potentially unobserved, individual manager features play an important role in determining their productivity.

Table 2: Regression of EB-Adjusted Manager Fixed Effects on Observable Characteristics

	Company A			Company B	
	(1)	(2)	(3)	(1)	(2)
log manager salary	-0.045	0.011	0.042		
	(0.031)	(0.035)	(0.037)		
log energy efficiency				0.021	0.079***
				(0.029)	(0.020)
log FT/PT emp. ratio	0.058**	0.094***	0.098***	-0.026***	-0.044***
	(0.024)	(0.027)	(0.029)	(0.010)	(0.006)
EB-adjusted store FE		-0.531***	-0.542***		-0.727***
		(0.080)	(0.084)		(0.060)
Consumer satisfaction (1PC)			-0.003		
			(0.005)		
N	1,222	738	624	2,120	1,541
Within R^2	0.025	0.334	0.363	0.012	0.587

Observations are weighted by the number of periods each manager is in the data. Standard errors are clustered at the connected set level and shown in parenthesis. ***, ** and * indicate that coefficients are significantly different from zero at the 1%, 5% and 10% levels, respectively.
Note: omitted controls include the number of hours open in the week and over the weekend, and the distance to each store's closest competitor.

[10]Note that this is the *within-group* R^2; that is, it measures the explanatory power of the model *inside* each connected set. The overall R^2 in all regressions is above 0.99 – but note that a lot of the variation in manager fixed effects across connected sets is in some sense artificial, as it is created by the normalization of store/manager effects within each connected set.

Store quality. We perform a similar exercise for stores, with results shown in table 3. Column (1) regresses the store's EB-adjusted fixed effect on attributes of the store's location, namely population density and income per capita.[11] Columns (2) adds store specific features, including the manager's compensation and full-time to part-time employment ratio, as explanatory variables. Column (3) adds the average EB-adjusted fixed effect of all managers that worked at the store, as well as a location fixed effect to capture potential geographical features other than income and population density. Finally, column (4) includes the residualized first principal component of the consumer satisfaction survey (only available for company A). As before, all regressions also include a fixed effect for each connected set. For company B, we also include a dummy for the store's format (flagship store, small local, airport, etc).

More productive stores tend to be located in places with higher average income. While not surprising, recall that because the regression includes connected set fixed effects, all identifying variation resides within those sets. If stores in the same connected set are also located in similar places, our coefficients understate the actual correlation between store quality and local features across the full sample.

We also find that high quality stores have managers that make higher earnings, regardless of their quality. However, when the average quality of the managers in each store is included as an explanatory, the coefficient on managerial salaries drops by more than half and becomes statistically indistinguishable from zero. The coefficients on energy efficiency, full-time to part-time employment ratio, and average manager quality all follow the same pattern as in table 2. Those results could just reflect similar underlying economic forces, or the fact that those attributes are influenced by both the manager and the store (e.g., through company policy or location).

5.2 Robustness Checks

We estimate several alternative models to asses the robustness of our findings above. Our robustness checks can be roughly classified into three different categories, based on possible issues that arise in estimating the effect of manager or store characteristics in their quality.

[11] For Company A, data refers to the county where the store is located, and is obtained through IPUMS-USA (Ruggles et al., 2021). For Company B, data refers to the store's NUTS3-level region and is obtained from the UK's Office of National Statistics. Local income is measured by average wages in the American Community Survey (Company A), and by the regional gross disposable household income (Company B).

Table 3: Regression of EB-Adjusted Store Fixed Effects on Observable Characteristics

	Company A				Company B		
	(1)	(2)	(3)	(4)	(1)	(2)	(3)
log population density	0.025	0.005			-0.011	-0.010	
	(0.019)	(0.175)			(0.013)	(0.015)	
log income per capita	0.262**	-0.013			0.088	0.036	
	(0.102)	(0.094)			(0.118)	(0.119)	
log manager salary		0.444***	0.183	0.152			
		(0.121)	(0.149)	(0.151)			
log energy efficiency						0.149***	0.148***
						(0.031)	(0.018)
log FT/PT Emp. Ratio		0.141***	0.129**	0.141***		-0.060***	-0.059***
		(0.041)	(0.052)	(0.052)		(0.016)	(0.012)
EB-adjusted manager FE			-0.859***	-0.842***			-0.938***
			(0.195)	(0.197)			(0.027)
Consumer satisfaction (1PC)				-0.027			
				(0.020)			
Omitted controls	No	Yes	Yes	Yes	No	Yes	Yes
Location FE	No	No	Yes	Yes	No	No	Yes
N	278	276	229	229	919	826	880
Within R^2	0.043	0.374	0.588	0.595	0.023	0.113	0.576

Observations are weighted by the number of periods each store is in the data. Standard errors are clustered at the connected set level and shown in parenthesis. ***, ** and * indicate that coefficients are significantly different from zero at the 1%, 5% and 10% levels, respectively.
Note: omitted controls include: number of hours open in the week and over the weekend, distance to the closest competitor. Store size (log floor area) is also included as an omitted control for company A. All regressions for company B include dummies for the store format (flagship, local store, etc.)

Omitted variables. There are several potential features in a manager or store that might affect their impact on productivity. In our results above, we have focused on the ones that we consistently observe across both companies in our data. Among other observed characteristics that are either only available for a select number of managers or for only one company, we have a manager's tenure (both in the company and in his or her current job), and the manager's gender. We do not find statistically significant effects of manager tenure (or log-tenure, or tenure/log-tenure squared) in the manager's FE. The same is true when including a manager's gender, which is only available for company B.

Alternative measurements. Instead of including only the first principal component of the outcomes in consumer satisfaction surveys as a regressor in each regression, we can also

include each individual outcome. Our results are not affected by this.

Alternatively, one could argue that our strategy of regressing $CS^{(1PC)}_{s(m),t}$ on time and store (or manager) fixed effects and using the residual $v_{m,t} = \delta_m + u_{s(m),t}$ as the contribution of the manager to the consumer satisfaction is prone to attenuation bias because is also includes the "pure" residual from the original regression. To address this concern, we can also directly estimate a manager fixed effect δ_m in

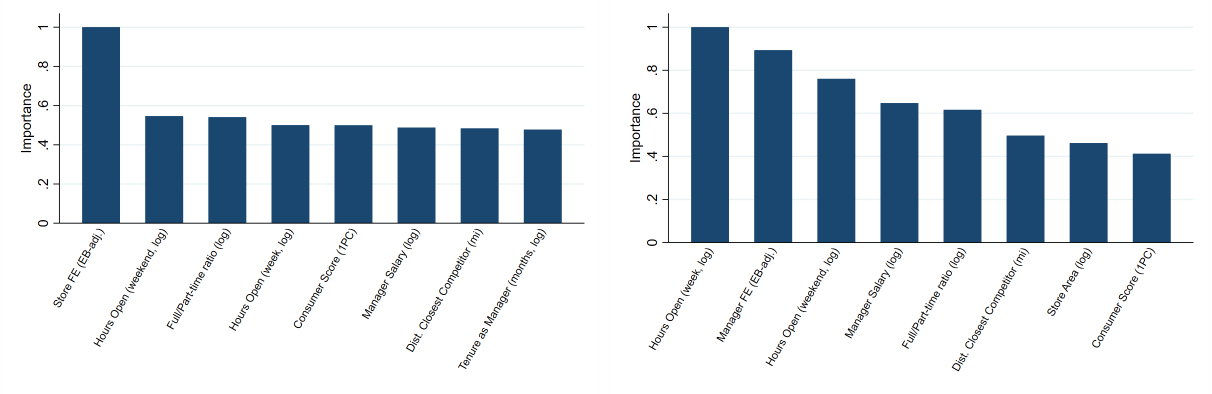
$$CS_{s(m),t} = X_{s,t}\beta + \delta_m + \delta_t + u_{s,t},$$

where $X_{s,t}$ are store-level characteristics are are meant to approximate the store's contribution to consumer satisfaction (including location, store size, number of employees, number of hours open, and distance to the closest competitor). The resulting impact of δ_m on the manager's productivity is very similar to the one found above.

Non-linear relationships. Finally, another possibility is that the linear model simply isn't appropriate to explain the variation in the productivity of managers or stores. For that reason, we also estimate a random forest model that leverages the same variables shown in tables 2 and 3 to predict manager/store productivity. To account for the normalization issues across connected sets, we first de-mean all variables by their respective connected set (which is equivalent to the inclusion of a connected set fixed effect in the linear model). Using the predicted values from the random forest model, we can calculate an "R^2" statistic and compare its predictive power with the linear model.

As expected, we find that the random forest model does a substantially better job at predicting manager and store productivity than the linear model does (in company A, the "R^2" increases to approximately 0.87 and 0.91 when explaining the variance in manager and store productivity, respectively). However, impact of each explanatory follows the same pattern as in the linear model, which we can see by plotting the importance scores in figure 7 (for brevity, we show only the results for company A).

Figure 7: Importance Matrix from Random Forest Model to Predict Manager and Store EB-Adjusted FE (Company A)



(a) Manager Productivity (b) Store Productivity

Note: highest importance score normalized to 1 in each model.

6 Previous Shocks and Adaptation to COVID-19

The COVID-19 pandemic has severely disrupted life across the globe. Companies have had to adapt quickly and, often, substantially. In this section, we investigate whether good managers were able to better adapt to the COVID-19 pandemic restrictions, minimizing the potential negative effects on their store's performance. We begin by detailing some of the ways the pandemic affected managers and stores, and then investigate how the exposure to exogenous pre-COVID shocks (e.g., natural disasters, unsafe conditions, etc.) affected managers' ability to deal with the pandemic shock. Note that because of data availability our results apply only to Company A.

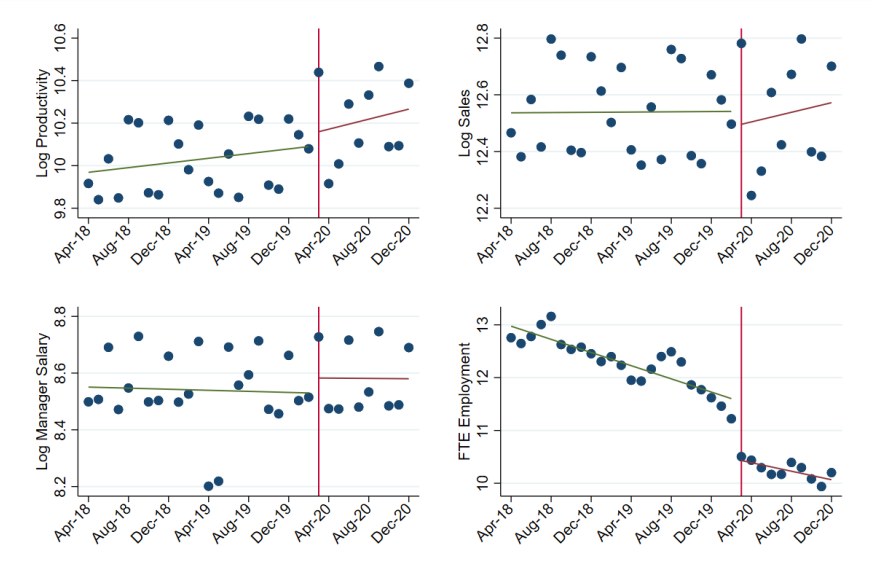
6.1 The Impact of COVID-19 Pandemic

Data for Company A is available in a monthly basis between April of 2018 and December of 2020. We define the *COVID-19 period* as starting in February of 2020, when some of the first non-pharmaceutical interventions (e.g., shelter-in-place and social distancing orders) were put in place in the US. This gives us 23 months of data before and 10 months of data after the pandemic started.

Manager turnover: Compared to pre-COVID periods, manager turnover was slightly lower after the COVID-19 pandemic started. The probability that a store changed its manager in a given was 3.4% before the pandemic and fell to 2.4% during the pandemic.

