Managers and Productivity in Retail

Robert D. Metcalfe
Columbia University & NBER

Alexandre B. Sollaci
International Monetary Fund

Chad Syverson
*University of Chicago & NBER**

Abstract

In many sectors, research has established that management explains a significant portion of productivity differences across organizations. A remaining question, however, is whether it is managers themselves or firm-wide management practices that matter. We shed light on this question by analyzing store-level data from two multi-billion-dollar retail companies. In this setting, managers move between stores, but management practices are set by firm policy and store attributes are largely fixed, allowing us to isolate managers' personal roles in determining store performance. We have three high-level findings: (i) managers affect and explain 25-35% of the variance of store-level productivity; (ii) negative assortative matching between managers and stores, which may reflect both firms' decisions and a selection-driven bias that we characterize as novel and argue might apply in other settings using AKM-style movers designs; and (iii) moving from a below median to an above median manager raises store-level productivity by around 10% after several months.

****Acknowledgments**: We thank participants at the Empirical Management Conference, NBER Summer Institute Personnel Economics, European Economic Association Meetings, Chicago, CMU, Duke, IMF, LSE, MIT, Princeton, Stanford, as well as Tim Armstrong, Vittorio Bassi, Nick Bloom, Stephane Bonhomme, Leah Boustan, Thomas Chaney, Oren Danieli, Michela Giorcelli, Bob Hahn, Nathan Hendren, Jonas Hjort, Mitch Hoffman, Matt Kahn, Rem Koning, Alex Mas, Simon Quach, Raffaella Sadun, and John Van Reenen for excellent 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 authors 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 explains a notable portion of 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.[1] Management's effects on performance can operate through two channels that are not mutually exclusive: (1) the manager herself and (2) firm-level management practices (Syverson, 2011). This is an important distinction. In principle, a firm's mandated management practices can be part of the intangible capital of an organization and, as such, are transferable over time and space. An individual manager's performance effects, on the other hand, are inherently tied to the manager. This distinction has implications regarding firms' optimal strategies for improving performance and productivity, the distribution of the rents yielded by superior management, and how government interventions to improve management might increase national productivity levels.

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' personalities, preferences, and innate abilities. Some work has looked at managers' personal effects (such as innate ability and personality styles), although this has focused mostly on corporate CEOs.[2] One potential confounding feature of the corporate CEO effects literature is that it relies on CEOs moving across firms for identification, opening the possibility that firm-level unobservables may shape the selection of certain managers or their influence on performance. Moreover, it has been found that even within the same firm, there is substantial variation in productivity that is not explained by the CEO (Chew, Clark and Bresnahan, 1989).

In this paper, we shed additional light on the influence of individual managers on business performance. Our empirical setting grants us dimensions of sharpness and resolution unavailable in prior research. For instance, unlike the CEO effects literature, we can control for firm-level influences on performance, such as governance, corporate cul-

[1]Examples of relevant and important studies include, among others, papers by Ichniowski, Shaw and Prennushi (1997); Ichniowski and Shaw (1999); Bertrand and Schoar (2003); Ichniowski and Shaw (2003); Bloom and Van Reenen (2007); Bloom, Sadun and Van Reenen (2012); Bloom et al. (2013, 2014); Kaplan and Sorensen (2021); Bruhn, Karlan and Schoar (2018); Bloom et al. (2019); Giorcelli (2019); Janke, Propper and Sadun (2019); Bandiera et al. (2020); Gosnell, List and Metcalfe (2020); Sandvik et al. (2020); Bianchi and Giorcelli (2022).

[2]See Barberis et al. (1996); Bertrand and Schoar (2003); Fee, Hadlock and Pierce (2013); Bandiera et al. (2015); Bender et al. (2018); Bandiera et al. (2020); Huber, Lindenthal and Waldinger (2021); Acemoglu, He and le Maire (2022); Rubens (2023); Baltrunaite, Bovini and Mocetti (2023).

ture, pricing, branding, product assortment, supply chain structure, and firm policies regarding management practices. This allows us to focus more on the roles of particular people (managers) rather than practices. Despite company-wide policies, managers in our setting still have multiple margins to affect store-level productivity, like motivating employees; choosing which employees to hire, fire, and promote; selecting the optimal full-time/part-time ratio; enforcing and managing employee schedules and store technology; providing a pleasant physical environment and culture; reducing stock-outs; and implementing company-level changes in price and product assortment.[3] Of course, this list makes clear we cannot eliminate all variations in practices across managers. However, our within-firm design lets us go considerably farther than most prior studies, and several of these remaining practice differences may in any case be less codifiable and separable from the manager than corporate policies.

Additionally, we are able to empirically separate the performance effects of the manager from productivity variation driven by attributes inherent to the establishment they manage. In other words, we can measure if an establishment (retail store, in our setting) has high productivity because the manager is talented, or instead because there are elements of that store's operating environment (a favorable location, for instance) that lead it to perform well. This further empirically isolates managers' productivity effects. It also lets us capture how managers might influence stores of different types, and whether and to what extent the firm's allocation of managers across stores harnesses any complementarities between store and manager types. In this way, we can zoom out beyond measuring manager heterogeneity and explore more broadly the optimality of manager allocation, an atypical addition in this literature.

Our data includes information on the performance and management of thousands of stores owned by two multi-billion dollar retail companies, one in the US and one in the UK.[4] Our data are rich enough not only to measure how much individual managers matter, but also to allow us to characterize some potential 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, which reduces a host of measurement biases and validity issues.

[3]Previous research suggests that management discretion in how some of these are implemented influences productivity (Hoffman, Kahn and Li, 2018).

[4]The managers in our data are quite representative of the 8.9 million managers in the United States (US Census Bureau) and the 3.5 million managers in the United Kingdom (Office of National Statistics) with respect to salary, education, and the scope of management responsibilities.

Estimating individual manager and store effects on productivity poses methodological challenges. First, not all manager and store fixed effects can be separately estimated (Abowd, Kramarz and Margolis, 1999). Identification of both effects relies on managers moving between stores. A benefit of working with two large companies is that we have a number of large connected sets through which we can measure many manager fixed effects. Across the two companies, we can estimate 3,791 manager fixed effects and 2,935 store fixed effects. 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 moments of the fixed effects distribution. To address these biases, we implement the correction procedures described in Andrews et al. (2008) and Gaure (2014). We also apply an empirical Bayes (EB) 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).

These adjustments allow us to learn about the joint distribution of manager and store fixed effects. Its shape and range in particular tell us much about the extent to which managers influence stores' productivity. This is another element that distinguishes our approach from previous studies, which have 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). Moreover, another advantage of our inside-the-firm measurement is that we can abstract away from across-firm heterogeneity, which Sorkin (2018) and Di Addario et al. (2022) have shown to be large.

We report multiple findings. First, individual 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. We find in our two-way fixed effects movers design that, averaging across the four largest connected sets in the data, managers are responsible for about 30% of store-level variation in productivity in the US-based company and almost 70% of the variation in productivity in the UK-based company.[5]

These results are robust to how productivity is measured (whether labor productivity, capital productivity, or TFP). Also, our model works well out of sample. Manager and

[5]Note that because the measured covariance between the fixed effects of managers and stores is negative, the sum of their variance can be larger than the variance in productivity. On average, the variance of the fixed effect of managers is about half of the variance of the fixed effect of stores.

store fixed effects fit on the first two-thirds of our data period do very well at predicting store-level productivity in the last one-third.

We also find that manager and store fixed effects tend to be negatively correlated in our data. That is, better managers tend to work at worse stores and vice versa. While contrary to our expectations, we posit three potential explanations for this negative assortative matching (NAM). One, there might be a benefit of positive assortative matching due to complementarities, but firms are not aware of it and do not take advantage of it in their manager allocation decisions. Two, alternatively, NAM may reflect the firms' optimal allocations in the absence of substantial complementarities between managers and stores (Becker, 1973).

A third possibility is that NAM reflects a measurement issue that may apply broadly in many empirical settings beyond our own. We show that productivity-based selection on manager-store pairings may negatively bias the estimated covariance, masking true complementarities. Specifically, if the likelihood of a manager-store pair breaking up (resulting in the manager moving) decreases in the manager and store types, this induces a negative correlation between manager and store fixed effects among movers. Because manager and store types cannot be separately identified for non-movers, the correlation between types may be estimated as negative, even if the true correlation among all manager-store pairings is zero or positive. This effect is economically structural; it is not driven by statistical measurement issues, so simply increasing the size of the data would not address it. Beyond our sample, we expect this mechanism may be at work in many of the empirical settings in the two-way fixed effects literature using AKM-type (Abowd, Kramarz and Margolis, 1999) estimators, as the conditions that cause it are likely to hold in many different settings. This is a separate, methodological contribution of our work here. We discuss this more below.

Next, we estimate manager effects in a difference-in-differences specification where we match above- and below-median mover managers on pre-move store productivity levels and trends. We find shifting from a below-median to an above-median manager raises productivity by 7 to 16%, depending on the company. Conversely, moving from an above-median to a below-median manager reduces productivity by 10 to 14%. While we do not have random assignment of managers to stores, we show that issues like match-specific effects, permanent manager effects, or manager improvements over time (which can lead to endogeneity issues—see Card, Heining and Kline, 2013) do not drive any of our results. This may reflect in part what we know from interviews with firm executives. Managerial moves are seldom driven only by managers' preferences; staffing rules and company needs are important determinants, making the moves less likely to be correlated with productivity shocks.

All of the results above are largely consistent across the two companies, despite considerable differences in their products, store scales, and geographic footprints.

Our study touches on several literatures. There are several papers related to our result on managers' influence on outcomes. Lazear, Shaw and Stanton (2015) shows that supervisors of computer-based test graders matter, where 10th-to-90th percentile supervisor difference has an equivalent effect on output as adding one worker to a team of 9. Janke, Propper and Sadun (2019) shows hospital managers, on the other hand, do not measurably influence health-related productivity metrics. Fenizia (2022) uses a movers design and shows that within the Italian social security administration, manager fixed effects explain 9% of the total variation in claims productivity at the office level. Giardili, Ramdas and Williams (2022) examine car manufacturing plants and managers and show that managers explain about 7% of production differences. Minni (forthcoming) shows that better managers shape firm output per worker in her setting primarily by better matching workers to jobs within the firm, highlighting how managerial discretion over within-firm factor allocations affects output.

Our work differs from these important papers along several margins. First, we examine managers in a retail setting, where managers of customer-facing stores have may have considerable scope to influence realized productivity. Second, our revenue-based output metric can capture those multiple channels of potential managerial influence. Third, we have more and larger connected sets of managers and stores across both companies, increasing statistical precision for measuring manager and store types and the matching function between them. These differences might explain in part why our estimate of managers' effects on productivity is much larger than these other papers.

Our research also speaks to existing work on the channels through which managers can improve the productivity of high- and low-performing workers (Mollick, 2012; Lazear, Shaw and Stanton, 2015; Best, Hjort and Szakonyi, 2017; Adhvaryu, Kala and Nyshadham, 2019; Frederiksen, Kahn and Lange, 2020; Hoffman and Tadelis, 2021; Limodio, 2021; Cai and Wang, 2022; Friebel, Heinz and Zubanov, 2022; Patault and Lenoir, 2024; Minni, forthcoming). It also has ties to the matching literature in labor economics (Eeckhout and Kircher, 2018) and more specifically to the negative assortative matching results in Card et al. (2018), and Adhvaryu et al. (2020) who examine how managers choose workers to increase productivity. It is also related to choosing social movement (union) leaders on preferences for wage changes (Boudreau et al., 2021).

We extend the two-way fixed effects sorting models of labor across firms (Abowd, Kramarz and Margolis, 1999; Card, Heining and Kline, 2013; Lopes de Melo, 2018) by examining ties among productivity, managers, and stores rather than wages, workers, and firms. Productivity is an outcome of inherent interest. Plus, 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, Law and Manovskii, 2017). Such measurement and identification issues do not exist when revenues and costs are observed within each unit that the manager manages. On a broader methodological front, as we note above, to our knowledge we are the first in the two-way fixed effects literature to note the potential for selection-based bias in mover identification strategies (i.e., when the likelihood of a manager-store breaking up decreases in a function of the manager and store types). While we elucidate this selection-based bias in our setting, it will apply to all two-way fixed effect estimation settings with mover designs (worker-company, person-neighborhood, etc.) where the stability of pairs is an increasing function of both types. This issue is separate from the other biases in the two-way fixed effect literature, such as limited mobility bias (Andrews et al., 2008).

In the remainder of this paper, 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. Section 5 concludes.

2 Data

Our dataset comprises monthly operations data at the store level for two large retail companies. Our data for Company A, an American company, spans April 2018 to February 2020. Our data for Company B, a British company, runs from April 2014 to May 2017.

In both cases, the data includes hundreds 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 or region), some 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.[6]

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) 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.

2.1 Movers and Non-movers

AKM-type decompositions like those at the heart of our analysis only identify fixed effects for "movers." In our case, the fixed effects are the persistent productivity component of managers and stores, and the necessary condition for identification is that managers change stores or stores change managers during the sample. As a result, it is useful to briefly discuss what forces might cause manager-store pairs to change in the two firms in our sample.

Our data do not report why managers leave their stores. We therefore cannot observe directly whether manager departures are related to store characteristics that might otherwise induce contemporaneous changes in productivity (we explore this further in section 4). However, we see 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 (Appendix G contains a more detailed discussion on pre-trends). Still further, as discussed above, staffing rules and company needs separate from productivity shocks are important determinants of managers move. The combination of these factors ameliorate our concerns about movers experiencing systematically different productivity

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

shocks than non-movers.[7]

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.

3.1 Measuring Productivity

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 from 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 productivity measure 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. We consider these possibilities in Appendix C.1, which compares sales per employee with other potential measures of productivity, such as sales per floor area (which proxies for physical capital), and revenue-based TFP (which controls for labor, capital, and intermediate inputs). We find all measures are highly correlated and appear to capture the same underlying store characteristics. Productivity is also highly correlated with both store

[7] As an interesting side note, one clear influence on move probability, at least in Company B where we observe it, is the manager's gender (which we infer based on their title: Miss, Mr, Mrs). Female managers are about 7% less likely to move to another store than male managers, a gap that cannot be explained by manager tenure or store characteristics like scale or format (Table B.2). The gender gap in move probability shrinks and loses statistical significance among managers whose first store is located in a city with at least 10 stores, however. A possible explanation for this is that women are less likely to change residences for family reasons. If promotions are associated with moving to different stores and female managers were more apt to forgo store changes that require residence changes, it would explain why the gap shows up only in smaller metro areas. This would resonate with Costa and Kahn (2000) or Dauth et al. (2022), where large cities facilitate better matching.

sales and profits. Within Company A, for which we observe labor costs, the correlation between productivity and sales is about 0.99, and the correlation between productivity and operating profits (defined as sales − labor costs) is approximately 0.75. Almost all of the variation in sales or profits is also captured by productivity; a linear regression of either sales or profits on productivity yields a R^2 of over 0.99. Given those relationships, we focus on labor productivity in the remainder of the paper, though our results should apply to other performance measures as well.

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

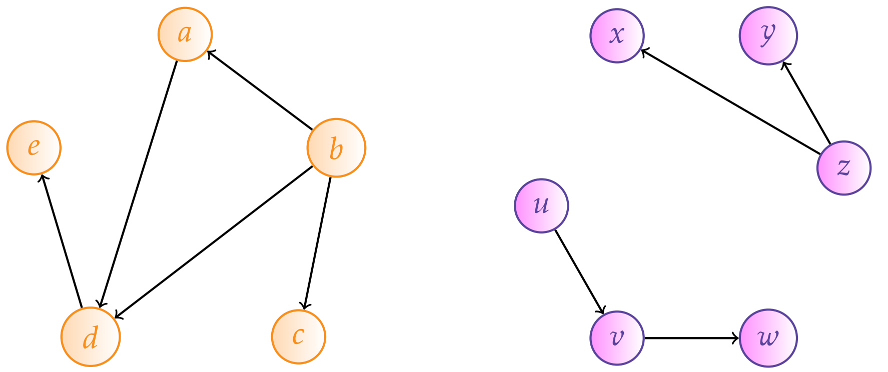
$$\log(prod_{s,t}) = \mu_s + \mu_{m(s,t)} + \mu_t + \varepsilon_{s,t} \tag{1}$$

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 manager fixed effect captures the manager average effect on the productivity of the stores where they worked. Throughout the paper, we interchangeably refer to the manager's fixed effect as their productivity, ability, or quality. The time fixed effect absorbs any company-wide productivity shocks.

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 the same manager for the duration of our sample, there is no way to separately identify each fixed effect, as the manager and store dummies are perfectly collinear. The approach therefore relies on managers that move across stores for identification.

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 the 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 a single *connected set*, meaning that all stores in the network have, directly or indirectly, shared at least one

Figure 1: Examples of Store Networks



(a) One connected set

(b) Two connected sets

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 within 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 United States. 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.[8]

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, because they are measured relative to different, unobserved baselines. This means that we can only compare the magnitudes of fixed effects inside the same connected set (see also Abowd, Kramarz and Margolis, 1999; Fenizia, 2022).

Our data contains about 890 connected sets in Company A and just over 1,200 connected sets for Company B. 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. The sets are much

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

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 estimates. 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) and, in some robustness checks, the `reghdfe` package in Stata (Correia, 2017).[9]

The third major challenge in our framework involves estimation error, as our fixed effects are typically estimated with relatively few observations. This may cause issues in the analysis of the fixed effects: For example, even if measurement error has mean zero and is independent from the true parameters, 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 in multiple ways below.

Lastly, the log-linear model (1) could be incorrectly specified, in which case the estimates of the fixed effects μ_s and μ_m would be correlated with the residual $\varepsilon_{s,t}$. One source of miss-specification are local shocks (not accounted for the time fixed effect) that may affect store-level productivity and bias the estimated manager and store fixed effects. We test this possibility in Appendix C.2 by running an alternative specification of equation (1) that replaces μ_t by a location-specific effect (city for company A and NUTS2 region for company B) and comparing the results with our baseline estimates, finding very similar results.

As argued in Card, Heining and Kline (2013), a correlation between the manager/store effects and the residual could also occur if there are match-specific effects in productivity. We also test for this in Appendix C.2 by running by including a match-specific term $\mu_{m,s}$ into equation (1). The manager and store fixed effects estimated in this alternative model show no notable differences compared to the baseline results. A third test looks at permanent effects of managers: if there were permanent effects from managers on stores, the inclusion of the fixed effects of previous managers should improve the model's fit.

[9]For consistency throughout the estimation process, we estimate fixed effects using the fixed-point iteration algorithm, which can be implemented in both packages.

Again we find very small differences relative to our baseline. Finally, we test whether managers' qualities are improving over time, in which case the inclusion of the interaction between the manager fixed effect and their tenure should again improve the model fit. We find very small differences here too. All of these issues are discussed in more detail in Appendix C.2.

One final source of endogeneity we consider is whether the likelihood that a manager leaves/joins a store is correlated with the shock $\varepsilon_{s,t}$. This would require that good and bad managers react differently to the same shock. If all managers are equally likely to leave a store after a bad shock, there is no correlation between $\varepsilon_{s,t}$ and $\mu_m(s,t)$ (we discuss this issue in a bit more detail in Appendix D). While we cannot directly address this concern in this paper, we again note that our conversations with firm management reveal that the decision to move from a store is seldom made by the manager alone and instead is often driven by staffing rules and company needs that are uncorrelated with store-level productivity shocks.

3.2.1 Variance Decomposition

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

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

The left-hand side of this equation adjusts the variance of log productivity 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.

Limited Mobility Bias. One issue we face when estimating manager and store fixed effects is measurement (or estimation) error, as μ_m and μ_s are estimated in a finite sample. In most cases, the presence of classical measurement error (uncorrelated with the true parameter value) is not cause for concern. However, in models with two or more fixed effects this may lead to a negative bias in the covariance between the estimated manager and store productivities. Intuitively, because the store and manager fixed effects must sum up to an observed quantity (i.e., log productivity, ignoring the time effect for illustration), when μ_s is slightly overestimated (underestimated), then μ_m will have to be slightly underestimated (overestimated) to preserve the overall total. As a result, the covariance between the estimated fixed effects of manager and store pairs will be downward biased. This issue is sometimes referred to as 'limited mobility bias,' as Abowd et al. (2004) suggested that this bias is larger when there are fewer movers in data.

Andrews et al. (2008) fully characterize this bias in a fairly general model and derive a formula for computing it using objects that can be estimated from the data. As a result, once the bias is computed, it can be used to adjust the original covariances to arrive at a consistent estimate for the relationship between manager and store quality. In practice, this process can involve computing inverses of impractically large matrices, so we employ the method developed by Gaure (2014), again through the `lfe` package in R.

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	1.82	1.02	-1.13
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	1.63	1.51	-1.21
2	1,979	0.72	0.22	-0.13
3	1,179	1.03	0.77	-0.60
4	1,008	0.39	0.30	-0.03

CS Rank is the rank of the connected set by size, and *Obs* indicates the number of observations (store×month) in each connected set. 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 and all estimates adjust for limited mobility bias.

After adjusting for limited mobility bias, 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 largest connected sets (CS) in each company. There are two key takeaways from this table. In all cases in the table, the store FE variance is larger than the manager FE variance. Thus stable store-level factors (including local demographics, competition from nearby retailers, etc.) explain a greater share of the observed productivity variation than do managers. The average share of the variance explained by manager fixed effects is still sizable, however. Averaging the shares across connected sets in the table, the variance of the manager fixed effects accounts for roughly 30% of the overall variance in productivity in company A, and almost 70% of the overall variance in company B. This suggests an important role for managers to influence outcomes.

The second takeaway is that the covariance between manager and store effects is almost always negative. This can also be seen in the “raw" data by plotting the number of matches by manager and store quality. As shown in Table 2, the most common matches are between a top manager and bottom store or a bottom manager and top store. 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. Indeed, most matching models would predict positive assortative matching between heterogeneous complementary inputs if productivity maximization were the objective (e.g., Becker, 1973).[10]

We estimate that Company A and Company B could increase their sales revenue by 0.65 and 1.83 percent each month (approximately USD 700,000 and GBP 4 million per month), respectively, by reallocating their managers across stores in accordance with positive assortative matching (see Appendix F for details). The negative covariance we find exists in both our sample companies, and is encountered when Fenizia (2022) investigates the matches between heterogeneous-quality Italian government offices and managers. This suggests other mechanisms might be responsible for what we observe. We discuss some possible explanations in the next section.

3.3 Negative Assortative Matching?

We discuss three possible explanations for the observed negative assortative matching (NAM). The first two have been found in other settings. The third possibility we identify, a selection-based NAM, is new to the literature and we believe an important source of the

[10]We estimated a separate specification for log productivity that included the interaction of the fixed effects and found that it too has a positive coefficient, providing further evidence of complementarities.

Table 2: Number of manager and store matches

		Stores					
		Company A			Company B		
		Top	Middle	Bottom	Top	Middle	Bottom
Managers	Top	33	51	97	116	229	461
	Middle	49	67	69	225	332	285
	Bottom	88	76	64	457	322	173

This table shows the number of manager-store pairs, split by manager and store quality. In each case, managers and stores are divided into terciles based on their position in the distribution of fixed effects (EB-adjusted estimates of the fixed effects are used for this exercise; see section 3.4). Note that the same manager can be matched to multiple stores along their career, and all of those are reflected above.

negative assortative matching in our setting and, quite possibly, many other applications of AKM-type decompositions.

(i) Lack of Awareness. 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. For example, a related paper by Cowgill et al. (2021) shows that the firm's (or CEO's) preferred match is negatively assortative, but a self-organized match is positively assortative. Given that central HR and regional managers have strong roles in determining matches, this might help explain our result. This unawareness might also reflect the findings of previous research suggesting companies do not always promote the best people to management (Lazear, 2004; Benson, Li and Shue, 2019). Either story indicates companies would benefit from measuring and learning more about managerial quality.

(ii) Learning & Reducing Productivity Variation. Another possibility is that our model is incomplete or misspecified. If the *true* production function does not exhibit complementarity between managers and stores, positive assortative matching is not the optimal configuration. For example, companies might assign better managers to worse stores to reduce productivity variation across locations. This would be desirable if risk-sharing priorities mattered (e.g., Adhvaryu et al., 2020), if they wanted to achieve a minimum standard across all locations (e.g., Fenizia, 2022) for branding purposes, or if they viewed store closures as particularly costly.

(iii) Selection-Based NAM. Finally, the observed NAM across managers and stores could indicate a sample-selection effect arising from a) the fact that our empirical model relies on movers for identification, and b) the ways store and manager types affect the likelihood that a store-manager match remains stable. The argument for this case evokes some elements of Simpson's paradox, and is consistent with what we know about the way matches are formed and incentivized within the two companies. We provide an intuitive explanation of this effect here, with more details in Appendix D.

To understand our argument, suppose that firms decide whether a manager should be relocated from their current store by looking at the store's overall productivity and determining if

$$\mu_s + \mu_m \geq k,$$

where $\mu_s + \mu_m$ is the store-level productivity (under our empirical model), and k is an arbitrary threshold. If the store's productivity is above the threshold, the company preserves the manager-store pair ("if it ain't broke, don't fix it"). If, however, productivity is below the threshold, the manager is relocated. Firms could adopt such a rule if, for example, it is costly to observe both store and manager quality separately, or if they can only do so imperfectly (i.e., with some noise).

The key point to realize is that, because the identification of manager and store fixed effects relies on movers, *we can only estimate fixed effects from manager-store pairs that were broken at least once*. Moreover, the rule for which pairs are broken is a function of both types. As a result, if the firm follows a selection rule resembling the one above, the correlation between the estimated fixed effects of managers and stores will be different than the correlation between the productivities of manager and store pairs in the full population (which includes those pairs that were never broken).

One way to understand this result is to realize that when we condition the sample on manager-store pairs that are broken up (call these "unstable matches"), the expected value of a manager's type decreases in the type of their corresponding store, μ_s: $\mathbb{E}[\mu_m|\text{unstable match}] = \mathbb{E}[\mu_m|\mu_m < k - \mu_s]$, which decreases with μ_s. In words, a high productivity manager could only be a mover if they were paired with a low productivity store. The higher the manager type, the lower the store type would have to be to result in a broken pairing.[11] This creates a negative relationship between manager type and

[11] The key idea behind this mechanism has counterparts in many other areas of study; for example, a similar reasoning can explain the apparent negative correlations between capital and productivity in models with firm exit: firms with low productivity only stay if they have high capital, and vice versa.

average store type among unstable matches. This imparts a source of negative type correlation into the (mover) manager-store pairs. For the same reason, the selection of movers will also generally cause the measured variance of manager and store fixed effects to understate their true dispersion across all manager-store pairs.[12]

Figure 2 presents this argument graphically. It shows 1000 independent manager and store fixed effects drawn from a standard bivariate normal distribution. By construction, the correlation between the manager and store effects here is zero. However, once we impose the selection rule—i.e., recognizing movers come from store-manager combinations with an insufficient sum of types (here, those with $k \leq 0$)—we see the correlation between manager and store effects is negative among the unstable pairs (below the threshold, shown in red x's). For that matter, the correlation is also negative among the stable pairs (above the threshold, shown in blue circles).

Figure 2: Random Matches of Managers and Stores

The chart shows the distribution of μ_s and μ_s when both are i.i.d. $\mathcal{N}(0,1)$ and $k = 0$. Unstable matches indicate those below the threshold k; stable matches indicate those above the threshold k.

[12] A competing explanation is that stores have varying degrees of market power, and because fixed effects are estimated based on revenue, stores with higher market power would appear to be more productive. In that sense, the negative correlation between manager and store fixed effects could instead be a consequence of differences in market power across stores (e.g., Bao, De Loecker and Eeckhout, 2022). This mechanism is unlikely to drive our findings for two reasons. First, pricing decisions in our sample companies are made at the corporate level, so revenue differences across stores reflect quantity differences rather than price variation. Second, we do not find a strong relationship between the measured store fixed effect and the distance to its closest competitors (see section 5, Figure 9 in the working paper version of this study, Metcalfe, Sollaci and Syverson, 2023). If competition were a defining factor for store productivity, we would expect that stores closer to competitors would do worse in revenue terms than stores facing no local competition.

Discussion and Extensions. The example above is relatively simple and misses some nuances of the true data generating process. However, the fundamental issues underlying our argument hold under more general cases, as we discuss at length in Appendix D. There, we emphasize the selection of stable and unstable matches by firms is not merely a theoretical notion but is also consistent with features of our data.

Our results are not driven by the particular functional form of the selection rule. In fact, under the assumption of joint normality of μ_m and μ_s and any selection rule of the form $\theta := f(\mu_s, \mu_m) < k$, we obtain an analytical expression for the bias and the conditional (truncated) covariance of types that belong to an unstable match (before they are potentially re-matched). We show (see Appendix D) that

$$\mathrm{Cov}(\mu_s, \mu_m | \theta < k) = \mathrm{Cov}(\mu_s, \mu_m) + \mathrm{bias},$$

where the sign of the bias (a) depends *only* on the properties of the selection rule θ, and (b) is negative under some regularity conditions.

We use the analytical solution above in concert with our data to calibrate a simple model and obtain the implied correlations among all store-manager pairs, including non-movers. This reveals the bias term can be quite large. Our calibration implies that while the observed correlation between manager and store fixed effects (average across the largest connected sets) is -0.57, the population correlation is -0.34. This bias of -0.23 is about 2/3 of the magnitude of the population correlation.

The analytical solution assumes that the initial matches are broken a single time and managers are not allowed to re-match. We turn to simulation exercises to approximate what actually happens in the data when separated managers re-match to different stores, as they typically do. In our first exercise, we vary the (unconditional) correlation between manager and store pairs, ρ, between -1 and 1 and calculate the size of the bias in each case. The bias is always weakly negative under joint normality, equaling zero at exactly $\rho = 1$ or $\rho = -1$. (In Figure 2, this would be equivalent to having all points lined up in a 45-degree, or -45-degree, line.) The bias is largest when ρ is between 0 and 0.5, and in many cases it is large enough to result in a negative observed correlation among movers even when the correlation for the overall population is positive.

In a second exercise, we extend the notion of manager and store matches to *rounds*: in each round managers and stores in unstable matches are identified, the matches broken, and those managers/stores are according re-matched; stable matches are not altered.[13] We show that, as managers and stores are re-matched in each round, the full-sample correlation between the FEs of manager-store pairs can go from negative to positive, even if the same correlation within stable *and* unstable matches remains negative. This can be thought of as a version of Simpson's Paradox, where the within-group correlations are negative, but each group is increasingly segregated on different quadrants of a plot (one in each side of the line $\mu_s + \mu_m = k$), leading to an overall positive correlation (see Figure D.3).

All of these exercises also help to clarify what we mean by "bias" in this context: the difference between the observed (conditional/truncated) covariance between manager and store fixed effects and its population (unconditional) counterpart. In this sense, there is no bias if one is interested only in calculating the covariances between the productivity of managers and stores that move. In many settings, however, we believe researchers would also want to characterize the unconditional correlation.

Another important clarification is that the bias we describe is fundamentally a function of the selection of stable and unstable matches by firms, not due specific assumptions about what firms or managers know. To motivate our simple case above, we assume that firms are not able to separately observe manager and store productivity, but this is not strictly needed for the argument to work. All that is required is that the firm follows the selection rule $\mu_s + \mu_m \geq k$ indicates a stable match.[14]

As a final point, we emphasize this type of selection-based bias we introduce is fundamentally different from other biases discussed in the AKM-related literature. For example, limited mobility bias can be understood as an incidental parameter issue. Measurement error from estimating fixed effects in a relatively small number of observations leads to a bias in the covariance of these effects when multiple fixed effects are included in a model. This bias disappears as the number of observations increases. The issue we present here is structural and based on the economics of how managers and stores are

[13]Note that in this case stable and unstable matches are redefined in each round. It also means that, except from managers and stores that are in stable matches from the beginning (round zero), which are dropped from the sample, we are able to identify the FEs of all managers and stores in both stable and unstable matches from rounds 1 and onward.

[14]There are, however, selection rules that do not produce a bias. For example, if the firms perfectly observes μ_m and μ_s, it could simply adopt a rule that matches the top manager to the top store; this would be equivalent to a match correlation of $\rho = 1$, a case for which we have already shown the bias is zero if μ_s and μ_m are jointly normal. A more plausible case would be that the firm observes the store productivity with some error $(\mu_s + \varepsilon)$, or that it knows μ_s but observes total productivity with a residual $((\mu_s + \mu_m) + \varepsilon)$. In these instances, the selection rule would look something like $(\mu_s + \varepsilon) + \mu_m \geq k$, which is plagued by the same issues as in our simpler example.

matched and unmatched. Increasing sample size or the number of times that managers move stores does not remove the bias, and may actually change its sign (per the discussion above and shown in Appendix D.3).

The selection-based bias is also different from typical biases related to endogeneity. In fact, the point estimates of the fixed effects are not affected at all. They are identified based on productivity changes that occur when a new manager matches with a store, regardless of the productivity type of the manager or store. It is possible that the selection *rule* induces endogeneity bias. We discuss this possibility in Appendix D.4, but argue it is not likely to be a severe problem in our setting.

We note that the selection-based bias we discuss also applies to any empirical design that relies on movers to estimate two-way fixed effects models. As a result, it potentially transfers to other research settings, including traditional AKM applications where wages are a function of worker and firm effects, depending on how firms choose how to retain or let go of specific workers.[15]

3.4 Correcting Measurement Error in Fixed Effects

The limited mobility bias 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 measurement error in our estimates.

One approach the literature has offered to this problem is to refrain from estimating individual fixed effects and instead focus on groups of similar managers or stores (see Bonhomme, Lamadon and Manresa, 2022). Appendix E.1 presents some results from our application of this method, though with varying success across companies. The method requires a first stage where the researcher can successfully group similar managers/stores based on a small set of observable characteristics, and we have unfortunately not found such groupings for both firms in our data.

We focus instead on Empirical Bayes (EB) adjustment of our estimates. This procedure attempts to attenuate the estimated fixed effects towards their true values (see Chandra

[15] For example, if the probability that worker-firm matches remain stable is an increasing function of each type (i.e., the combination of worker and firm productivity types yield a wage that is sufficiently high to keep either side from wanting to break up the pairing), the movers used in the actual AKM decomposition will have worker and firm types that are more negatively correlated than among all worker-firm pairs.

et al., 2016, for a similar application). The intuition behind EB adjustment is built around its assumption that the estimated fixed effects combine the true fixed effect and classical measurement error. The method uses the data to estimate the measurement error distribution and then, from this, backs out the true fixed effects that underlie the estimated ones. The true fixed effects distribution will exhibit less variance as a result.

Empirical Bayes Adjustment. EB-adjustment assumes measurement error causes the estimated fixed effects $\hat{\mu}_i$ (subscript i is used for ease of exposition and can denote either managers or stores) to be random variables that are independently and normally distributed around their true value, μ_i:

$$\hat{\mu}_i | \mu_i, \tau_i^2 \sim \mathcal{N}(\mu_i, \tau_i^2),$$

where τ_i^2 is the variance of the measurement error. We assume that μ_i is itself independently drawn from a normal distribution with mean $\lambda_{c(i)}$ and common variance σ^2. (Notation $c(i)$ indicates the connected set c that contains unit i.) The mean $\lambda_{c(i)}$ is constant within each connected set c, but can vary across connected sets, as the fixed effects are normalized by different values within each of them. For now, assume these parameters are known, so that

$$\mu_i | \lambda_{c(i)}, \sigma^2 \sim \mathcal{N}(\lambda_{c(i)}, \sigma^2).$$

Conditioning on the estimated fixed effects and given the assumptions above, one can derive the posterior distribution for μ_i as

$$\mu_i | \hat{\mu}_i, \tau_i^2, \lambda, \sigma^2 \sim \mathcal{N}(\mu_i^{EB}, (1-\kappa_i)\tau_i^2),$$

where

$$\mu_i^{EB} = (1-\kappa_i)\hat{\mu}_i + \kappa_i \lambda_{c(i)}$$

is the EB-adjusted fixed effect and $\kappa_i = \tau_i^2/(\tau_i^2 + \sigma^2)$. The EB-adjusted fixed effect is the expected value of the true fixed effect (μ_i) conditioning on the estimated fixed effect ($\hat{\mu}_i$). Intuitively, the adjustment attenuates $\hat{\mu}_i$ towards the unconditional mean $\lambda_{c(i)}$ by an amount determined by the variance of the estimated fixed effects (τ_i^2). Note that this procedure nests the case without measurement error, where $\tau_i^2 = 0$ and $\mu_i^{EB} = \hat{\mu}_i$.