Store outcomes: Unlike what was experienced by many businesses, we do not observe a negative effect on stores' revenues or managers' earnings (see figure 8). On the other hand, there were a clear and discontinuous decrease in full-time equivalent employment when the pandemic began (despite the downward trend before COVID-19). As a result, productivity at Company A's stores actually increased during COVID-19. This may reflect another instance of the "making do with less" effect (Lazear et al., 2016) sometimes observed during turbulent times or recessions.

Figure 8: Changes in Outcomes Since COVID-19



Manager quality: We also find that, despite the many challenges brought forth by COVID-19, managers who were better before the pandemic tend to perform better during the pandemic. We demonstrate this by testing the out-of-sample fit of model (1). Specifically, we regress stores' log productivity levels on the EB-adjusted manager fixed effects, EB-adjusted store fixed effects, and a set of dummies for each month. Because the manager and store fixed effects were estimated in the pre-COVID period, the coefficient on the EB-adjusted

fixed effects before COVID-19 is, by construction, one. When we fit the same model on the data during the COVID-19 period, we find that the coefficient on the EB-adjusted manager fixed effect is positive and significant (≈ 0.5), indicating that manager performance pre-pandemic is related to performance during the pandemic. Figure A.5 plots the results. The EB-adjusted manager FE coefficient being less than one in the latter period probably reflects a combination of mean reversion in manager performance (as suggested by Lazear (2004)) and the model's lower predictive power out of sample.

6.2 Exposure to Shocks and Productivity During the Pandemic

For firms, the COVID-19 pandemic and the resulting containment measures and changes in consumer behavior can be interpreted as a series of exogenous shocks that caused the temporary closure of stores, reductions in in-store customers and staff, employee absenteeism, adoption of new cleaning protocols, and so on. Such disruptions in daily operations had already been experienced by many managers, although to a smaller extent. Examples include weather events like winter storms and hurricanes, but also power outages, police activity (which might require the evacuation of the store's block), and physical facility problems. We have data on a large number and variety of such events both before the pandemic, and we take advantage of it to see whether managers' experience with such unexpected challenges prior to COVID-19 influenced their performance during it. (A full list of the shocks we observe and their effects on store operations are in table B.3.)

We start by regressing log productivity during COVID-19 on the previously estimated manager and store EB-adjusted fixed effects, a set of variables that record the number of shocks a store experiences in a given month (one variable for each type of shock), a set of calendar month dummies, and an indicator for each connected set. The coefficients are shown in columns (1) and (2) of table 4. Most recorded shocks do not have a statistically significant effect on store productivity; for the sake of space, we only show the coefficient for "unsafe conditions due to COVID-19," which has by far the largest productivity impact.

A store that is deemed unsafe to operate experiences a severe drop in productivity. The results in column (1) imply an average drop of 43 log points, approximately 35%. In column (2) we allow the effect of unsafe conditions to vary with manager quality. We find a large base effect (for a manager at the mean of the FE distribution) of 57 log points, corresponding to a 43% drop in store productivity. However, we also see that better managers mitigate part

of the shock's effect, at a rate of about 6 percentage points per "unit" of manager quality. Recall that in the largest connected set, the range of manager fixed effects for Company A is about 0.5 (see figure 2). This means that a top manager required to close their store due to unsafe conditions will experience a monthly productivity decline that is 3 percentage points (0.5×6) smaller than at a left-tail manager's store.

Columns (3) and (4) of table 4 show the results of a similar regression, but instead analyzing the relationship between store productivity during COVID-19 and the total number of pre-pandemic shocks experienced by managers. Here, we include manager tenure as a control variable, as managers with higher tenure would mechanically have had more exposure to shocks. The results in column (3) indicate that managers' previous experiences with shocks are related to better store productivity during the pandemic, though this is somewhat imprecisely estimated. However, when we interact the number of previous shocks with managerial quality in column (4), we see a positive and significant base estimate and, interestingly, a negative interaction between the managers' number of previous shocks and quality. One possible interpretation is that worse managers gain more from having had previous experiences with shocks, potentially because they had more room to grow in the first place.

The regression in column (4) also includes a dummy that indicates whether the store experienced any type of shock in the current month, as well as its interaction with the number of past shocks experienced by the manager. Unsurprisingly, shocks negatively affect store productivity during the pandemic period. We do not find evidence that managers with more experience with past shocks are able to deal with current shocks better than their peers.

6.2.1 Robustness Checks

We now briefly address some of the alternative interpretations or concerns about the results discussed above.

Severity of shocks. We saw that better managers appear to attenuate the productivity effect of an COVID-19 shock once it happens. Here, we further explore the connections between manager quality and the severity of shocks, by seeing whether a shock of a given type has differential effects on stores operating practices (rather than productivity). We first rank shock severity by their influence on store operations, as follows (from least to most

Table 4: Impact of Manager Experience with Previous Shocks on Store Productivity During COVID-19 (Company A)

	Effect of Shocks		Effect of Experience	
log Productivity	(1)	(2)	(3)	(4)
Manager FE (EB-Adj.)	0.503***	0.503***	0.498***	0.479***
	(0.043)	(0.043)	(0.045)	(0.046)
Store FE (EB-Adj.)	0.649***	0.650***	0.649***	0.653***
	(0.028)	(0.028)	(0.029)	(0.029)
Unsafe conditions (due to COVID-19)	-0.434***	-0.567***		
	(0.129)	(0.144)		
Manager FE × Unsafe conditions		0.059***		
		(0.015)		
# Previous shocks			0.010*	0.019**
			(0.006)	(0.007)
Manager FE × # Previous shocks				-0.004***
				(0.001)
Experienced shock (current month)				-0.034***
				(0.011)
Experienced shock × # Previous shocks				-0.015
				(0.013)
N	3,108	3,108	3,108	3,108
Within R^2	0.2212	0.2370	0.1788	0.1878

Robust standard errors in parenthesis and ***, ** and * indicate coefficients are significantly different from zero at the 1%, 5% and 10% levels, respectively. All specifications include a fixed effect for each time period and connected set. Columns (1) and (2) also have dummies for all other observed shocks, with their interactions with manager fixed effects also included in column (2). Columns (3) and (4) also control for manager tenure.

severe):

1: Open late, mid-day closure, close early

2: Full day closure (one day only)

3: Temporarily reducing operating hours, switching to curbside

4: Long-term closure (3 or more days)

We regress this index on the manager and store FEs that experienced each shock, a set of indicators for each type of shock (second column of table B.3), as well as indicators for each

month and connected set in our sample. In an alternative specification, we also interact manager quality and the type of shock. Because of the ordinal nature of the index, we estimate these specifications as an ordered logit.

We do not find evidence that manager (or store) quality has a statistically distinguishable relationship to the severity of shocks, at least as measured by stores' operations.

Exposure to shocks and the likelihood of leaving. Finally, we ask whether managers who are exposed to more shocks before the pandemic are also more likely to leave the company. If so, the managers with high exposure to shocks in our post-COVID sample might pose a selection issue.

We estimate a linear probability model of whether a manager left Company A before February of 2020 on the number of shocks the manager was exposed to during the pre-pandemic period plus other observables. Once again, because the variable of interest is fixed over time, all of the explanatory variables are first averaged/summed to facilitate estimation of a cross-sectional model.

Unsurprisingly, managers with high salaries are less likely to leave the company, even after controlling for manager and store quality. In contrast, managers with long tenures are more likely to leave (departures include retirements). We find limited evidence that managers of any type with high exposure to shocks are more likely to leave in Company A. However, there is a small directional interaction; the likelihood of a manager leaving as the number of shocks they experience rises slightly faster for high-FE than low-FE managers. This difference is very small, however. Spanning the entire range of the manager FE distribution in Company A corresponds to a chance in the likelihood of manager departure for managers having experienced 10 previous shocks (the 90th percentile of the distribution) is 0.5 percentage points. Thus this selection effect is unlikely to drive any of the results discussed before.

Table 5: Probability of Leaving Company A in the Pre-pandemic Period

	(1)	(2)	(3)
Total # previous shocks	-0.002	-0.002	-0.005***
	(0.002)	(0.002)	(0.001)
Avg. log manager salary	-0.435***	-0.384**	-0.384**
	(0.164)	(0.180)	(0.185)
log manager tenure (last period)	0.110***	0.058**	0.062**
	(0.023)	(0.025)	(0.024)
Last store FE (EB-Adj.)		-0.073	-0.080
		(0.138)	(0.140)
Manager FE (EB-Adj.)		0.089	0.080
		(0.028)	(0.229)
Manager FE × # previous shocks			0.001***
			(0.0004)
N	545	443	443
Within R^2	0.0773	0.0326	0.0696

Robust standard errors in parenthesis and ***, ** and * indicate coefficients are significantly different from zero at the 1%, 5% and 10% levels, respectively. All specifications include a fixed effect for each connected set.

7 Conclusion

A growing literature has established that management influences business performance and productivity. This work has spanned settings that vary both by industry/sector and the development level of the broader economy the business operates within.

In our paper we show that managers have an impact on productivity that is distinct, relatively quick, and separate from company-level management practices. Quantitatively, we find that establishment-level characteristics can explain a larger share of the variance in productivity than managers can, but that the ranges of the distributions of the two effects are comparable. In that sense, we estimate that replacing manager at the bottom of the distribution by one at the top could increase the productivity of an establishment by at least 50% and up to more than doubling it, depending on the company and the relevant connected set.

There are some obvious caveats to our research. First we focus on two retail firms only, so it would be good to managers' influence in other firms and markets. Firm-level management

practices are fixed in this paper. While advantageous for isolating the role of managerial personnel in determining productivity, if managers were also able to change company-level management practices (such as the Bloom and Van Reenen (2007) practices), then they could have an even greater effect on productivity than we measure here. Third, our sample period is relatively short, so we cannot fully characterize any dynamic effects. Still another limitation is that we do not have access to granular data of what managers do on a moment-to-moment basis or a measure of their social and strategy skills. This limits our ability to understand the particular mechanisms that underlie our findings, given that such variables have been found to be correlated with productivity (Antonakis et al., 2014; Bandiera et al., 2020; Alan et al., 2021; Hansen et al., 2021; Impink et al., 2021).

Overall, we think we have shown that there is much fruitful work to be done in understanding how managers influence the productivity of the establishments they manage.

References

Abowd, J. M., F. Kramarz, and D. N. Margolis (1999). High Wage Workers and High Wage Firms. *Econometrica 67*(2), 251–333.

Adhvaryu, A., V. Bassi, A. Nyshadham, and J. A. Tamayo (2020). No line left behind: Assortative matching inside the firm. Technical report, National Bureau of Economic Research.

Adhvaryu, A., N. Kala, and A. Nyshadham (2019). Management and shocks to worker productivity. Technical report, National Bureau of Economic Research.

Aghion, P., N. Bloom, B. Lucking, R. Sadun, and J. Van Reenen (2021). Turbulence, firm decentralization, and growth in bad times. *American Economic Journal: Applied Economics 13*(1), 133–69.

Alan, S., G. Corekcioglu, and M. Sutter (2021). Improving workplace climate in large corporations: A clustered randomized intervention.

Andrews, D., A. Charlton, and A. Moore (2021). Covid-19, productivity and reallocation: Timely evidence from three oecd countries.

Andrews, M. J., L. Gill, T. Schank, and R. Upward (2008). High wage workers and low wage firms: negative assortative matching or limited mobility bias? *Journal of the Royal Statistical Society Series A 171*(3), 673–697.

Antonakis, J., G. d'Adda, R. Weber, and C. Zehnder (2014). Just words? just speeches? on the economic value of charismatic leadership. *NBER Rep 4*.

Baltrunaite, A., G. Bovini, and S. Mocetti (2021). Managerial talent and managerial practices: are they complements? *Bank of Italy Temi di Discussione (Working Paper) No 1335*.

Bandiera, O., L. Guiso, A. Prat, and R. Sadun (2015). Matching firms, managers, and incentives. *Journal of Labor Economics 33*(3), 623–681.