The derivation above assumes the parameters τ_i^2, σ^2, and $\lambda_{c(i)}$ are known, but they must be estimated in practice. We follow the process outlined in Appendix C of Chandra et al. (2016) and set $\hat{\tau}_i^2$ to be the estimated variance of $\hat{\mu}_i$. We obtain the estimated means $\hat{\lambda}_{c(i)}$ through a weighted regression of the fixed effect estimates $\hat{\mu}_i$ on a full set of connected set dummies

$$\hat{\lambda}_{c(i)} = (C'WC)^{-1}C'W\hat{\mu}, \quad \text{and} \quad \hat{\sigma}^2 = \max\left\{0, \frac{\sum_i w_i \left\{\left(\frac{N}{N-N_C}\right)(\hat{\mu}_i - \hat{\lambda}_{c(i)})^2 - \hat{\tau}_i^2\right\}}{\sum_i w_i}\right\},$$

where $w_i = 1/(\hat{\tau}_i^2 + \hat{\sigma}^2)$ are weights, W is a diagonal matrix of the w_i's, C is a matrix with the connected set dummies, and μ is the vector that stacks all $\hat{\mu}_i$'s. N and N_C are the number of units and the number of connected sets, respectively. Under these conditions, the (feasible) EB-adjusted fixed effect is

$$\hat{\mu}_i^{EB} = (1 - \hat{\kappa}_i)\hat{\mu}_i + \hat{\kappa}_i\hat{\lambda}_{c(i)}$$

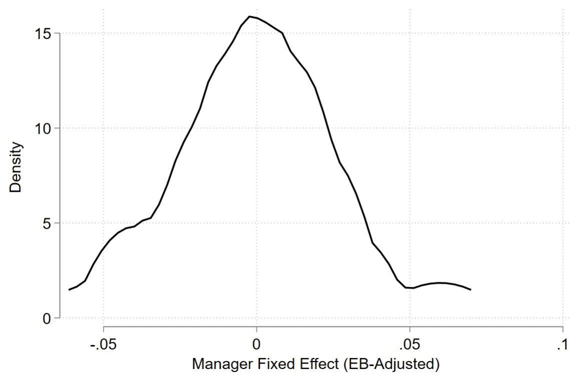
and $\hat{\kappa}_i = \frac{N-N_C-2}{N-N_C}\frac{\hat{\tau}_i^2}{\hat{\tau}_i^2+\hat{\sigma}^2}$. These calculations are made using Adam Sacarny's ebayes program in Stata. Additional discussion on the relationship between the EB adjusted and unadjusted fixed effects as well as on the impact of relaxing the independence assumption on the adjustment procedure can be found in Appendix E.2.

Results. Figure 3 plots the distributions of the EB-adjusted store and manager fixed effects inside each company's largest connected set. The fixed effects have been demeaned within their respective connected set, so we focus on the range of the distributions. Even after correcting for measurement error using this approach, there is a lot of variation in manager and store productivity types. For example, in Company A, replacing a very low quality manager by a very high quality manager would increase the store's productivity by over 10%. The implied change would be even larger in Company B, though there is also more variation in the store effects within this company. In fact, the manager and store fixed effects distributions 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 productivity are causal).[16]

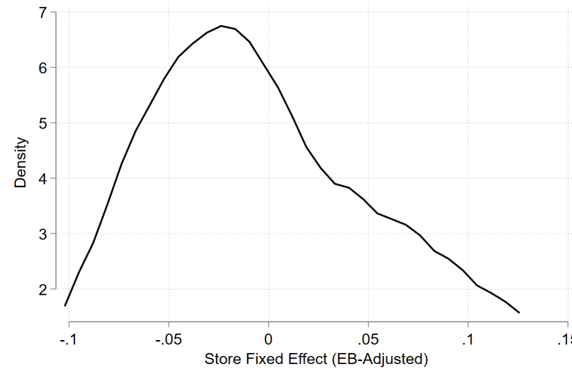
Figures A.2 and A.3 show that these findings apply more broadly. They plot the EB-

[16]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 3 count each store/manager only once.

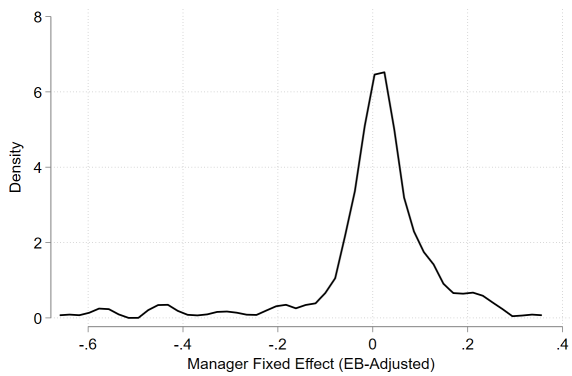
Figure 3: Distribution of Manager and Store Quality (EB-Adjusted)



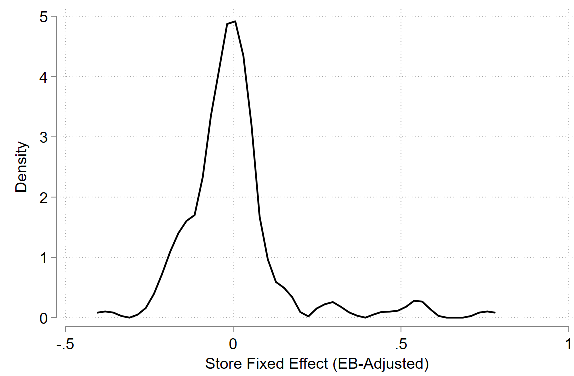
(a) Manager FE - Company A



(b) Store FE - Company A



(c) Manager FE - Company B



(d) Store FE - Company B

Note: the distributions shown in this chart refer to each firm's largest connected set only.

adjusted manager and store fixed effect distributions within each of the four largest connected sets for our two companies. Recall that, within our framework, we can only make such statements within connected sets (since the manager and store effects may be normalized by different quantities across connected sets). As a result, it is possible that our results understate the overall variation in the productivity of managers and stores—for example there is more variation in productivity across connected sets than within them. If that were the case, the potential impact of managers on store-level productivity would be even higher than the numbers we estimate here.

3.5 Out-of-Sample Fit

We close this section by presenting evidence that our estimates of manager and store productivity types perform well, despite the issues presented above, by conducting an out-of-sample fit test. Specifically, we re-estimate our main specification (equation 1) using only the initial two-thirds of the sample for each company to recover manager and store fixed effects. Next, we compute the estimated productivity of store s at time t by adding the estimated fixed effects

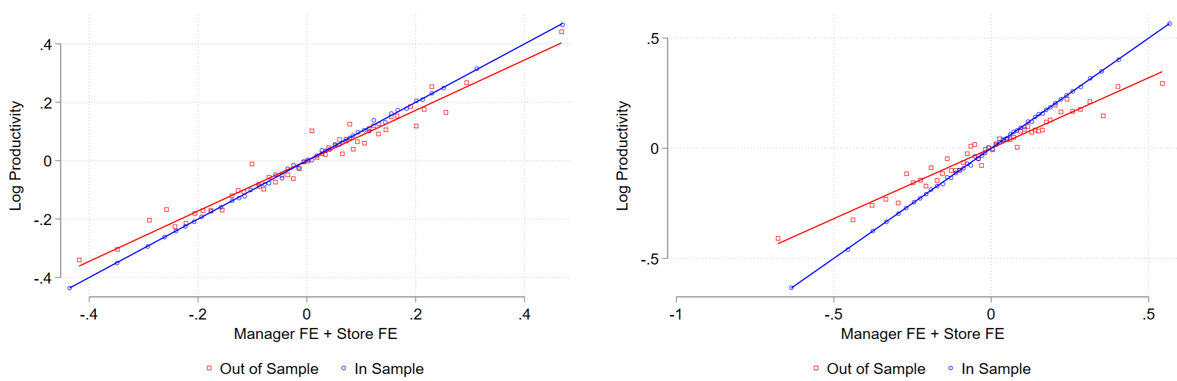
$$log(\widehat{prod}_{s,t}) = \hat{\mu}_{m(s,t)} + \hat{\mu}_s,$$

where $\hat{\mu}_{m(s,t)}$ is the estimated FE of the manager of store s at time t, and $\hat{\mu}_s$ is the estimated FE of store s. Then, to test the out-of-sample fit of the estimated FEs, we run regression

$$\log(prod_{s,t}) = \beta log(\widehat{prod}_{s,t}) + \eta_t + \varepsilon_{s,t},$$

where η_t is a time effect to capture aggregate shocks. The idea is simple: if one runs this regression in the in-sample period, the estimated coefficient $\hat{\beta}$ is, by construction, close to 1. If the out-of-sample predictions are any good, we should expect to see a coefficient that does not deviate much from the in-sample estimate.

Figure 4: Out of Sample Fit



(a) Company A (b) Company B

Note: the charts show how the estimated manager and store fixed effects fit log productivity (controlling for a time fixed effect) in-sample and out-of-sample. For this exercise, the sample is defined as the first 2/3 of observations in the data, and the remaining 1/3 is used to test the out-of-sample fit.

We plot the results in Figure 4. As expected, the coefficient β within the estimation sample is very close to 1. Out-of-sample, this coefficient drops to 0.86 (95% confidence interval [0.83,0.90]) in Company A and 0.64 in Company B (95% confidence interval [0.62, 0.66]). Mean reversion might explain part of this. The adjusted-R^2 for Company A is 0.52 and is 0.34 for Company B. While idiosyncratic store-by-time productivity shocks are clearly important in the out-of-sample periods, we take these results as indicating the model's predictive power is reasonably strong.

4 What Happens When Managers Move?

In this section, we explore what our data can say about the potential causes and consequences of manager changeovers at stores. We estimate a series of event studies to do so. Some attempt to bring such effects into sharper relief by focusing on instances where stores switch from a manager in the low end of the manager fixed effect distribution to one in the high end, and switches in the reverse direction.

We estimate the following specification:

$$y_{st} = \alpha_s + \alpha_t + \mathbf{1}\{s \in \mathbb{T}(t)\} \sum_{k=-6}^{k=4} \beta_k \mathbf{1}\{K_{st} = k\} + \varepsilon_{st}, \tag{3}$$

where y_{st} is a store outcome (productivity, sales, employment, and others); α_s and α_t are respectively store and time fixed effects; $\mathbf{1}\{s \in \mathbb{T}(t)\}$ indicates whether store s belongs to the treated group $\mathbb{T}(t)$ in period t (i.e., after it changed its manager); and $K_{st} = t - E_s$ is the number of months before/after the time of event E_s, defined as the month when the outgoing manager leaves store s.

Note that, unlike most event study designs, the coefficients β_k are estimated up to 6 months *before* the manager leaves the store (we discuss pre-trends more below). Managers' performance in the months immediately preceding their departure may be affected if they know of their move in advance, and store outcomes in the months prior to the managerial change may be related to that change (e.g., if the manager has underperformed). Including the 6 months prior to the move as potentially influenced by the treatment allows us to match treatment and control stores and check for pre-trends using periods less likely influenced by the move.

If managerial changes are primarily based on longer-term or broader-based considerations rather than short-term store dynamics, the coefficients β_k measure the month-by-month causal effects of a managerial change on store-level outcomes. While this assumption is not directly testable, it is bolstered by the nontrivial costs of moves for the company and the manager (including personal costs when moves involve switching to a store in another state or region). Changing stores also typically takes time and requires availability of a slot at the receiving store, so company-wide guidance on staffing tends to discourage frequent moves. In absence of this assumption, the coefficients nevertheless are instructive about any systematic shifts that occur around managerial changes.

To estimate the specification, we follow Borusyak, Jaravel and Spiess (2024) (BJS). This framework addresses many of the potential issues that may arise when estimating event studies, including negative weights on observations (de Chaisemartin and D'Haultfœuille, 2020), while allowing for arbitrary heterogeneity on treatment effects. In simple terms, the strategy in the BJS estimation consists of an imputation procedure where counterfactual outcomes for treated stores are calculated as $\hat{y}_{st}|_{\mathbf{1}\{s\in\mathbb{T}(t)\}} = \hat{\alpha}_s + \hat{\alpha}_t$, with the time and store fixed effects estimated using untreated units only. The average treatment effect is thus a weighted average of the difference between actual outcomes (y_{st}) and the counterfactual $\hat{y}_{st}|_{\mathbf{1}\{s\in\mathbb{T}(t)\}}$. We use the Stata program `did_imputation` to implement.

Figures 5 and 6 plot the estimated coefficients from (3) for six outcome variables: log productivity, log sales, log full-time employment, log part-time employment, the full-to-part-time employment ratio, and log manager salary (only available for Company A) and log energy consumption (only available for company B). We note that these results include all managerial moves in our sample, while our findings in Figures 8 and 9 focus on a subset of moves.

The results on productivity are consistent across companies, displaying a marked decline before the managerial switch, and quick recovery after the new manager comes in (though the month-specific coefficients are often not statistically significant at the 5% level). Comparing this to the sales and employment results indicates the decline is mostly driven by lower sales, while the recovery after the managerial move is a combination of higher sales and fewer employees.

In company A, any impact of a new manager on sales is slow and does not break the previous periods' trend. In contrast, the effect of the new manager on full-time employment is very pronounced in the first few months after the managerial change. However, the employment shift is also temporary, returning to previous levels after three months, but the growth in sales in the meantime keeps productivity at a higher level than before the managerial switch. The event study also reveals a sharp and sudden drop in the manager's salary (over 15%) just as they prepare to leave the store, followed by an immediate recovery as the new manager joins.

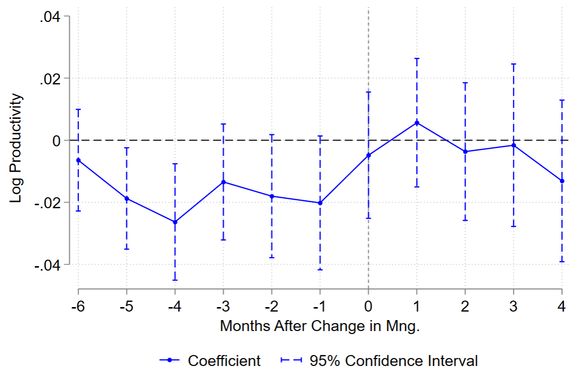
In company B, the sales movements are much more pronounced, with the average store's total revenue increasing about 2% in two months. New managers reduce part-time employment in company B, rather than full-time employment as in company A, and the effects seem persistent (though with large confidence bands). Energy consumption does not follow any obvious pattern as the store manager changes, despite the shift in personnel.

These results highlight important differences in what managers might influence across the two companies, even if any impacts on productivity are relatively similar. Because both companies are in the same sector (though in different countries), these differences might reflect different company cultures, policies that determine manager mobility, or even regulatory environments.

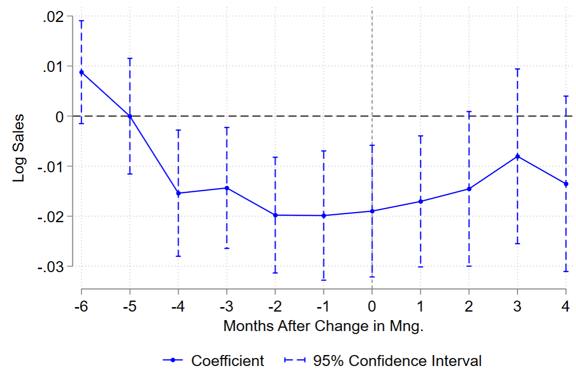
Appendix G contains a number of analyses supporting our specification and results. For one, it shows evidence regarding the absence of pre-trends in the 7-12 months preceding a manager change. We also perform two additional robustness checks. First, our baseline event study regression (3) estimates the impact of managers on stores up to 6 months before the old manager actually moves and 4 months after the new manager joins the store. This is equivalent to an event study with 11 periods (since period 0 is included), but where the "event" is defined 6 months before the old manager leaves. However, within the BJS framework, this distinction can be important because the counterfactual outcomes $\hat{\alpha}_s + \hat{\alpha}_t$ are estimated using untreated observations only, and stores within the 6 months leading up to a managerial change are not included in this group. To test if this distinction is relevant, we run an alternative specification of the event study where stores that are within 6 months of a managerial change are also included in the untreated group, finding very similar results.

Second, the patterns above could be related to the reason for the manager's departure from a store in the first place, say if bonuses drop due to poor performance. We test in Appendix G whether the outcomes are different for departing managers who leave the company altogether (excluding retirements, which we observe in the data) than for those who switch to a different store within the company. We find stayers have better sales and productivity performance than leavers, and the drop in stayers' earnings right before switching stores is considerably attenuated. This is consistent with the hypothesis that some of the leavers were terminated or counseled out of the company.

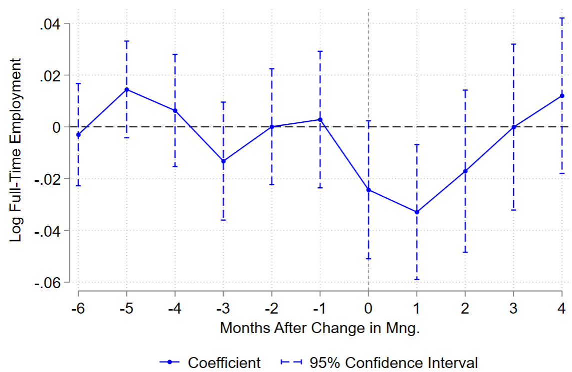
Figure 5: Event Studies: Changes in Outcomes Around (Any) Manager Move
Company A



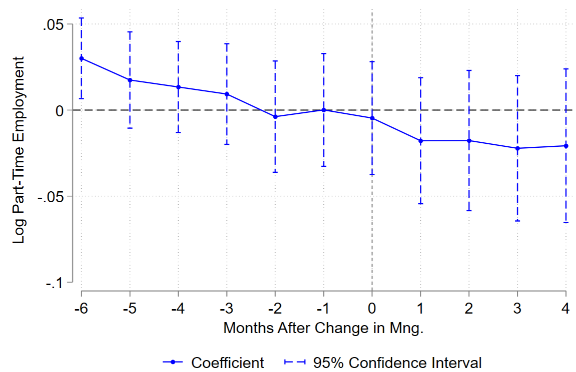
(a) Productivity (log)



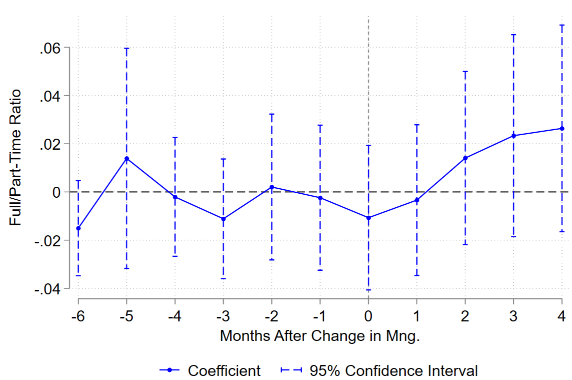
(b) Sales (log)



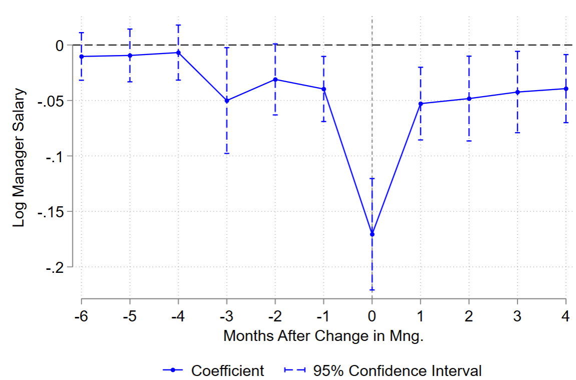
(c) Full-Time Employment (log)



(d) Part-Time Employment (log)



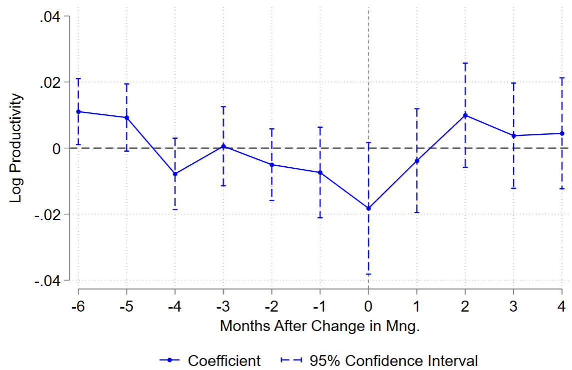
(e) Full/Part-Time Employment (ratio)



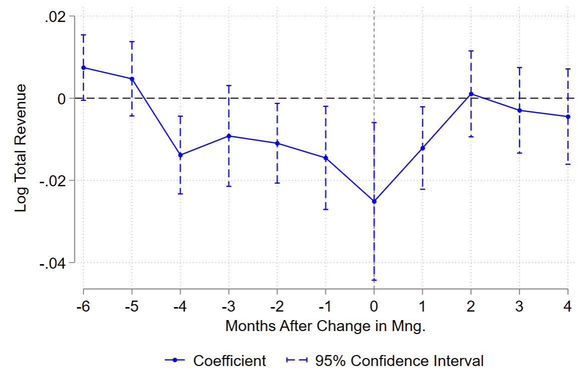
(f) Manager Salary (log)

Note: Dots show the value of each coefficient β_k while whiskers indicate the 95% confidence interval constructed with standard errors that are clustered at the store level (estimated following Borusyak, Jaravel and Spiess, 2024). Sample includes all managerial changes in the data.

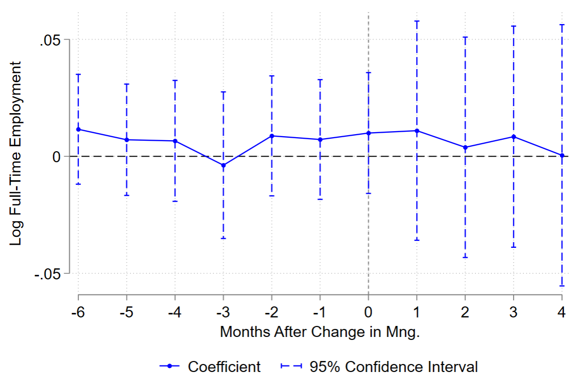
Figure 6: Event Studies: Changes in Outcomes Around (Any) Manager Move
Company B



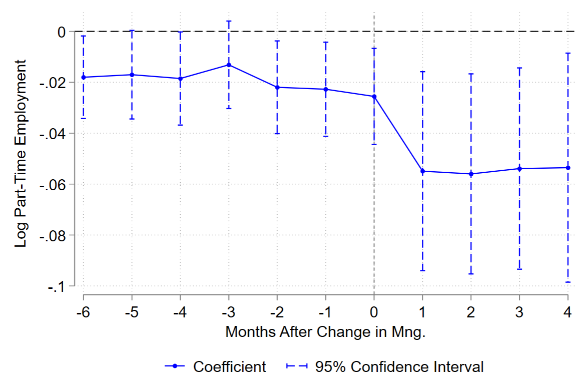
(a) Productivity (log)



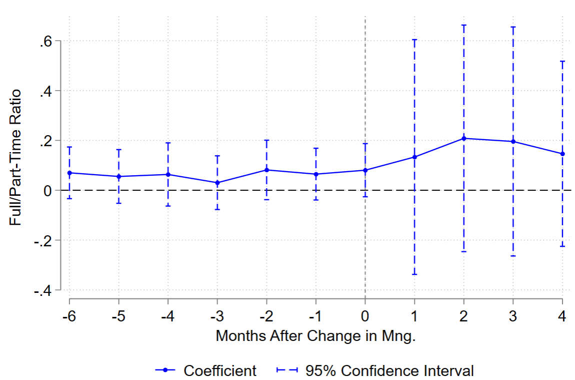
(b) Sales (log)



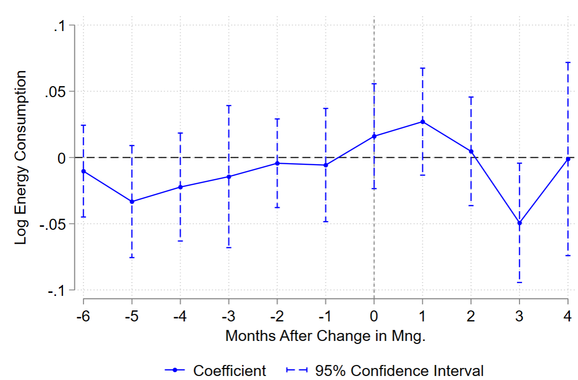
(c) Full-Time Employment (log)



(d) Part-Time Employment (log)



(e) Full/Part-Time Employment (ratio)



(f) Energy Consumption (log)

Note: Dots show the value of each coefficient β_k while whiskers indicate the 95% confidence interval constructed with standard errors that are clustered at the store level (estimated following Borusyak, Jaravel and Spiess, 2024). Sample includes all managerial changes in the data.

Lastly, we acknowledge that the number of periods used to measure the full impact of a new manager over store-level outcomes may be limited, which could miss some longer term effects. However, we also note that many of the impacts we measure in the event studies are surprisingly fast, often immediate, and level off after a few months since a new manager comes into a store. The same observation is true on the event studies below, which show the effects of switching from good to bad (or vice versa) managers. Because of that, we argue it is likely that much of the manager's impact is felt over the first few months, and thus likely captured in our exercises.

4.1 Are Stores Switching to Better Managers?

The event studies shed some light on what might lead to separations between managers and stores. We now investigate whether new store managers are on average different from those they replaced. We consider two manager attributes: their tenure in the company and their productivity fixed effect.

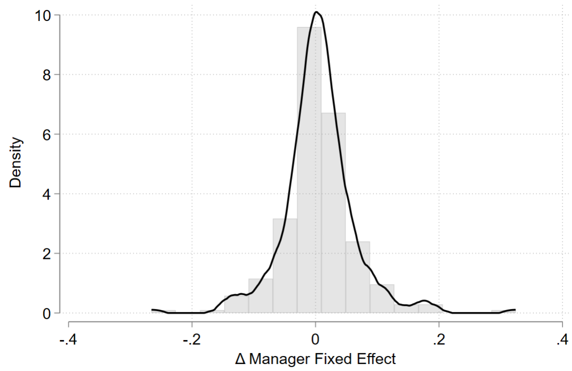
In both companies, incoming managers are on average less experienced than outgoing managers. In Company A the tenure gap is 14 months and significantly different from zero at the 1% level. The gap is smaller in Company B, only 3.4 months and not statistically significant at the usual levels. Compared to the range of the distribution of changes in tenure, those average differences are modest (see Figure A.5).

Regarding their productivity types, Figure 7 shows the distribution of the differences in incoming and outgoing managers' EB-adjusted fixed effects. Panel (a) plots the results for Company A, and panel (b) does so for Company B. Because managers who worked in the same store always share a connected set, we can directly compare their fixed effects. In both companies, the distribution is approximately symmetric around zero, with a mean difference not statistically different from zero at a 10% significance level. Thus, whatever the circumstances that might lead a manager to leave a store, the new manager is on average no different from the one they replaced.[17]

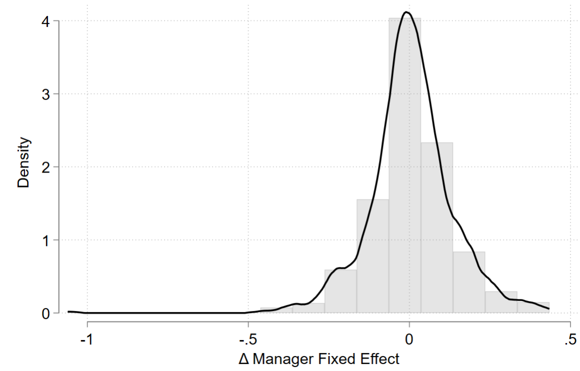
The same exercise can be done from the managers' viewpoints. That is, compute the difference in the productivity fixed effects between the stores they move from and to. Panels (c) and (d) of Figure 7 show, for Companies A and B respectively, the distribution

[17] Note that the sample used to calculate the average managerial change in tenure and in productivity may be slightly different: tenure is only available for managers that move from within the company, but fixed effects can be estimated for some managers come from outside the firm.

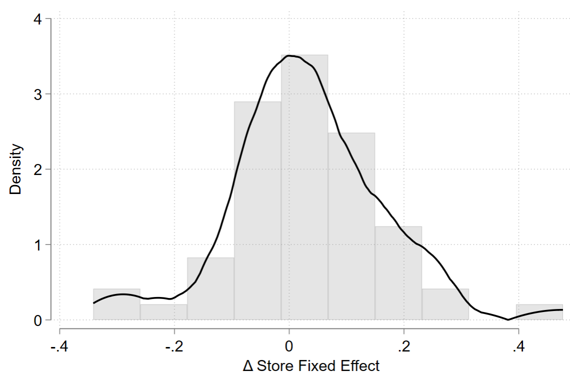
Figure 7: Distribution of Changes in Manager and Store FE



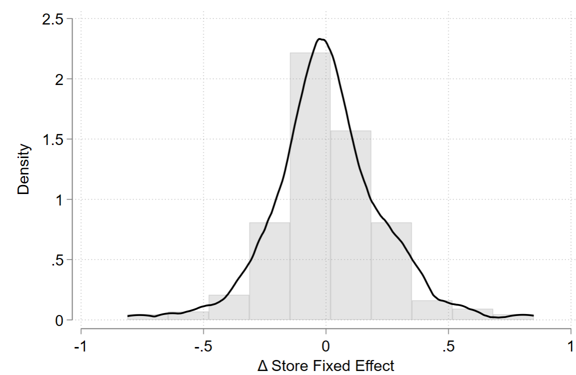
(a) Manager FE (Company A)



(b) Manager FE (Company B)



(c) Store FE (Company A)



(d) Store FE (Company B)

of changes in store fixed effects when managers move. On average, managers move to slightly higher quality stores in both companies, but the differences are small. Indeed, Table 3 shows that managers that start in a low (high, middle) productivity store tend to move to other low (high, middle) productivity stores as well.

4.2 Moving Between Good and Bad Managers

The results in the previous section suggest that when stores switch managers, they are, on average, no better or worse than the manager who left. But what happens when there *is* a significant change in the manager's productivity? To answer this question, we estimate the change in a store's outcomes when it switches from a low to a high productivity manager, and vice versa. To do so, we group managers based on their position relative to the connected set median. We repeat the event study analyses above for two

Table 3: Transition matrix of managerial moves

		Stores					
		Company A			Company B		
		Top	Middle	Bottom	Top	Middle	Bottom
Managers	Top	65	18	17	69	19	12
	Middle	23	61	16	16	62	22
	Bottom	14	20	66	13	18	69

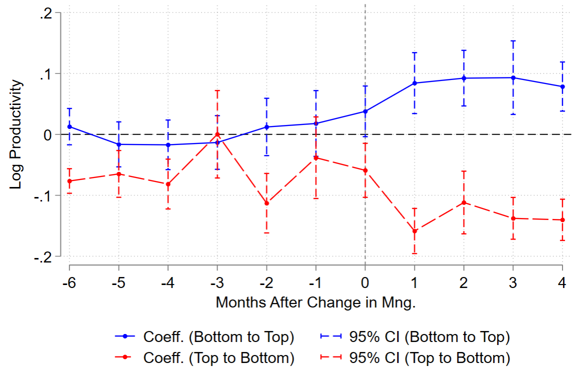
This table shows the transition probabilities (in percent) of managerial changes across stores, based on their relative productivity. For this, managers and stores are split into terciles (top, middle, bottom) of the productivity distribution within their connected sets, and the share of managerial moves is calculated across each group. Note that while stores are split into equally sized groups, the table represents the share of *moves* from one group to another. Hence, if multiple managers move to/from the same store, that store will feature multiple times in the calculation.

store groups: those that start with a top (above-median) manager and switch to a bottom (below-median) manager, and those that start with a bottom manager and switch to a top manager.

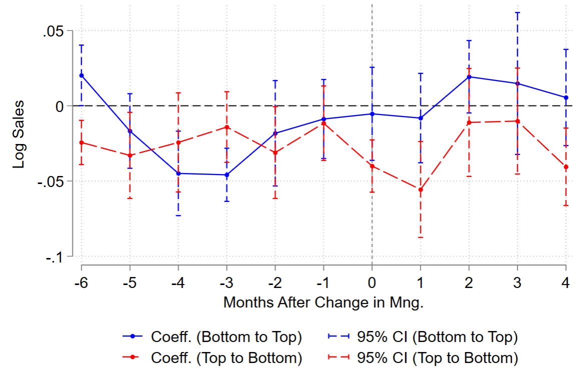
The results are in Figures 8 and 9, with red lines reflecting outcomes at stores switching from managers in the top half of the FE distribution to those in the bottom half, and blue lines reflecting stores that switched from a manager in the bottom half of the distribution to one in the top half. In both companies, though particularly clearly in Company B, there is an immediate increase in store level productivity after a top-end manager arrives, and an immediate productivity decrease when a bottom-end manager arrives.

Interestingly, while some separation is still visible, the difference of the changes in sales for high-low vs low-high manager switches is relatively muted. This indicates that managers achieve productivity gains mostly through changes in staffing. The charts for full- and part-time employment confirm this. Top-end managers in Company A tend to reduce the number of part-time employees in their stores, while top-end managers in Company B reduce both full-time *and* part-time employment. Either way, they do it without negatively impacting sales. Bottom-end managers do the opposite. Their arrival corresponds to an increase in employment and a reduction in sales. Importantly, note that the number of hours that each store is open does not change, nor does there seem to be any impact on the stores' energy consumption (in Company B), which indicates a real gain in productivity from the remaining staff. The productivity gains also do not seem to be reflected on the manager's compensation (company A); apart from a one-month reduction when the new manager moves in, we do not observe large changes in salary compared to the previous manager.