Bandiera, O., A. Prat, S. Hansen, and R. Sadun (2020). Ceo behavior and firm performance. *Journal of Political Economy 128*(4), 1325–1369.

Barry, J. W., M. Campello, J. Graham, and Y. Ma (2022). Corporate flexibility in a time of crisis. Technical report, National Bureau of Economic Research.

Becker, G. S. (1973). A theory of marriage: Part i. *Journal of Political Economy 81*(4), 813–846.

Bell, B. D., N. Bloom, and J. Blundell (2021). This time is not so different: Income dynamics during the covid-19 recession. Technical report, National Bureau of Economic Research.

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

Best, M. C., J. Hjort, and D. Szakonyi (2017). Individuals and organizations as sources of state effectiveness. Technical report, National Bureau of Economic Research.

Bloom, N., E. Brynjolfsson, L. Foster, R. Jarmin, M. Patnaik, I. Saporta-Eksten, and J. Van Reenen (2019). What drives differences in management practices? *American Economic Review 109*(5), 1648–83.

Bloom, N., P. Bunn, P. Mizen, P. Smietanka, and G. Thwaites (2020). The impact of covid-19 on productivity. Technical report, National Bureau of Economic Research.

Bloom, N., B. Eifert, A. Mahajan, D. McKenzie, and J. Roberts (2013). Does Management Matter? Evidence from India. *Quarterly Journal of Economics 128*(1), 1–51.

Bloom, N., R. S. Fletcher, and E. Yeh (2021). The impact of covid-19 on us firms. Technical report, National Bureau of Economic Research.

Bloom, N., C. Genakos, R. Martin, and R. Sadun (2010). Modern management: good for the environment or just hot air? *Economic Journal 120*(544), 551–572.

Bloom, N., R. Lemos, R. Sadun, D. Scur, and J. Van Reenen (2014). The new empirical economics of management. *Journal of the European Economic Association 12*(4), 835–876.

Bloom, N., R. Sadun, and J. Van Reenen (2012). The organization of firms across countries. *Quarterly Journal of Economics 127*(4), 1663–1705.

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

Bonhomme, S., K. Holzheu, T. Lamadon, E. Manresa, M. Mogstad, and B. Setzler (2020). How Much Should we Trust Estimates of Firm Effects and Worker Sorting? NBER Working Paper 27368.

Bonhomme, S., T. Lamadon, and E. Manresa (2019). Discretizing Unobserved Heterogeneity. University of Chicago, Becker Friedman Institute for Economics.

Borusyak, K. and X. Jaravel (2016). Revisiting event study designs. Working Paper.

Boudreau, L., R. Macchiavello, V. Minni, and M. Tanaka (2021). Union leaders: Experimental evidence from myanmar.

Boyd, G. A. and E. M. Curtis (2014). Evidence of an energy-management gap in us manufacturing: Spillovers from firm management practices to energy efficiency. *Journal of Environmental Economics and Management 68*(3), 463–479.

Bruhn, M., D. Karlan, and A. Schoar (2018). The impact of consulting services on small and medium enterprises: Evidence from a randomized trial in mexico. *Journal of Political Economy 126*.

Card, D., J. Heining, and P. Kline (2013). Workplace heterogeneity and the rise of west german wage inequality. *Quarterly journal of economics 128*(3), 967–1015.

Chandra, A., A. Finkelstein, A. Sacarny, and C. Syverson (2016). Health care exceptionalism? performance and allocation in the us health care sector. *American Economic Review 106*(8), 2110–44.

Correia, S. (2017). Linear models with high-dimensional fixed effects: An efficient and feasible estimator. Technical report.

Decker, R. A., J. C. Haltiwanger, R. S. Jarmin, and J. Miranda (2018). Changing business dynamism and productivity: Shocks vs responsiveness. Technical report, responsiveness. Technical report, National Bureau of Economic Research.

Eeckhout, J. and P. Kircher (2011). Identifying sorting – in theory. *Review of Economic Studies 78*(3), 872–906.

Eeckhout, J. and P. Kircher (2018). Assortative matching with large firms. *Econometrica 86*(1), 85–132.

Fee, C. E., C. J. Hadlock, and J. R. Pierce (2013). Managers with and without style: Evidence using exogenous variation. *Review of Financial Studies 26*(3), 567–601.

Fenizia, A. (2020). Managers and productivity in the public sector. Working Paper.

Fenizia, A. (2022). Managers and productivity in the public sector. *Econometrica 90*(3), 1063–1084.

Frederiksen, A., L. B. Kahn, and F. Lange (2020). Supervisors and performance management systems. *Journal of Political Economy 128*(6), 2123–2187.

Gaure, S. (2013). lfe: Linear group fixed effects. *The R Journal 5*(2), 104–117. User documentation of the 'lfe' package.

Gaure, S. (2014). Correlation bias correction in two-way fixed effects linear regression. *Stat 3*(1), 379–390. Description of the limited mobility bias correction method used in 'lfe'.

Giorcelli, M. (2019). The long-term effects of management and technology transfers. *American Economic Review 109*(1), 121–52.

Gosnell, G. K., J. List, and R. Metcalfe (2020). The impact of management practices on

employee productivity: A field experiment with airline captains. *Journal of Political Economy 128*(4), 1195 – 1233.

Hagedorn, M., T. H. Law, and I. Manovskii (2017). Identifying equilibrium models of labor market sorting. *Econometrica 85*(1), 29–65.

Hansen, S., T. Ramdas, R. Sadun, and J. Fuller (2021). The demand for executive skills. Technical report, National Bureau of Economic Research.

Hoffman, M. and S. Tadelis (2021). People management skills, employee attrition, and manager rewards: An empirical analysis. *Journal of Political Economy 129*(1), 243–285.

Huber, K., V. Lindenthal, and F. Waldinger (2021). Discrimination, managers, and firm performance: Evidence from aryanizations in nazi germany. *Journal of Political Economy 129*(9), 2455–2503.

Ichniowski, C. and K. Shaw (1999). The effects of human resource management systems on economic performance: An international comparison of us and japanese plants. *Management science 45*(5), 704–721.

Ichniowski, C. and K. Shaw (2003). Beyond incentive pay: Insiders' estimates of the value of complementary human resource management practices. *Journal of Economic Perspectives 17*(1), 155–180.

Ichniowski, C., K. Shaw, and G. Prennushi (1997). The effects of human resource management practices on productivity: A study of steel finishing lines. *American Economic Review 87*(2), 768–809.

Impink, S. M., A. Prat, and R. Sadun (2021). Communication within firms: Evidence from ceo turnovers. Technical report, National Bureau of Economic Research.

Janke, K., C. Propper, and R. Sadun (2019). The impact of ceos in the public sector: Evidence from the english nhs. Technical report, National Bureau of Economic Research.

Kahn, M. E. and N. Kok (2014, February). Big-Box Retailers and Urban Carbon Emissions: The Case of Wal-Mart. NBER Working Papers 19912, National Bureau of Economic Research, Inc.

Kaplan, S. N. and M. Sorensen (2017). Are ceos different? characteristics of top managers. Technical report, National Bureau of Economic Research.

Lazear, E. P. (2004). The peter principle: A theory of decline. *Journal of political economy 112*(S1), S141–S163.

Lazear, E. P., K. L. Shaw, and C. Stanton (2016). Making do with less: working harder during recessions. *Journal of Labor Economics 34*(S1), S333–S360.

Lazear, E. P., K. L. Shaw, and C. T. Stanton (2015). The value of bosses. *Journal of Labor Economics 33*(4), 823–861.

Leuven, E. and B. Sianesi (2003, April). PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. Statistical Software Components, Boston College Department of Economics.

Limodio, N. (2021). Bureaucrat allocation in the public sector: evidence from the world bank.

Lopes de Melo, R. (2018). Firm wage differentials and labor market sorting: Reconciling theory and evidence. *Journal of Political Economy 126*(1), 313–346.

Martin, R., M. Muûls, L. B. De Preux, and U. J. Wagner (2012). Anatomy of a paradox: Management practices, organizational structure and energy efficiency. *Journal of Environmental Economics and Management 63*(2), 208–223.

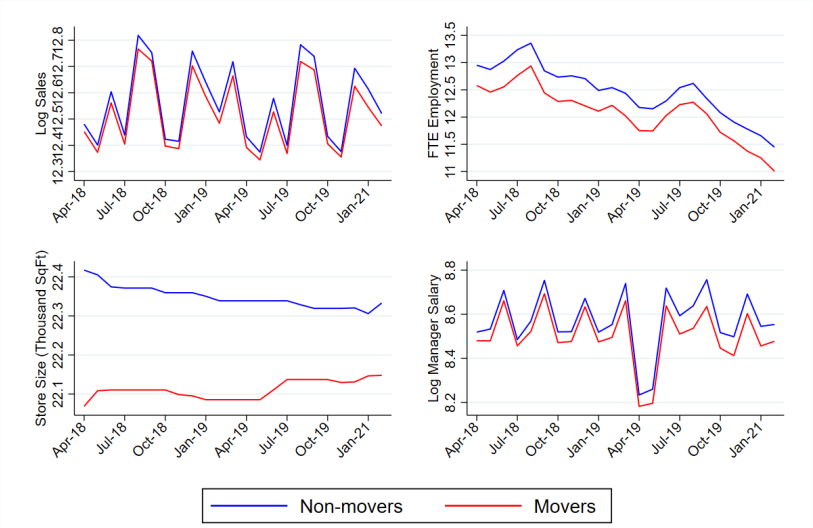
Ruggles, S., S. Flood, S. Foster, R. Goeken, J. Pacas, M. Schouweiler, and M. Sobek (2021). Ipums usa: Version 11.0 [dataset]. Technical report, Minneapolis, MN: IPUMS.

Syverson, C. (2011). What determines productivity? *Journal of Economic Literature 49*(2), 326–65.