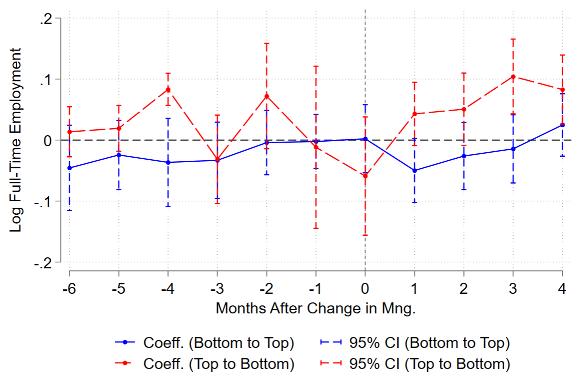
Figure 8: Event Studies: Changes in Outcomes After Switching To Top/Bottom Managers
Company A



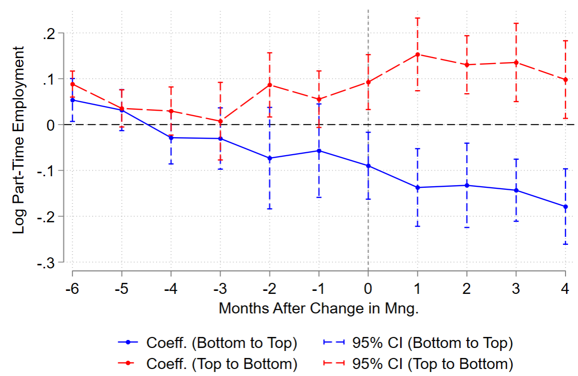
(a) Productivity (log)



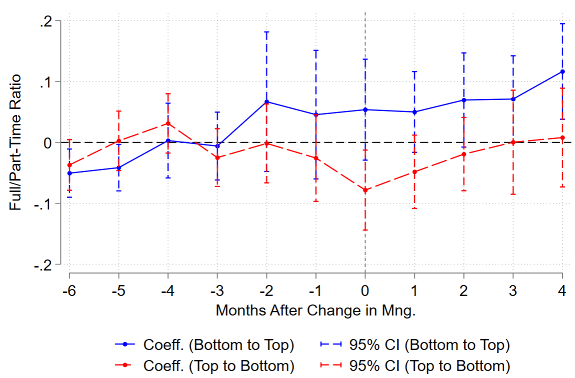
(b) Sales (log)



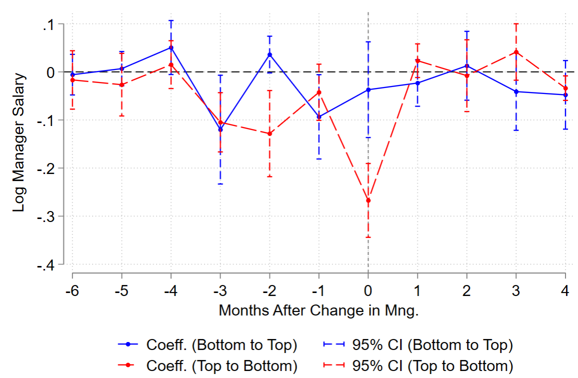
(c) Full-Time Employment (log)



(d) Part-Time Employment (log)



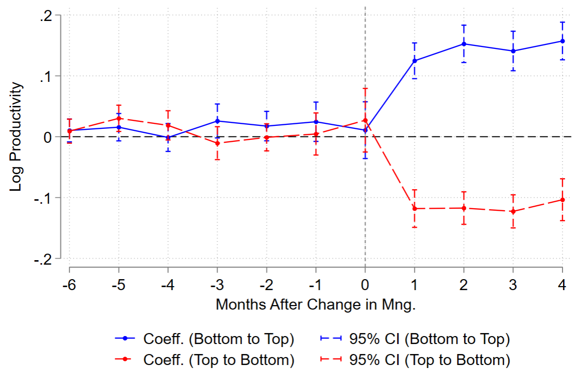
(e) Full/Part-Time Employment (ratio)



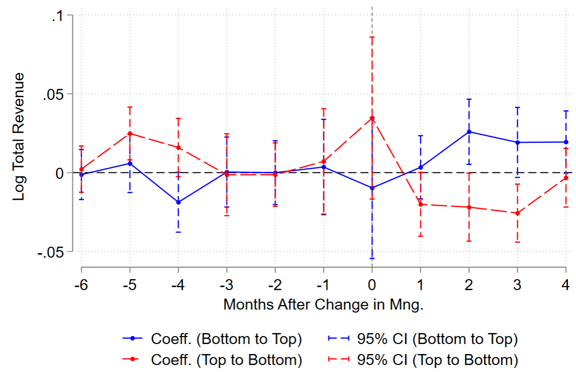
(f) Manager Salary (log)

Note: Dots show the value of each coefficient β_k while whiskers indicate the 95% confidence interval constructed with standard errors that are clustered at the store level (estimated following Borusyak, Jaravel and Spiess, 2024). Sample is restricted to below-to-above and above-to-below median managerial changes in the data.

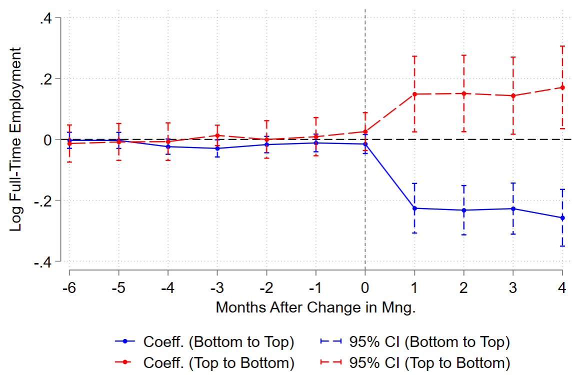
Figure 9: Event Studies: Changes in Outcomes After Switching To Top/Bottom Managers Company B



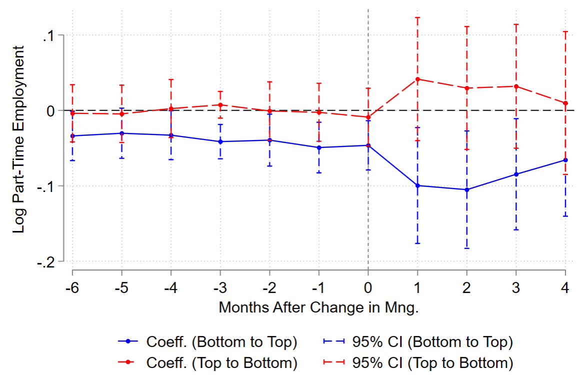
(a) Productivity (log)



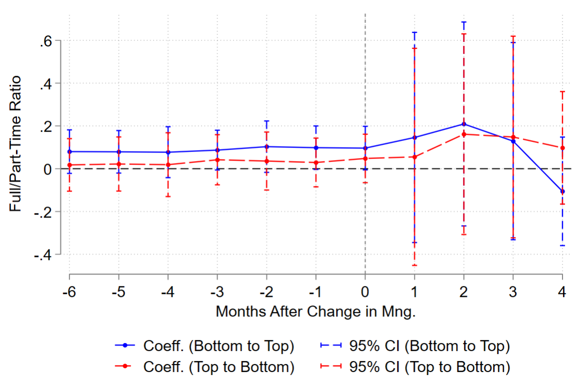
(b) Sales (log)



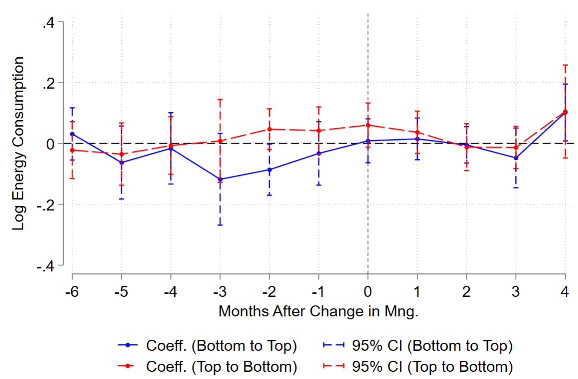
(c) Full-Time Employment (log)



(d) Part-Time Employment (log)



(e) Full/Part-Time Employment (ratio)



(f) Energy Consumption (log)

Note: Dots show the value of each coefficient β_k while whiskers indicate the 95% confidence interval constructed with standard errors that are clustered at the store level (estimated following Borusyak, Jaravel and Spiess, 2024). Sample is restricted to below-to-above and above-to-below median managerial changes in the data.

Table 4: Productivity Effects From Top and Bottom Managers

Panel A: Productivity Effects β_k Before and After Managerial Switch

	Company A			
	Top to Bottom		Bottom to Top	
	Before Change	After Change	Before Change	After Change
Mean	-0.062	-0.121	-0.001	0.077
Standard Deviation	0.039	0.039	0.017	0.023
Difference in Means	-0.059		0.078	

	Company B			
	Top to Bottom		Bottom to Top	
	Before Change	After Change	Before Change	After Change
Mean	0.008	-0.087	0.015	0.117
Standard Deviation	0.014	0.064	0.010	0.061
Difference in Means	-0.095		0.102	

Panel B: Change in (EB-adjusted) Manager Fixed Effects After Managerial Switch

	Company A		Company B	
	Top to Bottom	Bottom to Top	Top to Bottom	Bottom to Top
Mean	-0.05	0.06	-0.100	0.125
Standard Deviation	0.05	0.06	0.08	0.105
Number of Changes	43	40	110	128

Note: Panel A shows the average and standard deviation of the point estimates of β_k in periods $k = -6, \ldots, -1$ (before change) and $k = 0, \ldots, 4$ (after change) when stores switch from an above median manager to a below median manager (top to bottom) and when stores switch from a below median manager to an above median manager (bottom to top). Panel B shows the average and standard deviation of the changes in the EB-adjusted manager fixed effect under the same top to bottom and bottom to top switches.

Magnitude of Productivity Changes. The results presented above suggest relatively large changes in store-level productivity once a new manager is hired. To better understand the scale of the productivity gains or losses that stores experience, we calculate the average of the coefficients β_k (equation 3) in periods $k = 0, \ldots, 4$ and periods $k = -6, \ldots, -1$ (Figures 8 and 9) and take the difference between those averages (see panel A of Table 4). The average increase in productivity experienced by stores in company A when they switch from a bottom manager to a top one is 7.8%, while the same number in company B is approximately 10%. Similarly, when stores in company A move from a top manager to a bottom one, they can expect to become about 6% less productive, while stores in company B are 9.5% less productive—*just from switching managers*!

Panel B Table 4 shows the average change in the EB-adjusted manager fixed effect under the same top-to-bottom/bottom-to-top switches. Overall, the average changes in productivity mentioned above are in line with the average change in the manager's EB-adjusted fixed effect, suggesting that managers are indeed responsible for the observed changes in productivity.

5 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 in which the business operates.

In this paper we show that managers have an effect on productivity that is distinct, relatively quick, and separate from company-level management practices. Quantitatively, we find that while persistent establishment-level attributes explain a larger share of the variance in productivity than managers can, persistent managerial effects still explain a lot, and the total ranges of the two effects' distributions are comparable. We estimate that replacing a manager at the bottom of the distribution by one at the top could increase a store's productivity by at least 10% and perhaps as much as double it, depending on the company and the relevant connected set (Figures 3 and A.2).

There are some obvious caveats to our study. First, we only have data for two retail firms. While it is not particularly difficult to imagine at least some of our findings applying more generally, we cannot test this supposition. Second, while our ability to hold firm-level management practices fixed is an advantage that allows us to zoom in on the roles of managers as individuals, we cannot impose common practices within those firm-level bounds. Relatedly, we do not have access to granular data on what managers do on a moment-to-moment basis or measures of their social and strategic 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 and satisfaction in the workplace (Bandiera et al., 2020; Hansen et al., 2021; Impink, Prat and Sadun, 2024; Antonakis et al., 2022; Dube, Naidu and Reich, 2022; Alan, Corekcioglu and Sutter, 2023). In this way, managers and practices are to some extent still inherently linked in our setting. Third, our sample period is relatively short, so we cannot fully characterize dynamic effects or estimate whether manager fixed effects experience low-frequency drifts due to learning or training. These caveats aside, we believe our results point to there being much more fruitful work to be done in understanding how managers influence the productivity of the establishments they manage.

References

Abowd, John M., Francis Kramarz, and David N. Margolis. 1999. "High Wage Workers and High Wage Firms." *Econometrica*, 67(2): 251–333.

Abowd, John M, Francis Kramarz, Paul Lengermann, and Sébastien Pérez-Duarte. 2004. "Are good workers employed by good firms? A test of a simple assortative matching model for France and the United States." *Unpublished manuscript*.

Acemoglu, Daron, Alex He, and Daniel le Maire. 2022. "Eclipse of Rent-Sharing: The Effects of Managers' Business Education on Wages and the Labor Share in the US and Denmark." National Bureau of Economic Research NBER Working Papers 29874.

Adhvaryu, Achyuta, Namrata Kala, and Anant Nyshadham. 2019. "Management and shocks to worker productivity." National Bureau of Economic Research.

Adhvaryu, Achyuta, Vittorio Bassi, Anant Nyshadham, and Jorge A Tamayo. 2020. "No line left behind: Assortative matching inside the firm." National Bureau of Economic Research.

Alan, Sule, Gozde Corekcioglu, and Matthias Sutter. 2023. "Improving workplace climate in large corporations: A clustered randomized intervention." *Quarterly Journal of Economics*, 138(1): 151–203.

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, John, Giovanna d'Adda, Roberto Weber, and Christian Zehnder. 2022. "Just words? Just speeches? On the economic value of charismatic leadership." *Management Science*, 68(9): 6355–6381.

Baltrunaite, Audinga, Giulia Bovini, and Sauro Mocetti. 2023. "Managerial talent and managerial practices: are they complements?" *Journal of Corporate Finance*, 102348.

Bandiera, Oriana, Andrea Prat, Stephen Hansen, and Raffaella Sadun. 2020. "CEO behavior and firm performance." *Journal of Political Economy*, 128(4): 1325–1369.

Bandiera, Oriana, Luigi Guiso, Andrea Prat, and Raffaella Sadun. 2015. "Matching firms, managers, and incentives." *Journal of Labor Economics*, 33(3): 623–681.

Bao, Renjie, Jan De Loecker, and Jan Eeckhout. 2022. "Are Managers Paid for Market Power?" National Bureau of Economic Research Working Paper 29918.

Barberis, Nicholas, Maxim Boycko, Andrei Shleifer, and Natalia Tsukanova. 1996. "How does privatization work? Evidence from the Russian shops." *Journal of Political Economy*, 104(4): 764–790.

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

Bender, Stefan, Nicholas Bloom, David Card, John Van Reenen, and Stefanie Wolter. 2018. "Management practices, workforce selection, and productivity." *Journal of Labor Economics*, 36(S1): S371–S409.

Benson, Alan, Danielle Li, and Kelly Shue. 2019. "Promotions and the Peter principle." *Quarterly Journal of Economics*, 134(4): 2085–2134.

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

Best, Michael Carlos, Jonas Hjort, and David Szakonyi. 2017. "Individuals and organizations as sources of state effectiveness." National Bureau of Economic Research.

Bianchi, Nicola, and Michela Giorcelli. 2022. "The dynamics and spillovers of management interventions: Evidence from the training within industry program." *Journal of Political Economy*, 130(6): 1630–1675.

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

Bloom, Nicholas, Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts. 2013. "Does Management Matter? Evidence from India." *Quarterly Journal of Economics*, 128(1): 1–51.

Bloom, Nicholas, Erik Brynjolfsson, Lucia Foster, Ron Jarmin, Megha Patnaik, Itay Saporta-Eksten, and John Van Reenen. 2019. "What drives differences in management practices?" *American Economic Review*, 109(5): 1648–83.

Bloom, Nicholas, Raffaella Sadun, and John Van Reenen. 2012. "The organization of firms across countries." *Quarterly Journal of Economics*, 127(4): 1663–1705.

Bloom, Nicholas, Renata Lemos, Raffaella Sadun, Daniela Scur, and John Van Reenen. 2014. "The New Empirical Economics of Management." *Journal of the European Economic Association*, 12(4): 835–876.

Bonhomme, Stéphane, Thibaut Lamadon, and Elena Manresa. 2022. "Discretizing Unobserved Heterogeneity." *Econometrica*, 90(2): 625–643.

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess. 2024. "Revisiting Event-Study Designs: Robust and Efficient Estimation." *The Review of Economic Studies*, 91(6): 3253–3285.

Boudreau, Laura, Rocco Macchiavello, Virginia Minni, and Mari Tanaka. 2021. "Union Leaders: Experimental Evidence From Myanmar."

Bruhn, Miriam, Dean Karlan, and Antoinette Schoar. 2018. "The Impact of Consulting Services on Small and Medium Enterprises: Evidence from a Randomized Trial in Mexico." *Journal of Political Economy*, 126.

Cai, Jing, and Shing-Yi Wang. 2022. "Improving management through worker evaluations: Evidence from auto manufacturing." *Quarterly Journal of Economics*, 137(4): 2459–2497.

Card, David, Ana Rute Cardoso, Joerg Heining, and Patrick Kline. 2018. "Firms and labor market inequality: Evidence and some theory." *Journal of Labor Economics*, 36(S1): S13–S70.

Card, David, Jörg Heining, and Patrick Kline. 2013. "Workplace heterogeneity and the rise of West German wage inequality." *Quarterly journal of economics*, 128(3): 967–1015.

Chandra, Amitabh, Amy Finkelstein, Adam Sacarny, and Chad Syverson. 2016. "Health Care Exceptionalism? Performance and Allocation in the US Health Care Sector." *American Economic Review*, 106(8): 2110–44.

Chew, William Bruce, Kim B Clark, and Timothy F Bresnahan. 1989. *Measurement, coordination and learning in a multi-plant network.* Division of Research, Harvard Business School.

Correia, Sergio. 2017. "Linear Models with High-Dimensional Fixed Effects: An Efficient and Feasible Estimator."

Costa, Dora L., and Matthew E. Kahn. 2000. "Power Couples: Changes in the Locational Choice of the College Educated, 1940–1990*." *The Quarterly Journal of Economics*, 115(4): 1287–1315.

Cowgill, Bo, Jonathan Davis, B Pablo Montagnes, and Patryk Perkowski. 2021. "Matchmaking Principals: Theory and Evidence from Internal Labor Markets."

Dauth, Wolfgang, Sebastian Findeisen, Enrico Moretti, and Jens Suedekum. 2022. "Matching in cities." *Journal of the European Economic Association*, 20(4): 1478–1521.

de Chaisemartin, Clément, and Xavier D'Haultfœuille. 2020. "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects." *American Economic Review*, 110(9): 2964–96.

Di Addario, Sabrina, Patrick Kline, Raffaele Saggio, and Mikkel Sølvsten. 2022. "It ain't where you're from, it's where you're at: hiring origins, firm heterogeneity, and wages." *Journal of Econometrics*.

Dube, Arindrajit, Suresh Naidu, and Adam D Reich. 2022. "Power and Dignity in the Low-Wage Labor Market: Theory and Evidence from Wal-Mart Workers." National Bureau of Economic Research.

Eeckhout, Jan, and Philipp Kircher. 2011. "Identifying sorting – in theory." *Review of Economic Studies*, 78(3): 872–906.

Eeckhout, Jan, and Philipp Kircher. 2018. "Assortative matching with large firms." *Econometrica*, 86(1): 85–132.

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

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

Frederiksen, Anders, Lisa B Kahn, and Fabian Lange. 2020. "Supervisors and performance management systems." *Journal of Political Economy*, 128(6): 2123–2187.