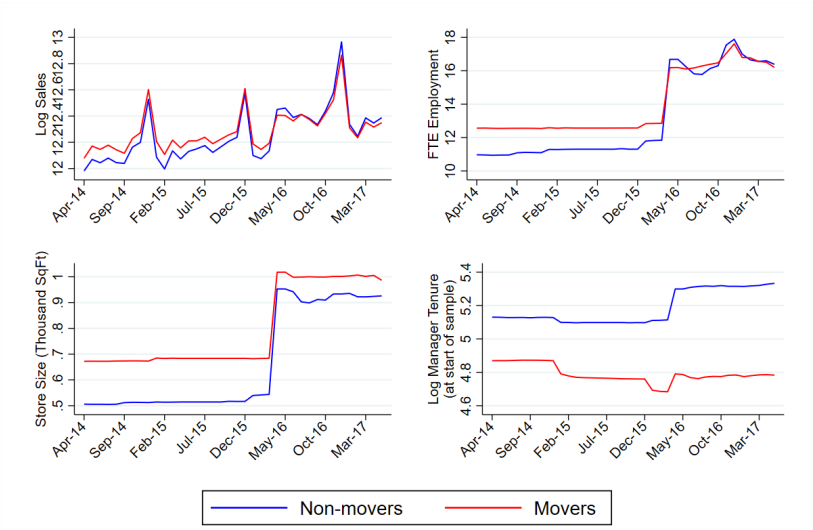
A Figures

Figure A.1: Pre-move Trends: Difference Between Movers and Non-movers

(a) Company A



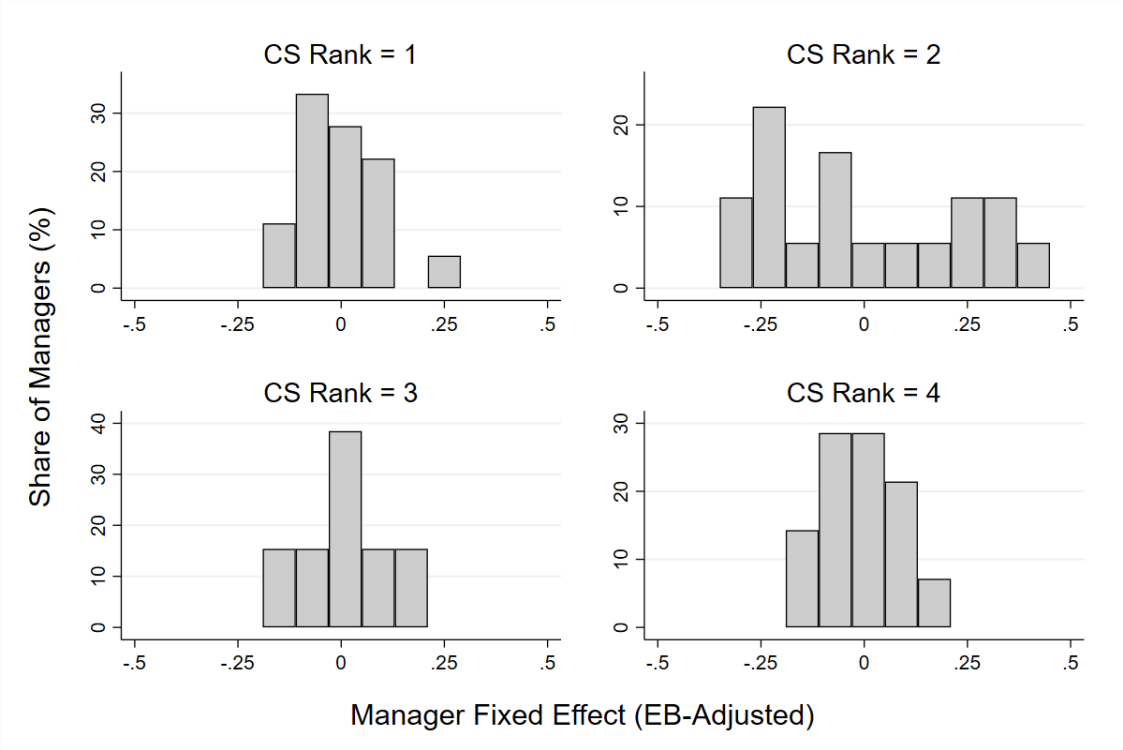
(b) Company B



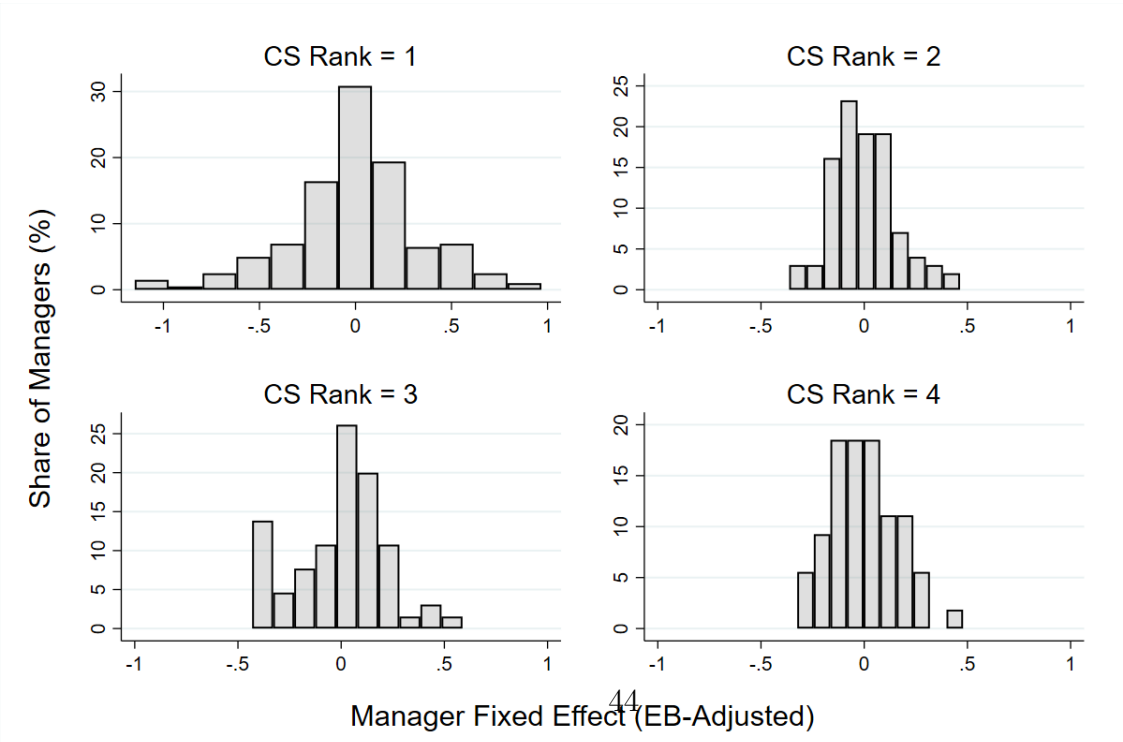
Note: starting in April of 2016, data for many of the smaller establishments of Company B is no longer available, which accounts for the shifts in average outcomes (for both movers and non-movers) at that date.

Figure A.2: Distribution of EB-Adjusted Manager Fixed Effects

(a) Company A



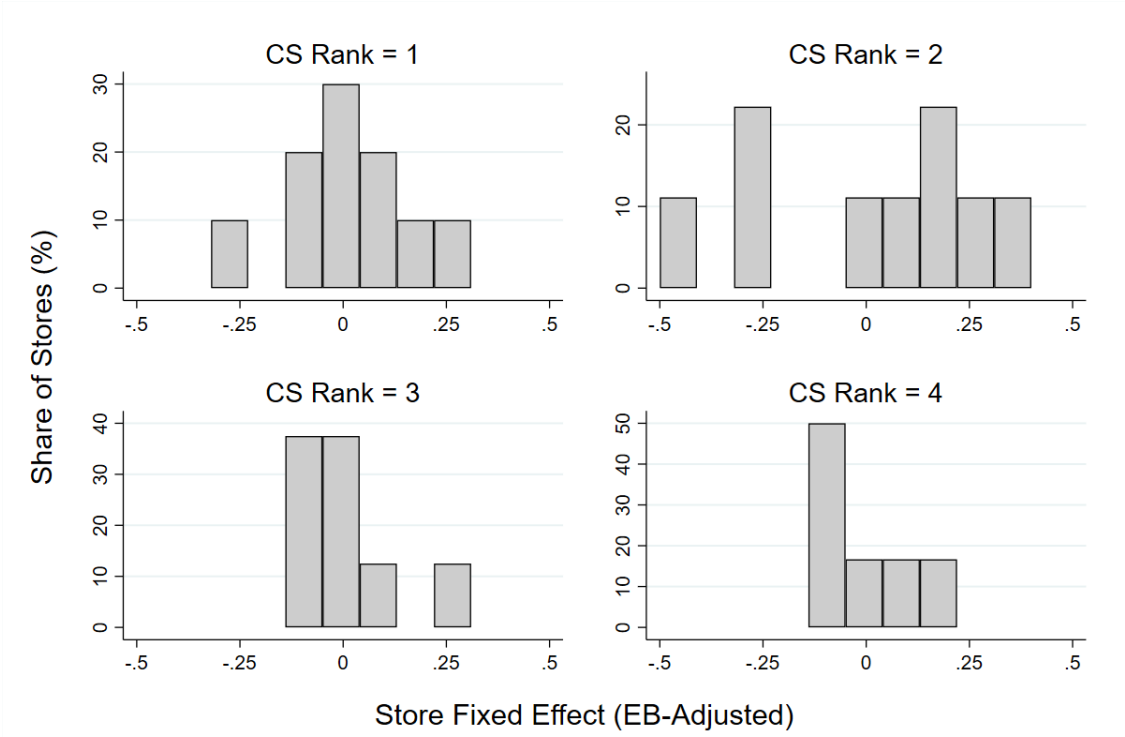
(b) Company B



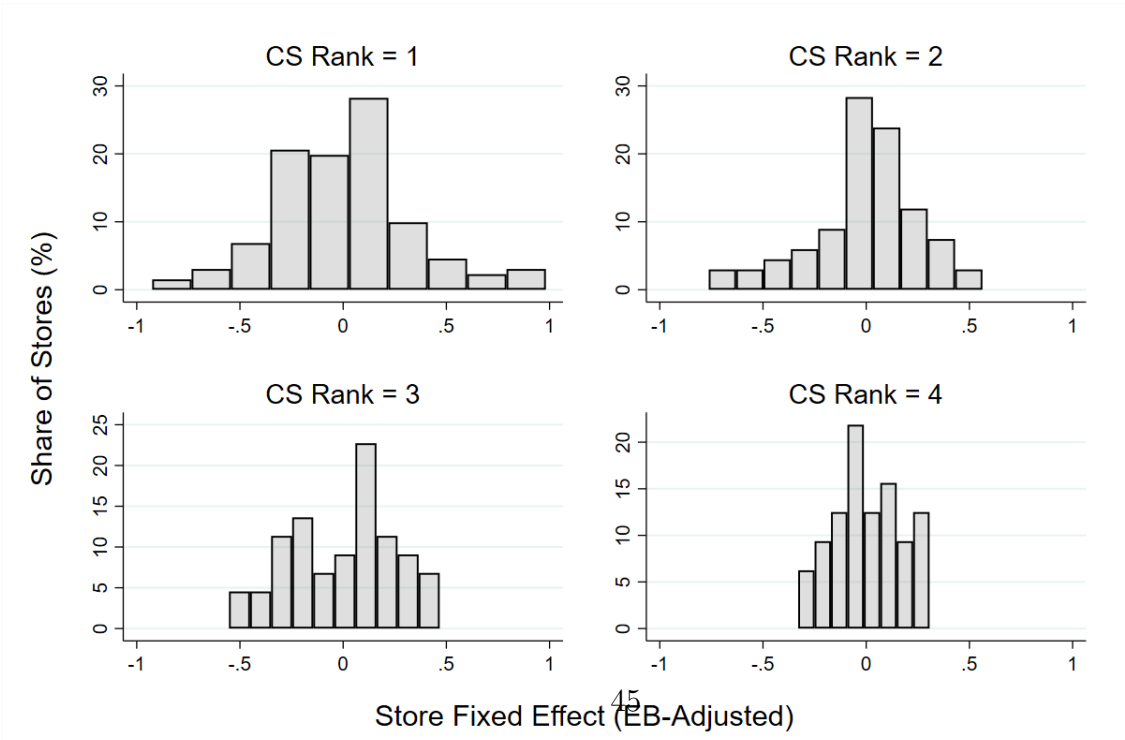
Note: FE's de-meaned by connected set for plotting.

Figure A.3: Distribution of EB-Adjusted Store Fixed Effects

(a) Company A



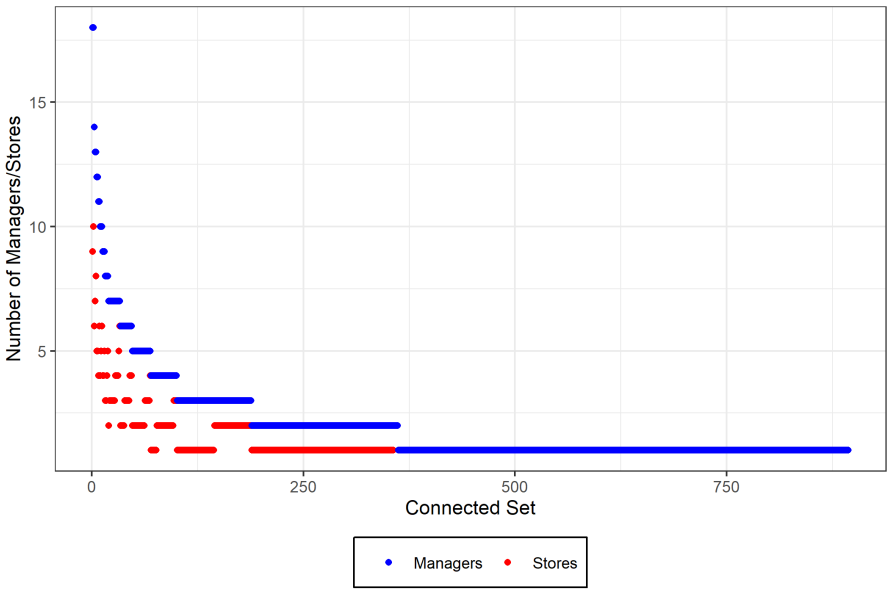
(b) Company B



Note: FE's de-meaned by connected set for plotting.