Friebel, Guido, Matthias Heinz, and Nikolay Zubanov. 2022. "Middle Managers, Personnel Turnover, and Performance: A Long-Term Field Experiment in a Retail Chain." *Management Science*, 68(1): 211–229.

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

Gaure, Simen. 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'.

Giardili, Soledad, Kamalini Ramdas, and Jonathan W Williams. 2022. "Leadership and productivity: a study of US automobile assembly plants." *Management Science*.

Giorcelli, Michela. 2019. "The Long-Term Effects of Management and Technology Transfers." *American Economic Review*, 109(1): 121–52.

Gosnell, Greer K., John List, and Robert 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, Marcus, Tzuo Hann Law, and Iourii Manovskii. 2017. "Identifying equilibrium models of labor market sorting." *Econometrica*, 85(1): 29–65.

Hansen, Stephen, Tejas Ramdas, Raffaella Sadun, and Joe Fuller. 2021. "The Demand for Executive Skills." National Bureau of Economic Research.

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

Hoffman, Mitchell, Lisa B Kahn, and Danielle Li. 2018. "Discretion in hiring." *Quarterly Journal of Economics*, 133(2): 765–800.

Huber, Kilian, Volker Lindenthal, and Fabian Waldinger. 2021. "Discrimination, managers, and firm performance: Evidence from aryanizations in nazi Germany." *Journal of Political Economy*, 129(9): 2455–2503.

Ichniowski, Casey, and Kathryn 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, Casey, and Kathryn 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, Casey, Kathryn Shaw, and Giovanna 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, Stephen Michael, Andrea Prat, and Raffaella Sadun. 2024. "Communication within Firms: Evidence from CEO Turnovers." *Management Science*, 71(1): 470–487.

Janke, Katharina, Carol Propper, and Raffaella Sadun. 2019. "The impact of CEOs in the public sector: Evidence from the English NHS." National Bureau of Economic Research.

Kaplan, Steven N, and Morten Sorensen. 2021. "Are CEOs different?" *Journal of Finance*, 76(4): 1773–1811.

Lazear, Edward P. 2004. "The Peter Principle: A theory of decline." *Journal of political economy*, 112(S1): S141–S163.

Lazear, Edward P, Kathryn L Shaw, and Christopher T Stanton. 2015. "The value of bosses." *Journal of Labor Economics*, 33(4): 823–861.

Limodio, Nicola. 2021. "Bureaucrat allocation in the public sector: evidence from the World Bank." *Economic Journal*, 131(639): 3012–3040.

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

Metcalfe, Robert D., Alexandre B. Sollaci, and Chad Syverson. 2023. "Managers and Productivity in Retail." National Bureau of Economic Research.

Minni, Virginia. forthcoming. "Making the Invisible Hand Visible: Managers and the Allocation of Workers to Jobs." *Quarterly Journal of Economics*.

Mollick, Ethan. 2012. "People and process, suits and innovators: The role of individuals in firm performance." *Strategic Management Journal*, 33(9): 1001–1015.

Patault, Berengere, and Clemence Lenoir. 2024. "How valuable are business networks? evidence from sales managers in international markets." *American Economic Journal: Applied Economics*, 16(4): 404–443.

Rubens, Michael. 2023. "Management, productivity, and technology choices: Evidence from US mining schools." *RAND Journal of Economics*, 54(1): 165–186.

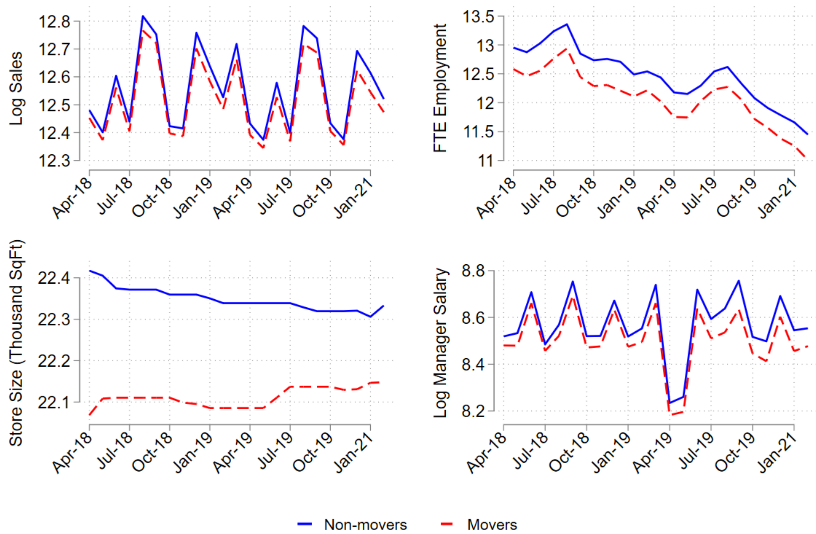
Sandvik, Jason J, Richard E Saouma, Nathan T Seegert, and Christopher T Stanton. 2020. "Workplace knowledge flows." *Quarterly Journal of Economics*, 135(3): 1635–1680.

Sorkin, Isaac. 2018. "Ranking firms using revealed preference." *Quarterly Journal of Economics*, 133(3): 1331–1393.

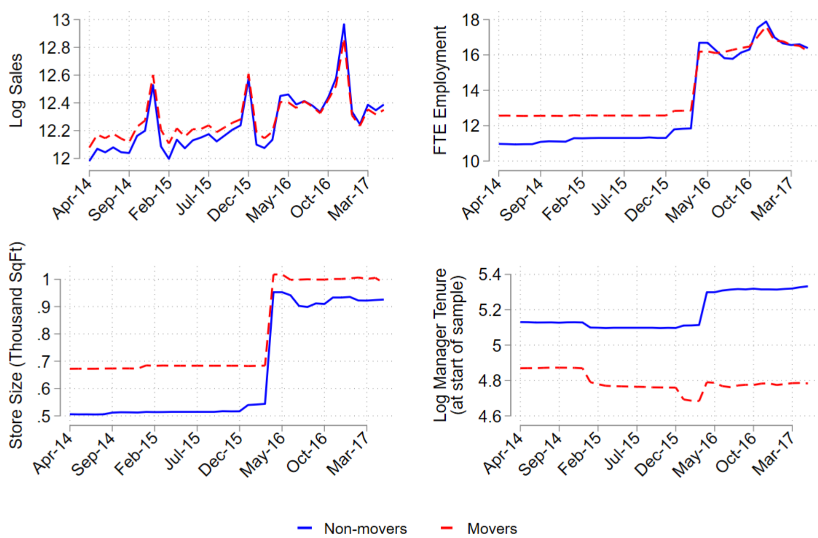
Syverson, Chad. 2011. "What Determines Productivity?" *Journal of Economic Literature*, 49(2): 326–65.

A Additional Figures

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



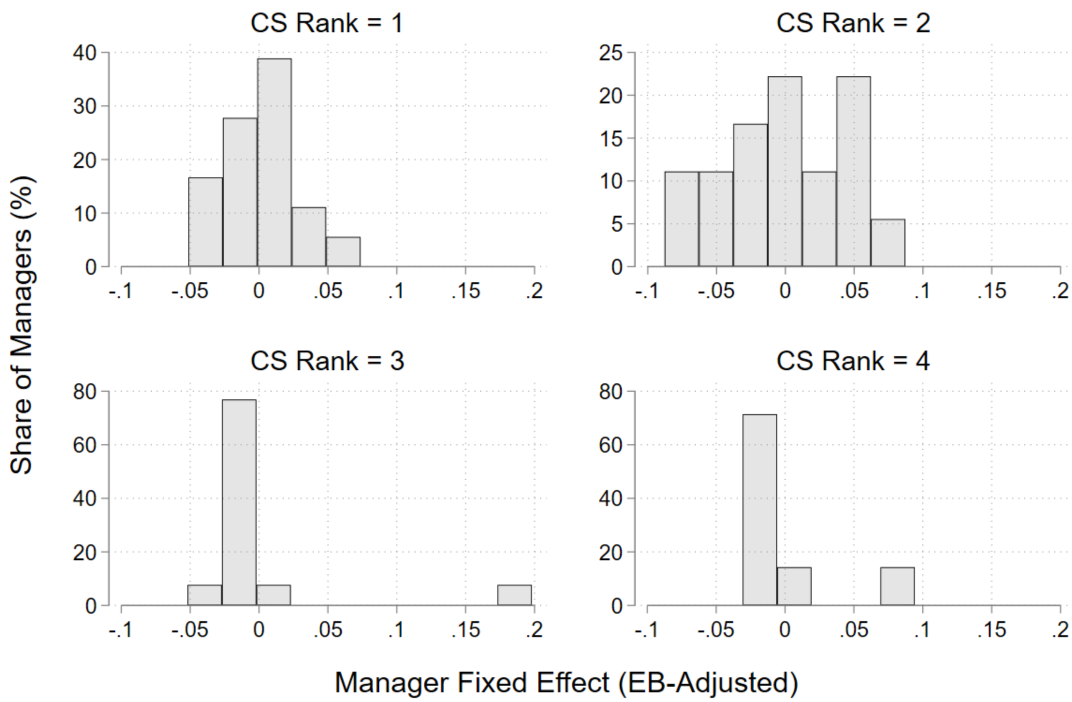
(a) Company A



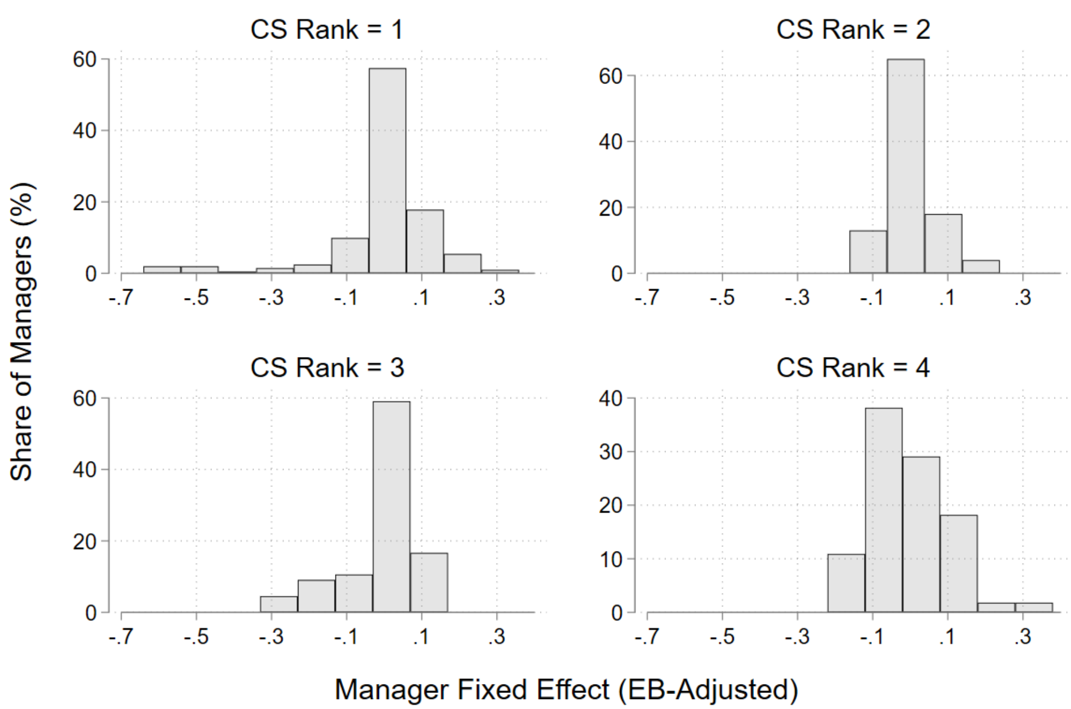
(b) Company B

Note: starting in May 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



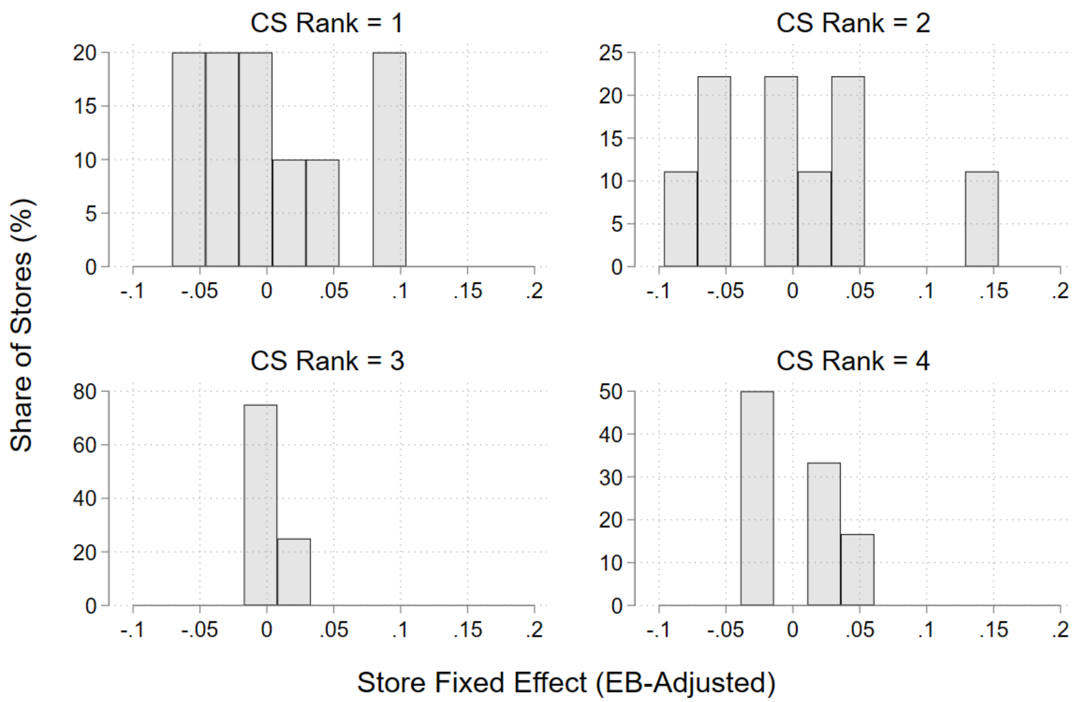
(a) Company A



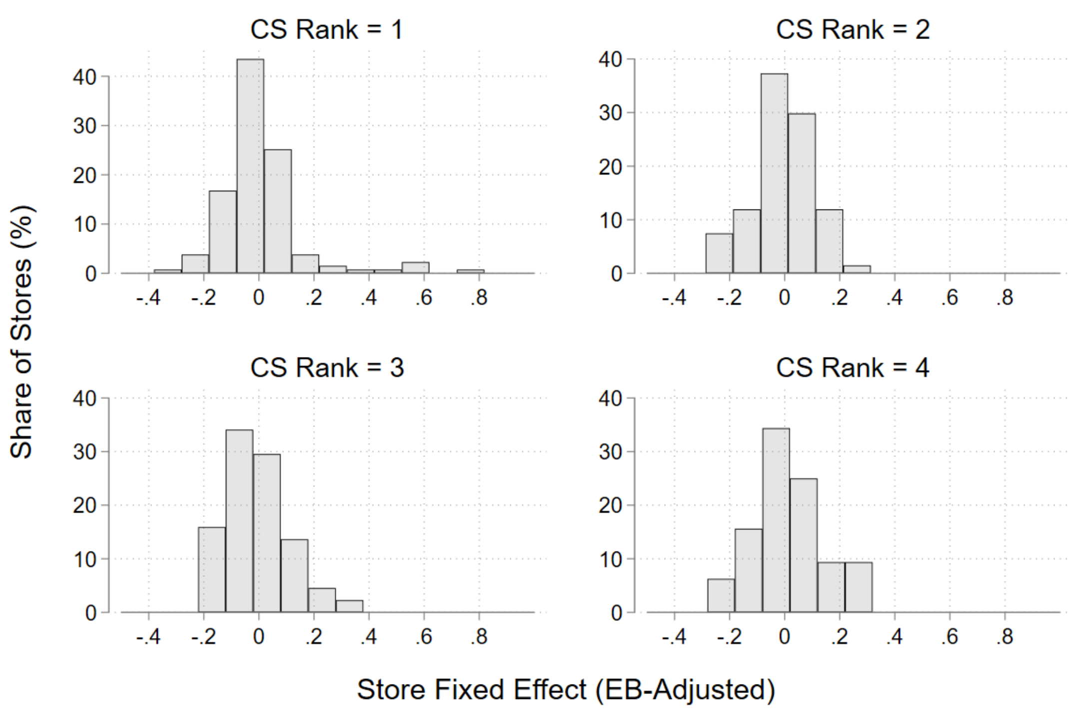
(b) Company B

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

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



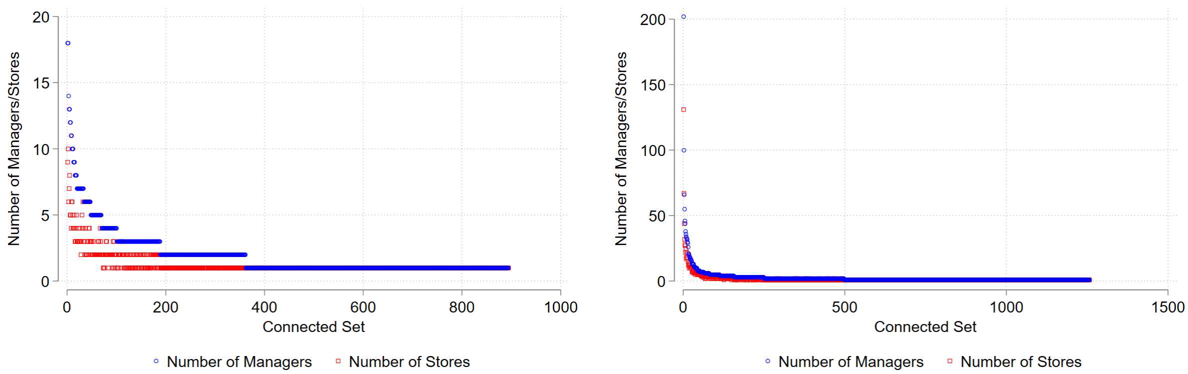
(a) Company A



(b) Company B

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

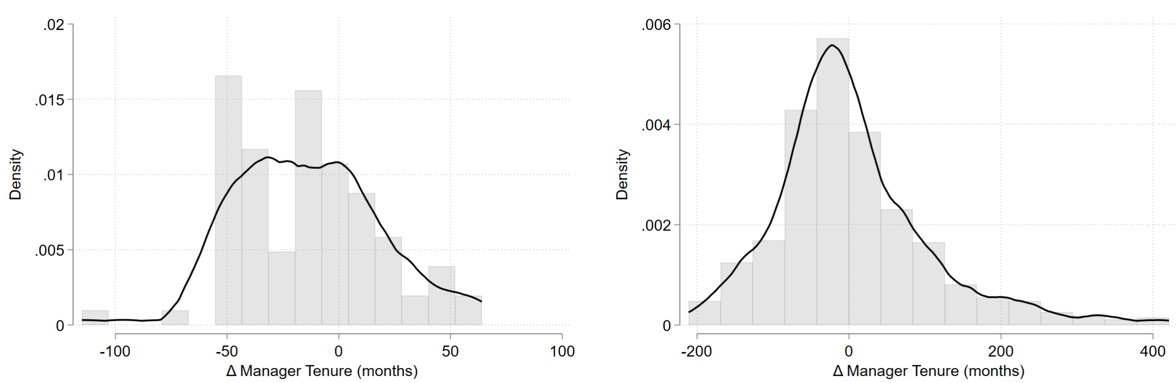
Figure A.4: Number of Stores and Managers by Connected Set



(a) Company A

(b) Company B

Figure A.5: Distribution of Changes in Manager Tenure



(a) Company A

(b) Company B

B Additional 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	1% 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	
Managers Tenure (movers vs. non-movers)		26% lower

Note: Numbers rounded to nearest integer. 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: Regression of Female Dummy of Manager and Store Characteristics

	(1)	(2)	(3)	(4)
Mover	-0.067***	-0.054***	-0.031	-0.051
	(0.018)	(0.019)	(0.037)	(0.038)
log Tenure		0.096***	0.110***	0.113***
		(0.012)	(0.027)	(0.022)
Sample	All managers	All managers	Did not leave	10+ stores in town
N	3,309	2,955	716	828
Within R^2	0.004	0.035	0.037	0.039

Note: Robust standard errors are shown in parenthesis. ***, **, and * indicate that coefficients are significantly different from zero at the 1%, 5% and 10% levels, respectively. Columns (2) - (4) control for log(revenue), log(FTE count), log(store area), as well as format (flagship, local, etc) and location (NUTS 2) fixed effects. In the case of movers, those variables are computed for the first store where the manager is observed in the data.