Figure A.4: Number of Stores and Managers by Connected Set

(a) Company A



(b) Company B

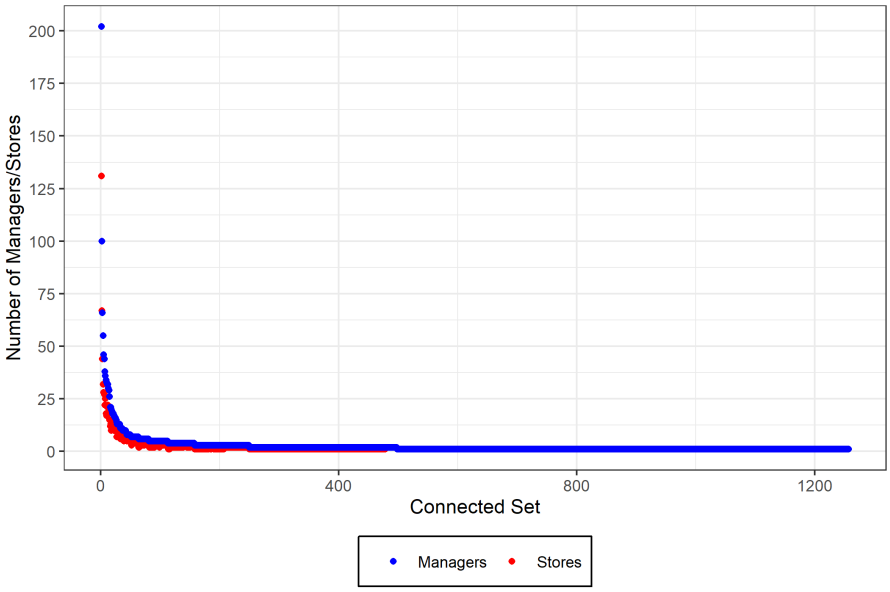
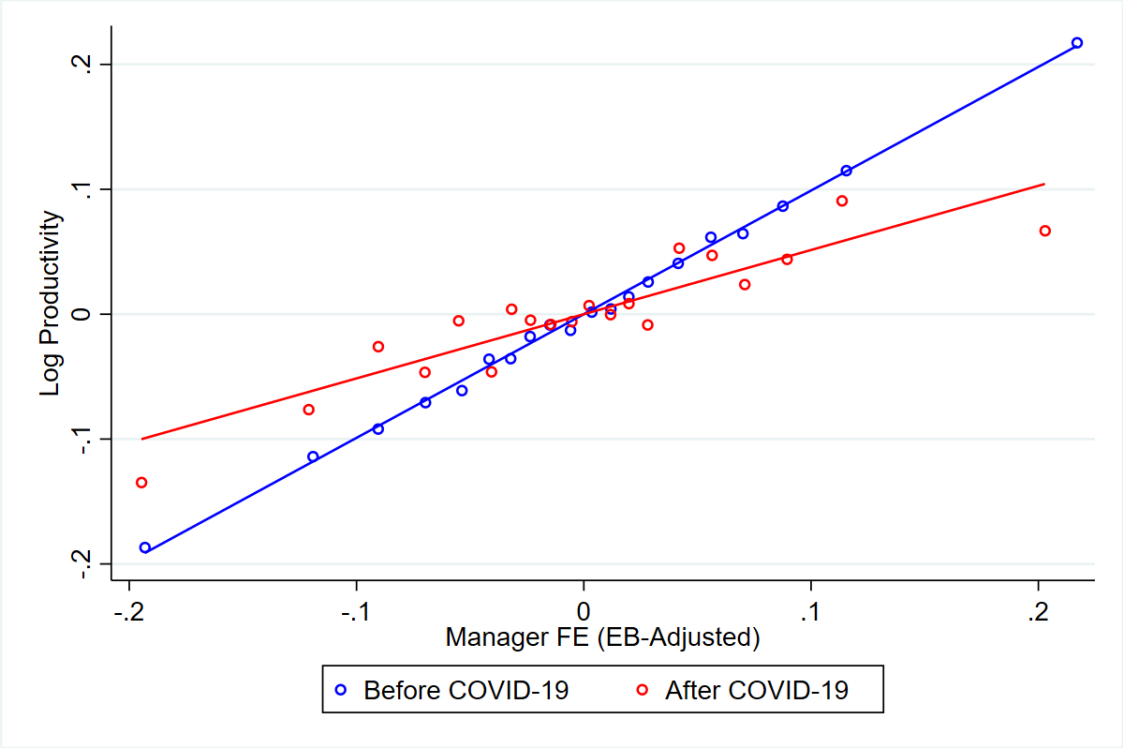


Figure A.5: Manager FE and Productivity: Before and During COVID-19



B Tables

Table B.1: Difference between mover and non-mover stores/managers

	Company A	**Company B**
% mover stores	53%	63%
Sales (movers vs. non-movers)	4% lower	0.7% lower
FTE (movers vs. non-movers)	3% lower	6% larger
Store size (movers vs. non-movers)	1% lower	24% larger
Manager salary (movers vs. non-movers)	6% lower	
Manager tenure (movers vs. non-movers)		26% lower

Note: Manager salary is not available for Company B, so we use the manager's tenure at the beginning of the sample instead.

Table B.2: Stores Managers Moved To vs Moved From – Company A

Variable	Store	Average	95% CI	
Log Sales	Moved to	12.65	[12.58	12.73]
	Moved from	12.49	[12.41	12.57]
	Difference	0.16	[0.08	0.24]
Log FTE Employment	Moved to	2.56	[2.5	2.63]
	Moved from	2.51	[2.44	2.59]
	Difference	0.05	[-0.01	0.11]
Log Manager Salary	Moved to	8.57	[8.54	8.6]
	Moved from	8.45	[8.40	8.49]
	Difference	0.13	[0.08	0.17]
Log Population Density	Moved to	8.04	[7.81	8.28]
	Moved from	7.89	[7.61	8.17]
	Difference	0.15	[-0.09	0.4]

Note: population density reflects the number of people per square mile (in 2010) in the county where the store is located. Data was obtained from the American Community Survey through IPUMS USA.

Table B.3: List of Observed Shocks and Potential Impacts on Stores

Impact on Stores	**Shocks/Reason for Impact**
Open Late	*Cleaning*
Mid-day Closure (close & reopen in same day)	*Hours Adjustment*
Close Early	Local Authority Evacuation
Full Day Closure	*Local Mandate*
Long-term Closure ($>$ 3 days)	Major Facility Issue
Temporarily Reducing Operating Hrs	Natural Disaster/Weather Condition
Switching to Curbside	Power Outage
Returning to Normal Operating Hrs	Protests
Returning From Curbside	*Staffing Issues*
	Unsafe Conditions
	Other

Note: this table lists all possible impacts on stores and all possible reasons (they are not necessarily matched). Shocks that can also be caused by COVID-19 interventions are shown in *italics*.

Table B.4: Impact of COVID-19 and Manager Experience on Store Sales

	(1)	(2)	(3)	(4)
Manager FE (EB-Adj.)	0.448***	0.447***	0.446***	0.434***
	(0.055)	(0.055)	(0.056)	(0.057)
Store FE (EB-Adj.)	0.997***	0.999***	0.997***	1.003***
	(0.036)	(0.036)	(0.036)	(0.036)
Unsafe Conditions (COVID-19)	-0.383***	-0.502***		
	(0.131)	(0.154)		
Hours Reduction (COVID-19)	-0.195***	-0.011		
	(0.079)	(0.074)		
Manager FE × Unsafe Conditions		0.053***		
		(0.016)		
Manager FE × Hours Reduction		2.209***		
		(0.450)		
Number of Shocks (pre-COVID)			0.017***	0.023***
			(0.005)	(0.006)
Manager FE × Number of Shocks				-0.003**
				(0.001)
N	3,108	3,108	3,108	3,108
Within R^2	0.3230	0.3336	0.2966	0.2977

Robust standard errors in parenthesis and ***, ** and * indicate coefficients are significantly different from zero at the 1%, 5% and 10% levels, respectively. All specifications include a fixed effect for each connected set.

C Estimating Store and Manager Productivity: Challenges and Alternatives

C.1 Alternative Measures of Productivity

There are a number of ways to measure productivity, and in this section we compare our baseline measure with other commonly used alternatives. Our baseline measure sets productivity equal to sales per employee. For the purposes of this section, we call this L-productivity:

$$\text{L-productivity}_{s,t} = \frac{\text{sales}_{s,t}}{\text{FTE employment}_{s,t}}.$$

K-productivity. The first alternative measure is constructed in a similar way, but replacing employment by each store's size. Since the store size is a measure of capital, we refer to this measure as K-productivity:

$$\text{K-productivity}_{s,t} = \frac{\text{sales}_{s,t}}{\text{Floor Area}_{s,t}}.$$

TFPR. The third and final measure we analyze is total factor productivity. Because we do not observe quantities sold, we must construct this measure using revenues and costs, thus TFPR.

We model a store's production function as

$$Y_{s,t} = A_{s,t} L_{s,t}^{\alpha_L} K_{s,t}^{\alpha_K} M_{s,t}^{\alpha_M},$$

where $Y_{s,t}$ is the store's output, $L_{s,t}$ is employment, $K_{s,t}$ measures the store's capital and $M_{s,t}$ measures materials and other inputs to production (overhead costs). Cost minimization implies

$$\frac{\alpha_j}{\sum_j \alpha_j} = \textit{cost share}_j, \quad j \in \{L, K, M\}$$

where $\textit{cost share}_j$ indicates the share of total costs taken by input j (e.g., the labor cost share is $\textit{cost share}_L = \frac{w_t L_{s,t}}{w_t L_{s,t} + r_t K_{s,t} + p_t^M M_{s,t}}$). While the labor and overhead costs are observable, the cost of capital goods for each store must be constructed. We approximate $K_{s,t}$ as the store floor area and r_t as the median rental rate at the county where the store is located. To

simplify computation, we assume constant returns to scale, $\sum_j \alpha_j = 1$.

Given the cost shares α_j, we compute a store's revenue-TFP as[12]

$$TFPR_{s,t} = \frac{(PY)_{s,t}}{L_{s,t}^{\alpha_L} K_{s,t}^{\alpha_K} (p^M M)_{s,t}^{\alpha_M}},$$

where $(PY)_{s,t}$ are sales and $(p^M M)_{s,t}$ are overhead costs of store s in month t.

C.1.1 Relationship Between the Three Measures

Figure C.1 plots the three productivity measures discussed above in a bin-scatter plot (after de-meaning to account for differences in scale), showing that all three of them are very closely related. The same can be seen in table C.1, which reports a very high correlation between each pairwise combination of the three productivity measures. All of this suggests that our baseline, L-productivity, captures the same features of the store that alternative productivity measures would.

Table C.1: Correlation Between Alternative Measures of Productivity

	L-productivity	K-productivity	TFPR
L-productivity	1		
K-productivity	0.60*	1	
TFPR	0.76*	0.69*	1

Note: * indicates that correlations are statistically significant at the 5% level.

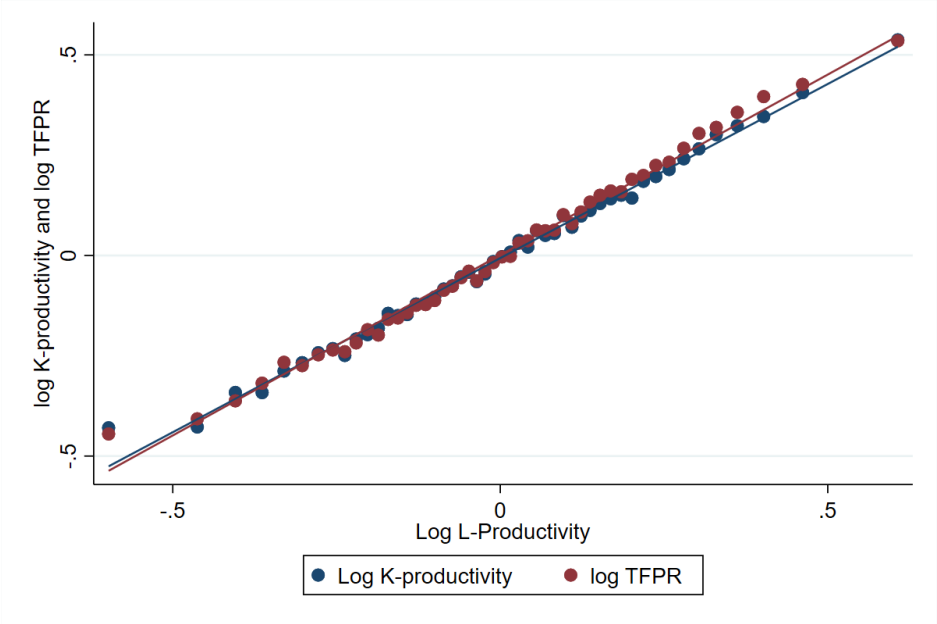
C.2 Allowing for Manager and Store Interactions

Our baseline decomposition of store productivity, equation (1), assumes a log-linear specification that does not explicitly allow for a sorting effect (complementarity) between the stores and managers. To test whether this would change our results, we run the alternative model

$$\log(prod_{s,t}) = \mu_s + \mu_{m(s,t)} + \mu_{s,m(s,t)} + \mu_t + \varepsilon_{s,t}$$

[12] Although we assume a Cobb-Douglas production function, the equation $A_{s,t} = \frac{Y_{s,t}}{L_{s,t}^{\alpha_L} K_{s,t}^{\alpha_K} M_{s,t}^{\alpha_M}}$ holds more generally as a first-order approximation of any production function. In those cases, one can compute α_j as the output elasticity of input j.

Figure C.1: Relationship Between Alternative Measures of Productivity



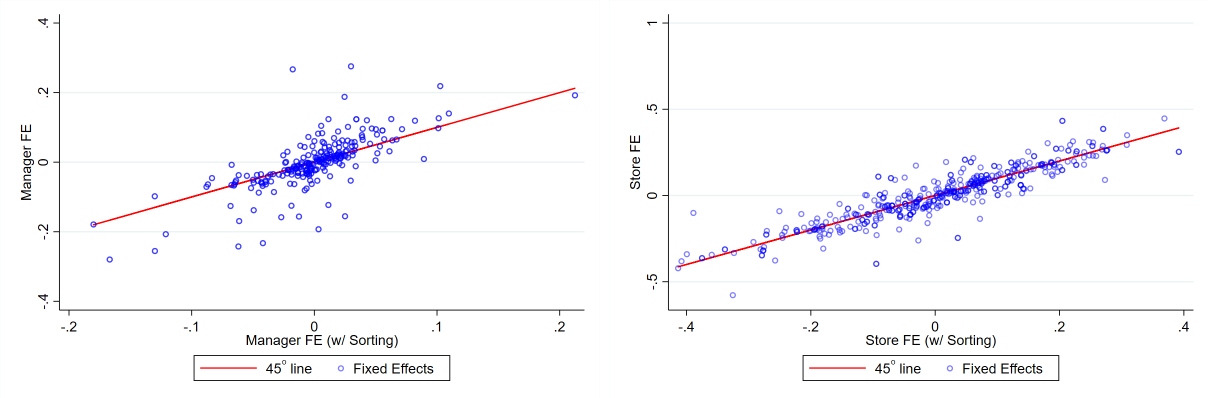
Note: data is binned in to 50 quantiles for plotting; each dot represents the average productivity in its respective bin. All three productivity measures have been de-meaned.

where $\mu_{s,m(s,t)}$ is a store-manager pair fixed effect, which captures any sorting effect between managers/stores. Figure C.2 compares the estimated μ_s and $\mu_{m(s,t)}$ obtained from equation above with the baseline effects estimated in the main text (equation 1). This figure shows that there is little change in the manager and store effects when a term is added to capture any type of complementarity between them.

C.3 Correcting for Measurement Error: Grouped Fixed Effects

A second approach that addresses the inherent measurement error in fixed effect estimation is to abandon the estimation of individual effects and instead focus on groups. In this section, we adopt the two-step procedure proposed by Bonhomme et al. (2019). In the first step, we cluster managers and stores into groups with similar units. Then, assuming that all managers and stores in the same group have a common fixed effect, we estimate the influence of each group on an establishment's productivity. The main advantage of this method is that it drastically reduces the number of estimated parameters, raising statistical

Figure C.2: Comparing Estimates With and Without Sorting



(a) Manager Fixed Effects

(b) Store Fixed Effects

precision. However, it also inevitably simplifies manager and store heterogeneity and requires us to take a stand on the way units are clustered.

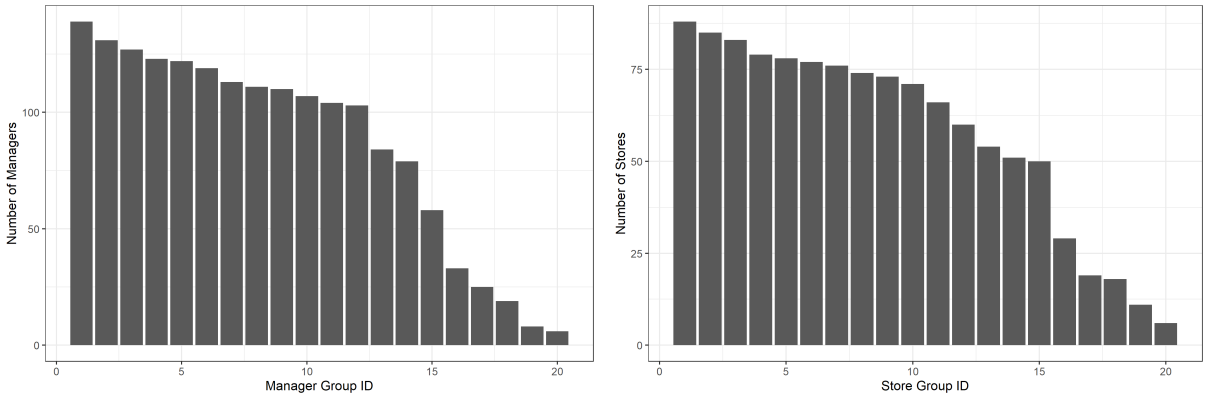
Step 1 We classify both managers and stores into 20 groups each, using a k-means algorithm. The number of groups is arbitrary, and chosen so that each group contains a relatively large number of managers/stores. Varying the number of groups does not meaningfully affect our results.

The key issue that arises in this step is the choice of variables used to cluster the units in our data. Ideally, all of the relevant heterogeneity across managers and stores would be captured by a few observable variables, so that there is little loss of information in the clustering process. We cluster managers in Company A based on their average log-wage across the sample. The intuition is straightforward: all else equal, more productive managers should be paid higher wages. Because this is not available for Company B, we use their average tenure instead.

We cluster stores by their average log FTE employment during the sample, based on the notion that more productive stores should have more employees working in them. While store employment could be guided by company policy, there is scope for managers to at least temporarily change the number of employees if he or she sees fit. This might imply that FTE employment also captures part of the manager's effect, however. To address this concern, we alternatively cluster stores by floor area (a measure of capital in retail) and find very similar results.

Step 2 Once groups are defined, we repeat the estimation of equation 1, with the exception that the effects on the right hand side no longer identify individual managers or stores, but rather the group to which they belong. Figure C.3 shows the number of managers and stores in each group. Figure C.4 below plots the distribution of the estimated grouped fixed effects. In the clustered data, there is a single connected set in the network of stores, so all fixed effects are normalized by the same value.

Figure C.3: Number of Managers and Stores in Each Group



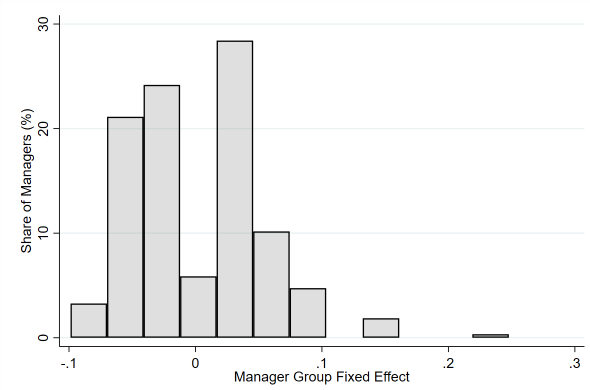
(a) Managers (b) Stores

Comparing figure C.4 with figures A.2a and A.3a, we see that the distribution of grouped fixed effects has most of its mass within a smaller range (this is not obviously the case *a priori*, as the number of connected sets also changes). The shape of the distributions in each of those three figures, however, is relatively similar, even if there are sometimes gaps in the distributions of fixed effects.

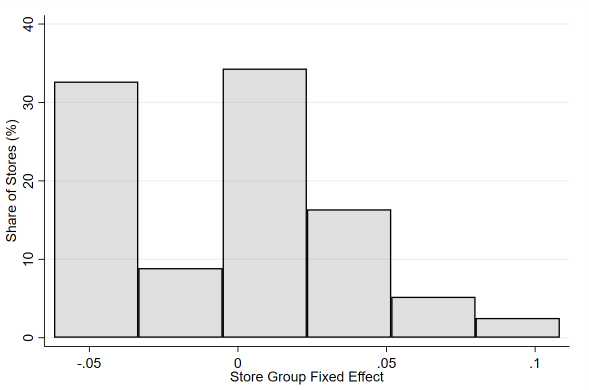
Finally, figure C.5 plots the relationship between the log of productivity (sales per full-time equivalent employee) and each of the fixed effects we estimate: unadjusted (i.e., the coefficients directly estimated from (1)), EB-adjusted, and grouped. As seen in the figure, the unadjusted and EB-adjusted FEs are very similar, with a noticeable difference only for store quality in Company B. The grouped fixed effects, however, have a significantly flatter relationship with productivity, even if still positive in all cases. This result could indicate that we loose much of the heterogeneity across managers and stores by grouping them along a single dimension.