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[18]

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

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	0.69*	0.76*	1

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

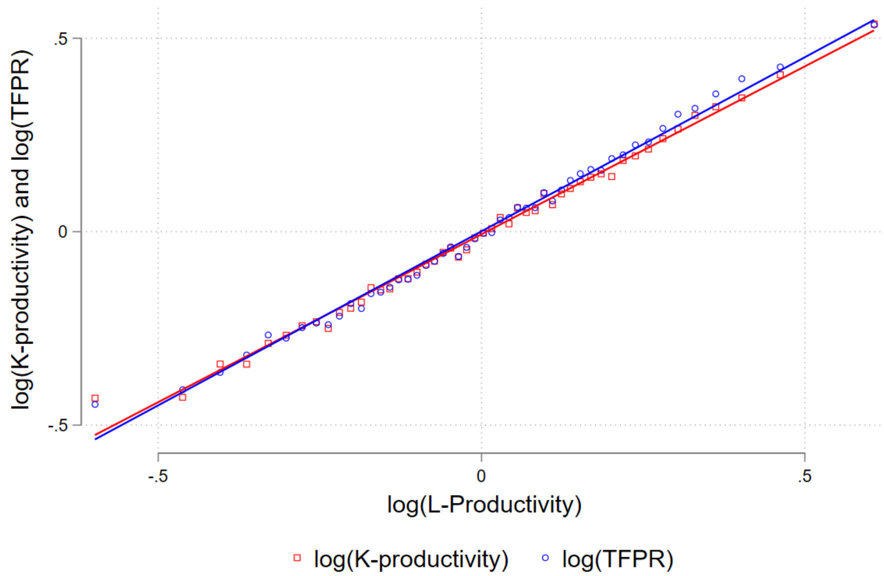
C.2 Augmented Models of Manager and Store Productivity

Local Shocks. Our baseline decomposition of store productivity, equation (1), controls for company-wide shocks through the time fixed effect. This section augments the model to include location-month fixed effects that account for local shocks:

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

[18] 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 Producivity



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.

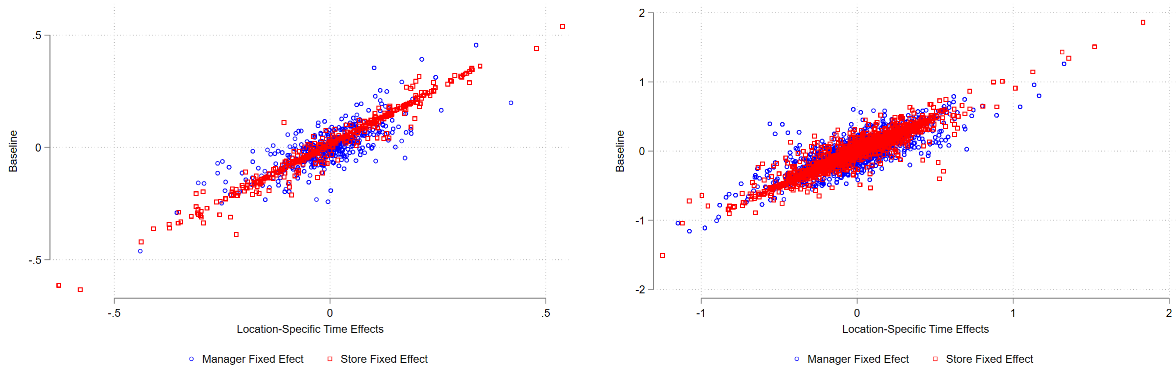
where $\mu_{l,t}$ is an area-month fixed effect (for Company A, area is the store's city, and for Company B it is the NUTS2 region). The estimated manager and store fixed effects under the baseline and in the specification above are highly correlated. In Company A the correlation between store fixed effects under the two methods is 0.98, and the correlation between manager fixed effects is 0.76. In Company B, these correlations are 0.93 and 0.80, respectively. Their overall distributions are very similar as well, as can be seen in this letter's Figure C.2 below.

Match-Specific Effects. The baseline model specification also 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

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

where $\mu_{s,m(s,t)}$ is a store-manager pair fixed effect, which captures any sorting effect between managers/stores. Figure C.3 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). For

Figure C.2: Effect of Controlling for Area-Specific Seasonality



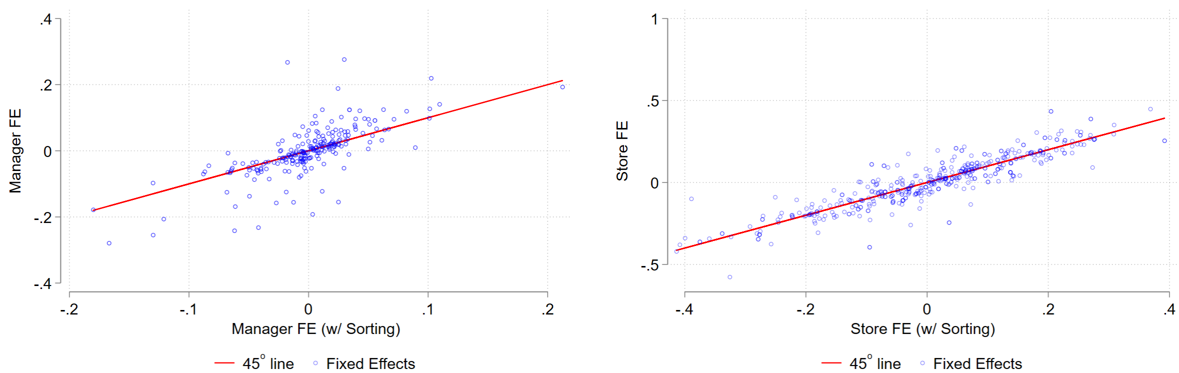
(a) Company A

(b) Company B

Note: the charts compare the estimated manager and store fixed effects under the baseline specification (including a time FE only) and under a location-specific seasonality effects (location-time FE), where location is measured as the city for Company A and the NUTS2 region for Company B.

brevity, we present the figures for company A only, though they are very similar when looking at company B. The 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.

Figure C.3: Comparing Estimates With and Without Sorting



(a) Manager Fixed Effects

(b) Store Fixed Effects

Note: the charts compare the estimated manager and store fixed effects under the baseline specification and under an alternative specific that accounts for match-specific (sorting) effects. The results presented focus on company A only, with similar findings for company B.

Persistence. One can also posit whether there is persistence of the manager effect, for example if the impact of a good manager lasts even after they leave a store. This would imply a model where

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

where $\mu_{m^*(s,t-k)}$ was the manager of the store k months ago. We adopt a simple test: if the “correct" model does indeed include the effect of past managers, then estimating the model above should result in a larger adjusted-R^2 than the “plain" model we consider in the text. Using $k \in \{6, 12\}$, we obtain an adjusted-R^2 of 0.8282 and 0.8071, respectively, which are barely above the baseline of 0.8064. We interpret this as limited evidence of a persistent direct effect of past managers in store-level productivity.

Learning. A final possibility we consider here is that managers learn over time. To test whether this would affect our baseline model, we again compare the adjusted-R^2 of our baseline model with the adjusted-R^2 of the alternative model

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

where $tenure_{m,t}$ measures the manager's tenure (we consider two measures: overall tenure at the company, and tenure as a manager).

Again, we find very small gains in the model's explanatory power, and therefore no strong evidence that this specification dominates our baseline. Following the strategy above, we can look at the adjusted-R^2 in the tenure-augmented model and compare it to the baseline model. Here we find adjusted-R^2 of 0.8110 and 0.8121 for the models with tenure (in the manager position and in the company, respectively), again very close to the baseline of 0.7861 (slightly different from the example above as the regression sample changes when tenure is included in the model). We note that this does not necessarily imply that managers do not learn over time, but perhaps that learning is not sufficiently captured by tenure only, or that our sample is too short to observe significant differences in performance over time.

D Selection-Based Matching

In section 3.3, the combination of a selection rule that breaks up any manager-store pair such that $\mu_m + \mu_s < k$, combined with the fact that we can only estimate the fixed effect of managers/stores that were part of an unstable (broken) match at least once, generates a bias in the correlation between the *observed* manager and store fixed effects. Here, we explore some points made in that section in more detail.

D.1 Empirical Evidence for the Selection Rule

This selection rule also implies two relationships that are consistent with elements of our data. First, there should be a negative relationship between the probability that a manager moves and the log-productivity of the store, as this would imply a larger $\mu_s + \mu_m$ (after controlling for time fixed effects). We test this relationship by estimating the relationship between managerial changes and store level productivity in a logit model. Due to the low probability that a manager moves multiple times within a short period, the data is aggregated to the store-quarter, and a full set of quarter-year dummies is included as controls. Table D.1 presents the results from this regression, showing a negative and significant impact of log productivity on the probability of moving. The lower coefficient for company B is likely driven by the fact that the vast majority of managerial moves in 2014 and 2015 hapenned only in December (see also Figure D.1).

Second, the average quality of managers in good stores should be increasing over time, because good managers and good stores form stable matches, while bad managers in bad stores are relocated. Figure D.1 shows the average EB-adjusted manager fixed effect within each tercile of the EB-adjusted store fixed effect distribution (top, middle, bottom). No clear pattern arises in company A (possibly due to the shorter time frame of the data). However, the results for company B suggest that the average quality of managers in top stores has increased, the average quality of managers in bottom stores has decreased.

D.2 A Multivariate Normal Example

To clarify the discussion on what we mean by a "bias" and its source, we solve for a specific case where the pair of fixed effects (μ_s, μ_m) is assumed to be jointly normally distributed. Note that we can set $\mathbb{E}[\mu_s] = \mathbb{E}[\mu_m] = 0$ without loss of generality, as both are fixed effects from a linear regression. We therefore have

$$\begin{pmatrix} \mu_s \\ \mu_m \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_s^2 & \rho\sigma_s\sigma_m \\ \rho\sigma_s\sigma_m & \sigma_m^2 \end{pmatrix} \right).$$

Table D.1: Probability of Moving and Store-Level Productivity

	Company A	Company B
log productivity	-1.663***	-0.212**
	(0.175)	(0.106)
N	9,484	17,746
Pseudo R^2	0.0329	0.3168

Note: The table shows the results of a logit regression of the probability of a managerial change on the log productivity of a store. Due to the low probability of sequential moves in a short period, the data is first aggregated to the store-by-quarter (as opposed to store-month) level. Standard errors in parenthesis and ***, **, and * indicate coefficients are significantly different from zero at the 1%, 5% and 10% levels, respectively. Both specifications include a full set of dummies for each quarter-year.

We also assume that the selection rule to identify unsuccessful matches is $f(\mu_s, \mu_m) := \theta < k$. For now, we let $f(\mu_s, \mu_m)$ be any arbitrary function of the productivity of managers and stores.[19]

Because we only measure the fixed effects of managers and stores that are separated, the *observed* covariance between these fixed effects is conditional on $\theta < k$ (truncated). Using the law of total covariance, this can be written as

$$\mathrm{Cov}(\mu_s, \mu_m|\theta < k) = \mathrm{Cov}[\mathbb{E}(\mu_s|\theta), \mathbb{E}(\mu_m|\theta)|\theta < k] + \mathbb{E}[\mathrm{Cov}(\mu_s, \mu_m|\theta)|\theta < k].$$

Since (μ_s, μ_m) are jointly normal, we have

$$\mathbb{E}[\mu_s|\theta] = \mathbb{E}[\mu_s] + \frac{\mathrm{Cov}(\mu_s, \theta)}{\mathrm{Var}(\theta)}(\theta - \mathbb{E}[\theta]),$$

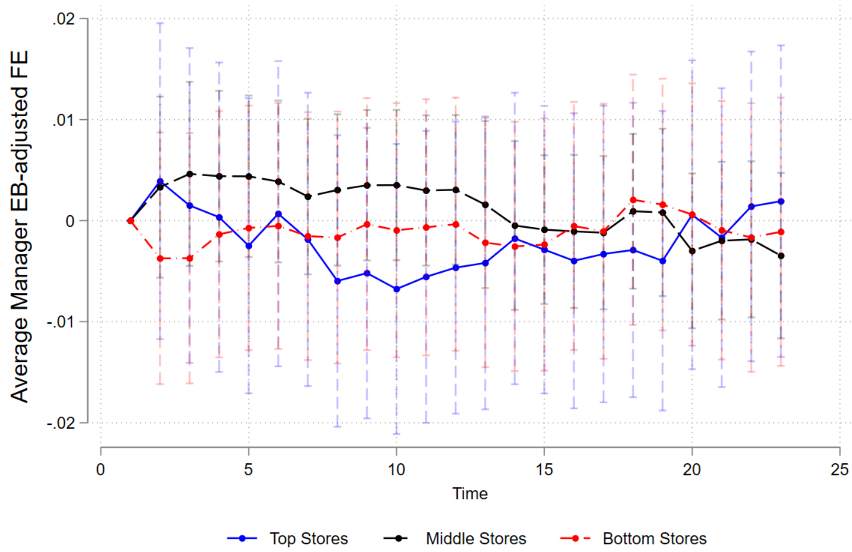
and similarly for $\mathbb{E}[\mu_m|\theta]$. Recall that $\mathbb{E}[\mu_s] = \mathbb{E}[\mu_m] = 0$; and because the selection rule depends only on the comparison between θ and k, we can also set $\mathbb{E}[\theta] = 0$ without loss of generality (as k can always be rescaled).

Next, and again using the properties of the multivariate normal distribution, the conditional covariance is

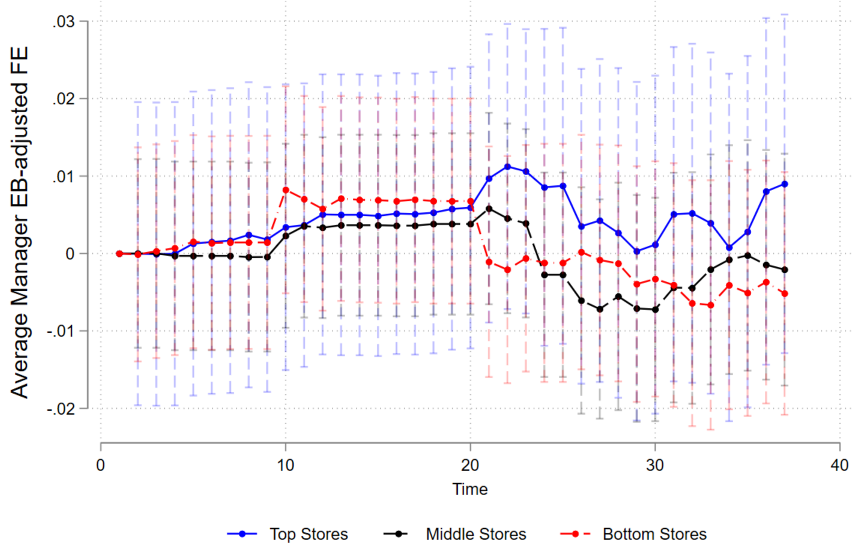
$$\mathrm{Cov}(\mu_s, \mu_m|\theta) = \mathrm{Cov}(\mu_s, \mu_m) - \frac{\mathrm{Cov}(\mu_s, \theta)\mathrm{Cov}(\theta, \mu_m)}{\mathrm{Var}(\theta)}.$$

[19]Note that the comparison $f(\mu_s, \mu_m) < k$ is also more general than might first look. Because the function f assumes real values, the selection rule must necessarily take the form $f(\mu_s, \mu_m) \in S$, for some set $S \subset \mathbb{R}$. Therefore, the implicit assumption made by our selection rule is that S is not the union of disjoint intervals; and because f can always be defined as a positive or negative function, we can simply pick $S = (-\infty, k]$.