Figure C.4: Distribution of Grouped Fixed Effects

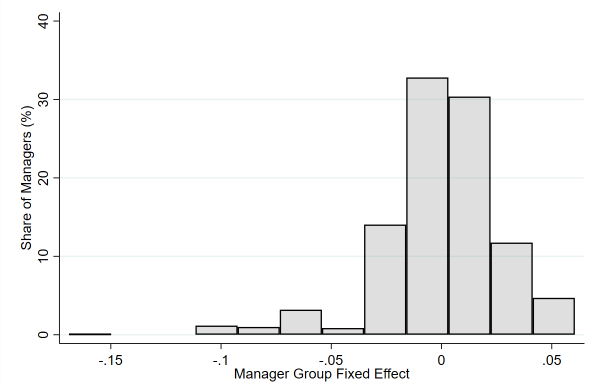
(a) Manager Fixed Effect (Company A)



(b) Store Fixed Effect (Company A)



(c) Manager Fixed Effect (Company B)



(d) Store Fixed Effect (Company B)

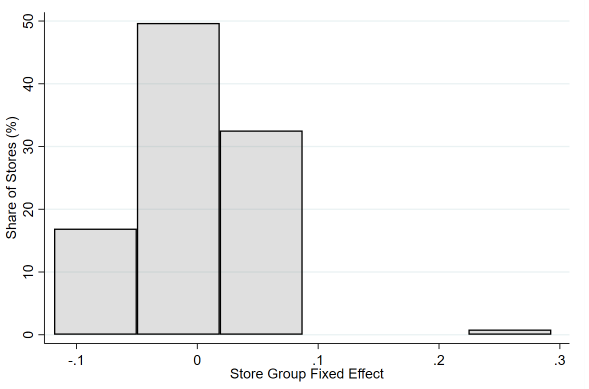
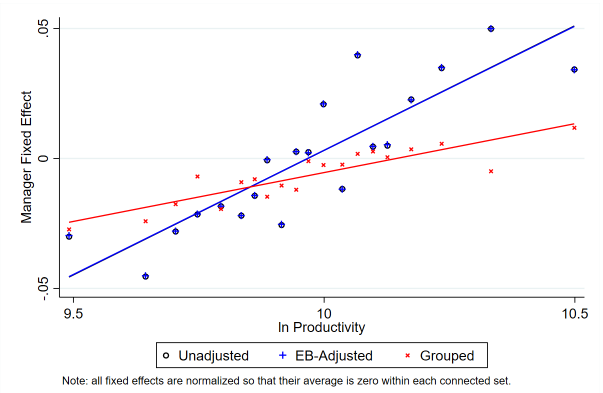
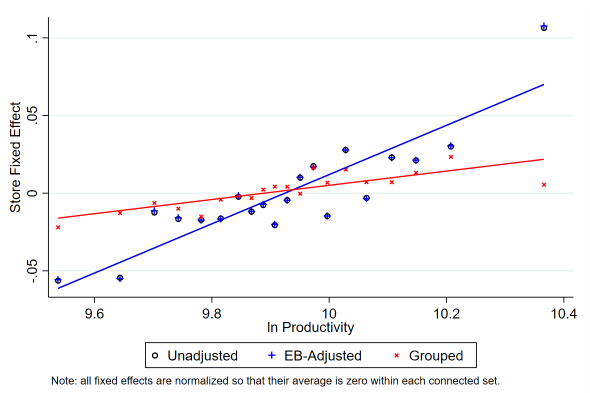


Figure C.5: Unadjusted, EB-Adjusted, and Grouped Fixed Effects

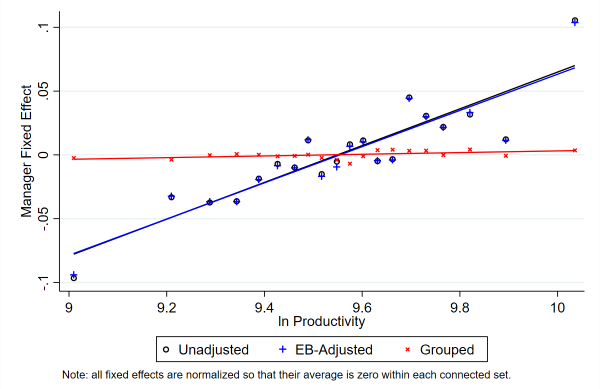
(a) Manager Fixed Effect (Company A)



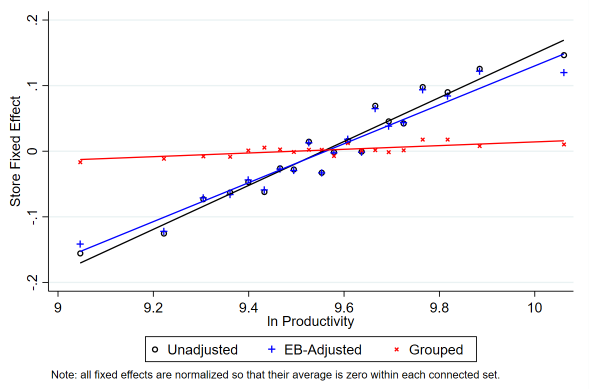
(b) Store Fixed Effect (Company A)



(c) Manager Fixed Effect (Company B)



(d) Store Fixed Effect (Company B)



Note: all fixed effects are normalized so that their average is zero within each connected set.

D Gains from Assortative Matching

As argued in section 3.2.1, there does not appear to be positive assortative matching between managers and stores. We discuss several reasons for why this might be the case, but the fact remains that, under certain assumptions, this is not the optimal allocation of managers to stores. In this section, we compute the potential revenue gain per year should companies assortatively match their managers and stores.

Given model (1), we can do this by simply ranking managers and stores based on their estimated fixed effects, match them based on their ranking, and compute the resulting productivity gain as

$$\Delta \log(prod_{s,t}) = \mu^*_{m(s,t)} - \mu_{m(s,t)},$$

where $\mu_{m(s,t)}$ is the FE of the current manager and $\mu^*_{m(s,t)}$ is the FE of the manager in the optimal match. Given the productivity increase, we can easily calculate the sales revenue increase as well. Because not all managers are present at all times nor can they be compared across connected sets, we only reallocate managers inside the same connected set and period. For Company A, we also differentiate between pre- and post-Covid-19 periods, as the pandemic could have changed the role of managers quite substantially (in which case the pre-Covid estimate is a better reflection of potential gains).

Our estimates are shown in table D.1, and suggest that Company A could have increased their revenue by 2.22 percent per month in the pre-Covid periods (or the equivalent of about USD 2.23 million/month) simply by optimally allocating their managers *within connected sets* – which means that the unconstrained reallocation could be even more beneficial. During the pandemic, the same exercise results in a slightly lower, but still relevant, estimate: 2.10 percent or USD 1.85 million per month. In Company B the gains are even larger: 6.07 percent gains per month, or about GBP 13.2 million! In each case, our results indicated that are substantial gains from reallocating the best managers to the best stores withing the company.

Table D.1: Sales Revenue Gain per Year with Assortative Matching

	Company A				Company B	
	Pre Covid-19		Post Covid-19			
	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI
Percentage	2.22	[1.82, 2.63]	2.10	[1.51, 2.68]	6.07	[5.55, 6.60]
Value (millions)	2.23	[1.90, 2.77]	1.85	[1.33, 2.38]	13.2	[12.0, 14.3]

Note: values are in USD for Company A, and GBP for Company B. Confidence intervals are obtained by bootstrapping sample 100 times.

E Event Study: Pre-Trends and Robustness

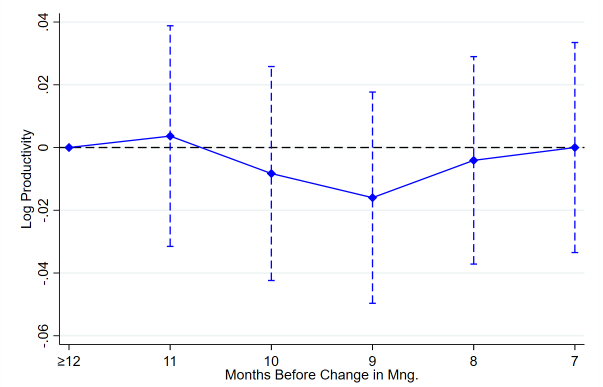
To test for the existence of pre-trends (i.e., before 6 months prior to a change in manager) in out study, figure E.1 runs model (3') in the periods used to match treated to control stores. As expected, none of the coefficients are statistically different from zero at the usual confidence levels, suggesting that stores were in a common trend before the change in manager became apparent.

E.1 Managers who Stay, and Managers who Leave

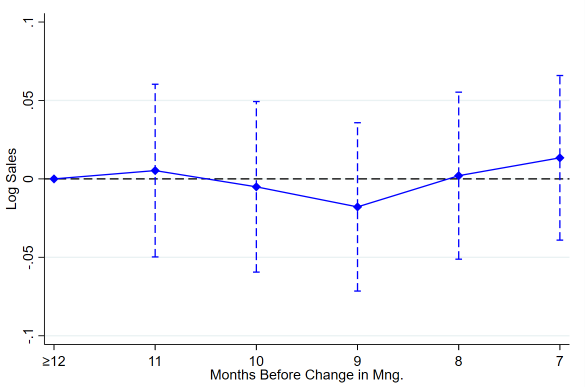
As mentioned in section 4, there are many reasons why a manager could leave a store. Among those, managers who leave a store but remain within the company (that is, move to another store) might be very different from those that leave the Company Altogether (are fired, retire, move to a different job). Figure E.3 tests this hypothesis by conducting each of the event studies on the subset of managers who stay within the company.

Most of the effects we have seen in figures 3 and 4 are also present in for managers who switch stores within the company, but are generally greatly attenuated. This is particularly true for Company A (figure E.3), where are all estimated coefficients are considerably closer to zero than in the full study – note in particular that the drop in wages is almost half as big as the drop in figure 3 (firms would plausibly not want to punish managers that transferring internally). The results for Company B (figure E.4) show a less pronounced decline in productivity and sales before the change in manager, and a quicker recovery afterwards, although sales seem to recover slower. We also see a higher increase in the ratio of full-time to part-time workers after the managerial change, although coefficients are not statistically

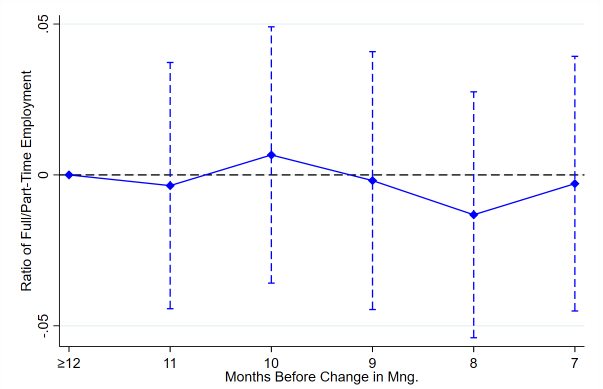
Figure E.1: Event Study: Outcomes During Matching Periods
Company A



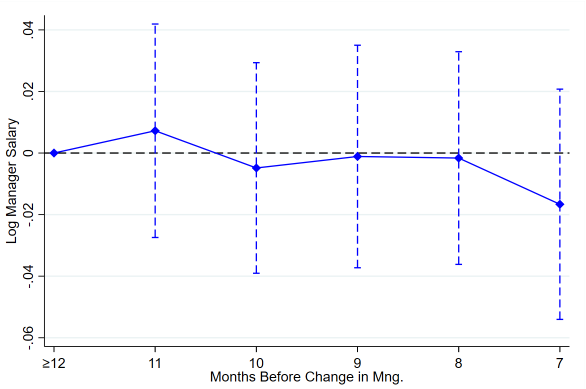
(a) Productivity



(b) Sales



(c) Full/Part-Time Emp. Ratio

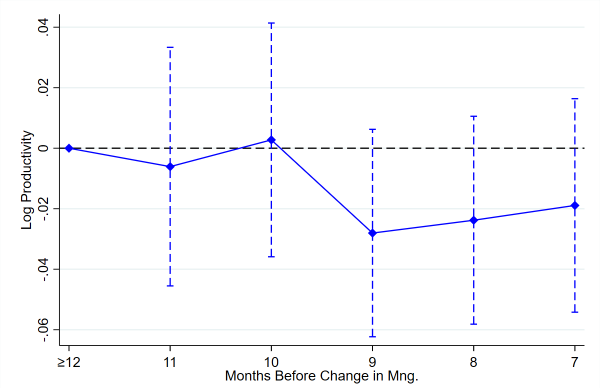


(d) Manager Salary

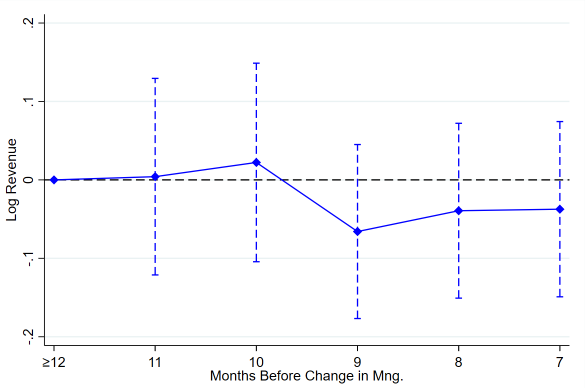
Note: dots show the value of each coefficient β_k, while whiskers indicate the 95% confidence interval (computed using robust standard errors).

significant (we should note that Company B has much fewer managers that move internally, relative to the full set of "movers").

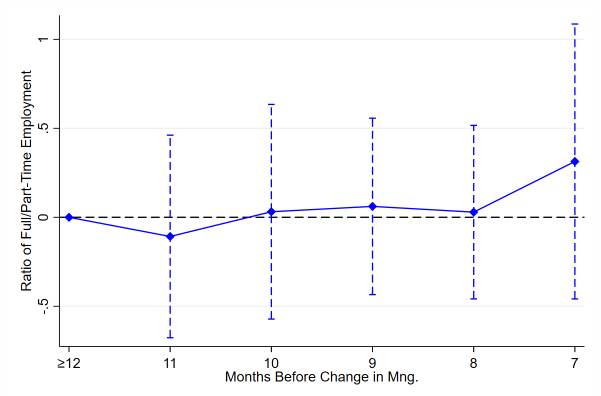
Figure E.2: Event Study: Outcomes During Matching Periods
Company B



(a) Productivity



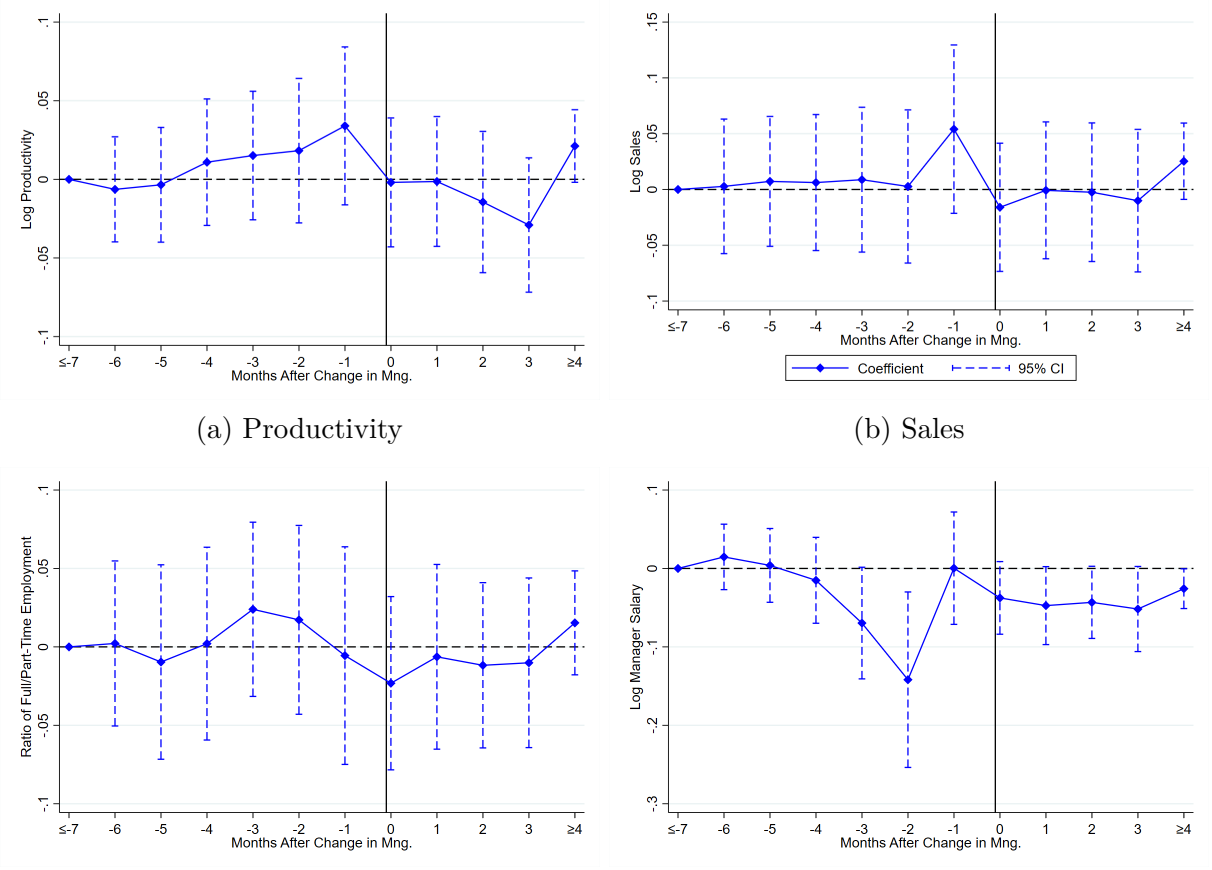
(b) Sales



(c) Full/Part-Time Emp. Ratio

Note: dots show the value of each coefficient β_k, while whiskers indicate the 95% confidence interval (computed using robust standard errors).

Figure E.3: Event Study: Managers Who Stay in Company A



(a) Productivity

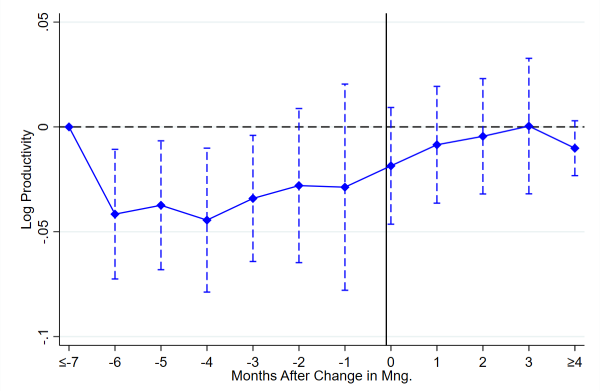
(b) Sales

(c) Full/Part-Time Emp. Ratio

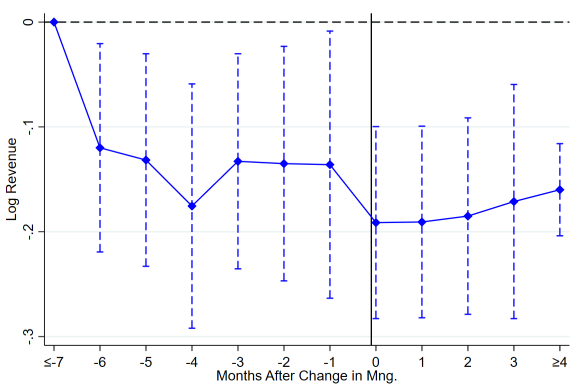
(d) Manager Salary

Note: dots show the value of each coefficient β_k, while whiskers indicate the 95% confidence interval (computed using robust standard errors).

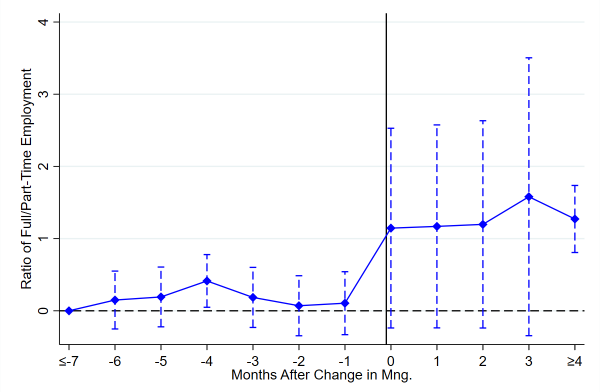
Figure E.4: Event Study: Managers Who Stay in Company B



(a) Productivity



(b) Sales



(c) Full/Part-Time Emp. Ratio

Note: dots show the value of each coefficient β_k, while whiskers indicate the 95% confidence interval (computed using robust standard errors).

F Manager Productivity and Energy Usage

In this section, we explore the impact of managers on their store's energy consumption. We do that by regressing energy usage on the manager's productivity (EB-adjusted fixed effect), as well as other store-level controls, and exploring the signs of the correlations we find. We focus on electricity consumption in stores in company B (we do not have energy use data for company A), as electricity use is almost universal across stores. In addition to that, about 40% of stores in our sample also use gas as a source of energy. Those stores tend to be concentrated in specific regions of the UK (perhaps reflecting local differences in the energy grid), are larger (in terms of store area), but also have lower revenues than similar stores that use solely electricity. Importantly, we find no correlation between the manager EB-adjusted fixed effect and the likelihood that a store uses gas. In any case, because electricity and gas might be substitutes, our regression includes a store-level dummy variable that equals 1 if a store records positive gas usage in at least one period of our sample.

We present the results in table F.1. Note that all specifications include a time fixed effect to capture seasonal movements, as well as store-level dummy variable that indicates each period in which a store adopts energy saving technologies, such as chillers and LED lights (energy tech FE).[13] The first two columns also include location (city) and connected set fixed effects, which are subsumed by a store fixed effect in column (3). As seen in the table, we find that a store's electricity consumption increases with it's size, both in terms of floor area and revenue, and decreases as the price of electricity increases or other sources of energy (gas) are available.[14] All of those findings are very intuitive, although we also highlight that there are clear endogeneity concerns in some coefficients, particularly on electricity prices and gas usage.

We also find that more productive managers tend to consume less electricity, regardless of movements in energy prices, availability of alternative energy sources, or the size of the store they manage. This result is also robust to the inclusion of the average number of hours that the store is open each week, which has a large positive impact on energy consumption,

[13] We crudely estimate the effect of energy savings technologies on energy usage by regressing log electricity consumption on the energy tech dummy, while controlling for electricity prices, average temperature, as well as time and store fixed effects. We find that adopting such technologies brings electricity consumption down by about 1.6 percent each month. This is consistent with results from Kahn and Kok (2014), who argue that newer buildings tend to be more energy-efficient than older ones.

[14] Energy prices reflect the average price of energy consumption, calculated as energy cost/energy consumption (both per month).

and the average monthly temperature in the store's vicinity, where we see a negative impact, indicating higher energy consumption in colder months.[15].

The last column replaces the location and connected set fixed effects by a store fixed effect, which absorbs a substantial amount of variation from the data (note that the within-R^2 drops from 0.8 to 0.05 in the third column). The store fixed effect also subsumes all of the variation in store area, gas usage, and the number of hours open each week. It almost completely removes the impact of the store's revenue on energy consumption as well. The effect of managerial productivity, however, is still negative and has a comparable size with the estimates in columns (1) and (2), even though it is no longer statistically significant.

[15] Average monthly temperature is obtained from measurements by the weather station that is closest to each store. We obtain data from each station from the Met Office website. Using the geographical coordinates for both stations and stores, we calculate the geodesic distance between each station-store pair and match each store to its nearest station. The average distance between a store and its nearest station is about 0.4 miles, with all stores located within 2 mile distance from its nearest station

Table F.1: Impact of Manager Productivity and Store Characteristics on Electricity Usage

	(1)	(2)	(3)
Manager FE (EB-Adj.)	-0.065*	-0.053*	-0.028
	(0.036)	(0.034)	(0.029)
Log Revenue	0.446***	0.327***	-0.001
	(0.036)	(0.034)	(0.017)
Log Store Area	0.641***	0.652***	-
	(0.025)	(0.024)	-
Log Electricity Price	-0.797***	-0.853***	-0.665***
	(0.145)	(0.134)	(0.136)
Uses Gas	-0.160***	-0.127***	-
	(0.026)	(0.026)	-
Log Hours Open (week)		0.991***	-
		(0.111)	-
Log Avg. Temperature		-0.123**	-0.144**
		(0.047)	(0.035)
Energy Tech FE	Y	Y	Y
Time FE	Y	Y	Y
Location FE	Y	Y	N
Connected Set FE	Y	Y	N
Store FE	N	N	Y
N	26,744	26,594	28,004
Within R^2	0.7931	0.8074	0.0512

Standard errors are clustered at the connected set level and shown in parenthesis. ***, **, and * indicate coefficients are significantly different from zero at the 1%, 5% and 10% levels, respectively.