Figure D.1: Evolution of the Average Manager FE by Store Quality Tercile



(a) Company A



(b) Company B

Note: Charts show the coefficients from regressing the manager quality (EB-adjusted fixed effect) on a time dummy and a connected set fixed effect. Regressions are run separately by store bin (top, middle, bottom), where bins are defined by tercile where each store is located within the store quality (EB-adjusted fixed effect) distribution. Dashed lines represent the 95% confidence interval constructed using robust standard error estimates.
The sharp changes in the average manager quality in company B are explained by the fact that almost all managerial changes in 2014 and 2015 occur in December (periods 9 and 21). Starting in 2016, managerial moves in other months become more frequent, with several smaller stores exiting the data in May of 2016 (see also Figure A.1).

Plugging these into the expression above gives

$$\begin{aligned}\text{Cov}(\mu_s,\mu_m|\theta<k) &= \frac{\text{Cov}(\mu_s,\theta)\text{Cov}(\mu_m,\theta)}{\text{Var}(\theta)^2}\text{Var}(\theta|\theta<k)+\text{Cov}(\mu_s,\mu_m)-\frac{\text{Cov}(\mu_s,\theta)\text{Cov}(\theta,\mu_m)}{\text{Var}(\theta)} \\ &= \underbrace{\text{Cov}(\mu_s,\mu_m)}_{\text{uncoditional covariance}} \underbrace{-\frac{\text{Cov}(\mu_s,\theta)\text{Cov}(\theta,\mu_m)}{\text{Var}(\theta)}\left[1-\frac{\text{Var}(\theta|\theta<k)}{\text{Var}(\theta)}\right]}_{\text{bias}}.\end{aligned}$$

This expression clarifies what we refer to as the "bias" in the main text: it is the difference between the unconditional covariance across manager and store fixed effects, and the truncated/conditional covariance that can be actually calculated by the researcher. In that sense, there is no bias if one is actually interested in estimating the truncated covariances, but that quantity may not reflect the relationships between the productivity of managers and stores in the population. The expression also shows that sign of the bias term depends only on the properties of the selection rule θ—specifically whether it is increasing/decreasing on (μ_s,μ_m) and the relative size of its truncated and unconditional variances. Under some regularity conditions, it is also possible to show that the bias term in the expression is negative. For example, if $f(\mu_s,\mu_m)$ is strictly increasing (or strictly decreasing) in both arguments, $\text{Cov}(\mu_s,\theta)\text{Cov}(\theta,\mu_m)>0$; if, in addition, $\text{Var}(\theta|\theta<k)<\text{Var}(\theta)$ (a statement that is true in most, but not all, cases[20]), then the bias term is negative.

In our specific example where $f(\mu_s,\mu_m)=\mu_s+\mu_m$, the expression above simplifies to

$$\text{Cov}(\mu_s,\mu_m|\theta<k)=\rho\sigma_s\sigma_m-\frac{(\sigma_s^2+\rho\sigma_s\sigma_m)(\sigma_m^2+\rho\sigma_s\sigma_m)}{\sigma_s^2+\sigma_m^2+2\rho\sigma_s\sigma_m}\left[\frac{k}{\sigma_\theta}\frac{\phi(k/\sigma_\theta)}{\Phi(k/\sigma_\theta)}+\left(\frac{\phi(k/\sigma_\theta)}{\Phi(k/\sigma_\theta)}\right)^2\right],$$

where $\sigma_\theta^2:=\sigma_s^2+\sigma_m^2+2\rho\sigma_s\sigma_m$, and we use the fact that

$$\text{Var}(\theta|\theta<k)=\sigma_\theta^2\left[1-\frac{k}{\sigma_\theta}\frac{\phi(k/\sigma_\theta)}{\Phi(k/\sigma_\theta)}-\left(\frac{\phi(k/\sigma_\theta)}{\Phi(k/\sigma_\theta)}\right)^2\right],$$

where ϕ and Φ are, respectively, the standard normal density and distribution. The expression for the truncated variance above follows from the fact that θ is also normally

[20]One counterexample is a distribution that takes values 100 with probability 0.9, -100 with probability 0.05, and -1000 with probability 0.05. In this case, truncating the distribution at, say, 0, clearly increases its variance, as the probability of an extreme value happening increases considerably.

distributed (since it is a linear combination of normally distributed variables). One can easily check that the bias term is indeed negative in this case.

D.2.1 Quantitative Exercise

The previous section establishes that observed (truncated) covariance between manager and store fixed effects is smaller than the unconditional (population) covariance, under the assumption that these fixed effects are jointly normally distributed. But how big is the difference/bias? Given that we now have expressions for each of the observed quantities in the data, we can perform a simple quantitative exercise to find out.

There are 4 parameters to calibrate in the expressions above: $\{\sigma_s, \sigma_m, \rho, k\}$. We use 4 data moments to do so: the observed variances of manager and store fixed effects, the observed covariance between those fixed effects, and the share of stores that switch managers, $P(\theta < k) = \Phi(k/\sigma_\theta)$. Since manager and store fixed effects are normalized within each connected set, we use the weighted average of variances and covariances computed in table 1 (the weights are the number of observations used to estimate each covariance).

While the truncated covariance is determined above, the variances of manager and store fixed effects are also affected by truncation. Using the law of total variance and joint-normality, we have

$$
\begin{aligned}
\operatorname{Var}(\mu_s|\theta < k) &= \mathbb{E}[\operatorname{Var}(\mu_s|\theta)|\theta < k] + \operatorname{Var}[\mathbb{E}(\mu_s|\theta)|\theta < k] \\
&= \mathbb{E}\left[\sigma_s^2 - \left(\frac{\sigma_s^2 + \rho\sigma_s\sigma_m}{\sigma_\theta}\right)^2 \middle| \theta < k\right] + \operatorname{Var}\left[\frac{\sigma_s^2 + \rho\sigma_s\sigma_m}{\sigma_\theta^2}\theta \middle| \theta < k\right] \\
&= \sigma_s^2 - \left(\frac{\sigma_s^2 + \rho\sigma_s\sigma_m}{\sigma_\theta}\right)^2 + \left(\frac{\sigma_s^2 + \rho\sigma_s\sigma_m}{\sigma_\theta^2}\right)^2 \times \sigma_\theta^2\left[1 - \frac{k}{\sigma_\theta}\frac{\phi(k/\sigma_\theta)}{\Phi(k/\sigma_\theta)} - \left(\frac{\phi(k/\sigma_\theta)}{\Phi(k/\sigma_\theta)}\right)^2\right] \\
&= \sigma_s^2 - \left(\frac{\sigma_s^2 + \rho\sigma_s\sigma_m}{\sigma_\theta}\right)^2\left[\frac{k}{\sigma_\theta}\frac{\phi(k/\sigma_\theta)}{\Phi(k/\sigma_\theta)} + \left(\frac{\phi(k/\sigma_\theta)}{\Phi(k/\sigma_\theta)}\right)^2\right],
\end{aligned}
$$

and similarly for $\operatorname{Var}(\mu_m|\theta < k)$, replacing σ_s^2 by σ_m^2.

The set of truncated variances and covariances are calculated from Table 1, while the share of stores that switch managers comes from Table B.1. This gives

$$\mathrm{Var}(\mu_s|\theta < k) = 0.85; \quad \mathrm{Var}(\mu_m|\theta < k) = 0.33; \quad \mathrm{Cov}(\mu_s, \mu_m|\theta < k) = -0.29; \quad \Phi(k/\sigma_\theta) = 0.53,$$

which we use to calibrate the parameters from our simple model by minimizing the sum of squared residuals between model and data. The values obtained from this process are

Parameter	σ_s	σ_m	ρ	k
Value	1.3	0.58	-0.34	0.09

Given that the truncated correlation from the data above is about -0.57, the bias amounts to approximately -0.23, or 2/3 of the value of the unconditional correlation.

These are some stark numbers, showing that the resulting biases in the observed covariance can be large. But they must be understood with some caveats. Joint normality, along with the specific selection rule we use, may be strong assumptions—though the same arguments likely hold in other scenarios. In addition, and perhaps more importantly, the exercise above is entirely based on comparing the conditional/truncated covariances with their unconditional counterparts. It does not allow, for example, for the broken pairs to be re-matched, which is what likely happens to most manager-store pairs (where the manager does not entirely leave the company). We address the latter issue with some simulation-based results in the following section.

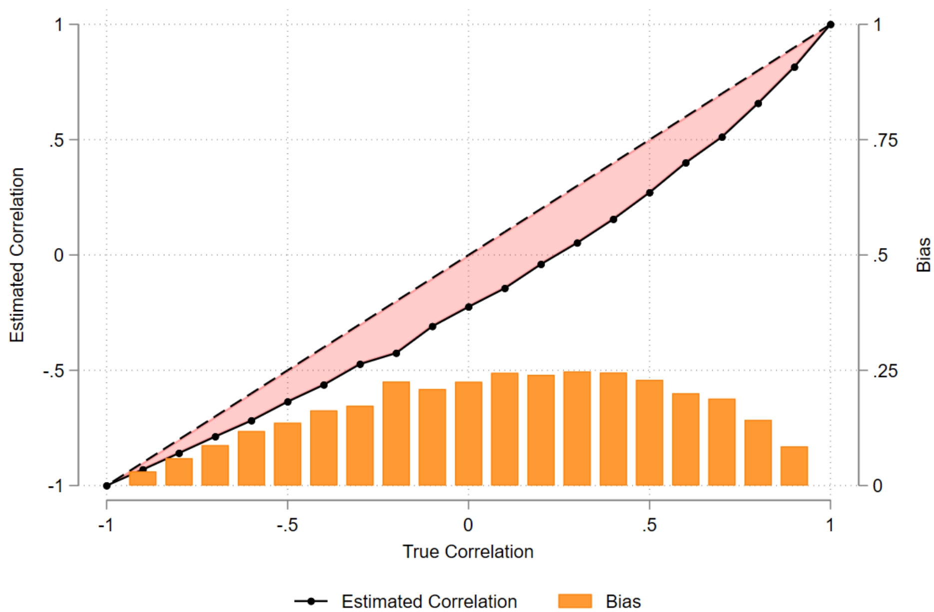
D.3 Simulation Exercises

Next, we discuss some simulation exercises that help to illustrate the behavior of the bias we identify. We start by generalizing the exercise in Figure 2 by allowing for an arbitrary "true" correlation between managers/stores in the initial match. Mechanically, we pick a correlation ρ and simulate 10,000 manager-store matches

$$\begin{pmatrix} \mu_m \\ \mu_s \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right).$$

Next, we select all stable matches $\mu_s + \mu_m \geq 0$ and drop them from the sample, as it would not be possible to estimate the quality on those pairs. It is useful to think of this step as a "first round" of matches. In a "second round," we break up the unstable matches (from round 1) and re-match them all according to same correlation ρ. Note that this may result in stable matches, as all managers are now paired with different stores. Combining the manager-store pairs from the first and second round results in what one might find in real data, consisting of several pairs of managers/stores, where each of them has been part of an unstable match at least once.

Figure D.2: Match Correlation and Size of Bias
(after two rounds of matches)



The chart shows the bias in the estimated correlation after simulating 10,000 pairs $\begin{pmatrix} \mu_m \\ \mu_s \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$ for a given "true" correlation ρ and allowing for two rounds of matching (see discussion in text). The correlation is estimated by stacking all pairs of managers and stores (after removing stable matches in round 1) and computing the correlation between the two stacked series.

The results of this simulation are shown in Figure D.2. Comparing the estimated and true correlation ρ, we see that the bias is always negative after two rounds, regardless of the value that ρ takes. This can be seen by noting that the estimated correlation is always below the 45-degree line, and the size of the bias is shown in the orange bars. The largest value of the bias occurs when the true correlation is positive but relatively low, at $\rho \approx 0.3$. In many cases it is large enough that the estimated correlation is negative even though the true correlation is positive, which could be an important contributing factor for the negative assortative matching results commonly found in the literature.

The only cases where the estimated correlation coincides with the true correlation are when $\rho = 1$ or $\rho = -1$. In those cases, μ_s is always a monotonic function of μ_s, so imposing $\mu_m + \mu_s \lesseqgtr k$ does not affect the correlation between the fixed effects (nor would re-shuffling them make any difference, as managers would always be matched with a stores with the same productivity level). In addition, note that selecting stable and unstable matches based on the rule we proposed can also bias the *variance* of the estimated manager and store fixed effects. This happens because the selection rule can restrict the range of the observed effects: for example, if the lowest manager productivity is $\underline{\mu_m}$, then any store $\mu_s > k - \underline{\mu_m}$ is necessarily in a stable match and its productivity never estimated. When this happens, the variance of the estimated store fixed effects is downwards biased as well.

What happens to the selection-based bias when we increase the size of the sample? A pure increase in the number of observations does not change the nature of the problem; for example, simulating more (μ_m, μ_s) in the exercise above does nothing to reduce the bias. But with more observations, one can also expect to find more movement of managers across stores. We again turn to our simulation: suppose that μ_m and μ_s are both i.i.d. $\mathcal{N}(0,1)$ and that the selection rule applies with $k = 0$. Now, however, we allow for several rounds of matching. We start with a "round 0," where 100,000 manager and store pairs are drawn from a standard normal distribution.[21] In this round, all stable matches are identified and removed from the data, as the productivities of those manager/stores cannot be separately identified. In each of the subsequent rounds, the following steps happen: (1) managers and stores that are not currently matches get randomly matched to another manager/store; (2) stable and unstable matches are identified using the threshold rule; and (3) unstable matches are broken, to be re-matched in the beginning of the next round.

This simulation is run for 20 rounds. In each round, we calculate the correlation between manager and store quality within currently stable matches, currently unstable matches, and across the full sample. To mimic what one might find in the data, all manager-store pairs formed in rounds 1-20 are kept in the sample. In this sense, the rounds can be thought of as the time periods in a panel, but in this case managers in unstable matches move to other stores in every period. The correlations are calculated at the end of step (2) in each round, before the unstable matches are broken. Note that a stable match is an absorbing state, as once a managers/store is part of stable match they will remain matched until the end of the simulation.

[21] We use a larger number of draws this time so the correlations remain stable throughout the exercise, as the number of pairs in stable and unstable matches changes.

This "absorbing" capacity of stable matches induces some interesting findings. Figure D.3 Panel (a) plots the share of stable matches[22] in each round in our exercise, showing a sharp increase in first 5-10 rounds, and leveling off after that. After these 10 rounds, the only managers and stores left are likely very unproductive, so they get stuck with unstable matches through the entire exercise. Figure D.3 Panel (b) plots the FE of managers and stores in stable and unstable matches after 20 rounds. Note that the distribution of FEs in stable matches looks a lot like the FE of managers and stores in stable matches in Figure 2, showing a negative correlation; however, the FEs of managers and stores in unstable matches are roughly uncorrelated, as managers and stores in unstable matches are randomly re-matched after each round.

Figure D.3 Panel (c) tracks the correlations between the FEs of managers and stores within stable and unstable matches over time. After two rounds, the correlations in Figure D.3 match the ones those in Figure D.2 when $\rho = 0$. Illustrating the mechanism described above, we see the correlation between managers and stores in stable matches quickly drop to about -0.65, remaining in that vicinity for the remainder of the exercise. The correlation between managers and stores in unstable matches also starts out negative, as shown in Figure 2, but quickly converges to the true correlation $\rho = 0$. As explained above, this is due to our assumption that unstable pairs are re-matched according to their true correlation.[23]

Interestingly, this generates a pattern where the correlation between the FEs of managers and stores in stable matches is negative; the same correlation for managers and stores in unstable matches is approximately zero; but the overall sample correlation is positive, as the two "groups" (stable and unstable pairs) become segregated in roughly the positive and negative quadrants in our plot.

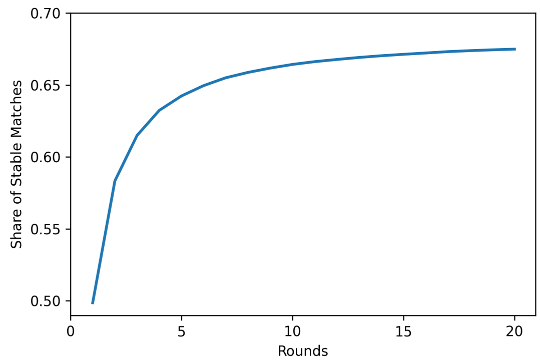
The *overall* correlation in the sample (considering all manager and store pairs each round) also starts out negative, but steadily increases and even turns positive after about five rounds! This is a very surprising result, given that the correlation between the FEs of managers and stores in stable *and* unstable matches is negative for most rounds in our exercise. The explanation, which can be though of as a version of Simpson's Paradox, is that after enough rounds the managers and stores in stable and unstable matches be-

[22]This share is calculated out of all matches, excluding the original stable matches from round 0.

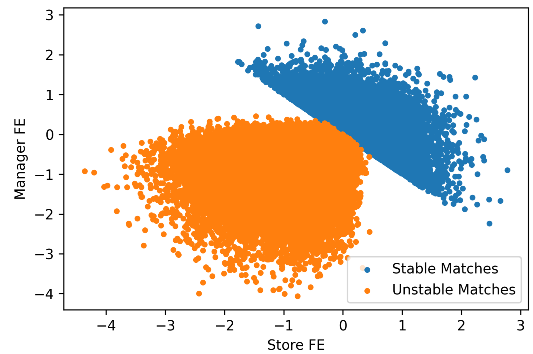
[23]In this very simple example, one could estimate the "true" correlation between manager and store FEs by looking only at the pairs formed after a manager moves. However, this result could easily change in more realistic scenarios, where managers may have other reasons to move across stores, could get fired for reasons not tied to performance, retire, or otherwise leave the company (also note that not all managers move stores at the same time, as is assumed in our example).

come segregated in roughly the positive and negative quadrants in our plot, as shown in Panel (b). This leads to within-group correlation that negative or approximately zero, but an overall correlation that is positive. Importantly, this exercise shows that even after multiple rounds of moves (many more than are likely to be found most datasets[24]), the correlation between *observed* fixed effects does not converge to the “true" value of that parameter.

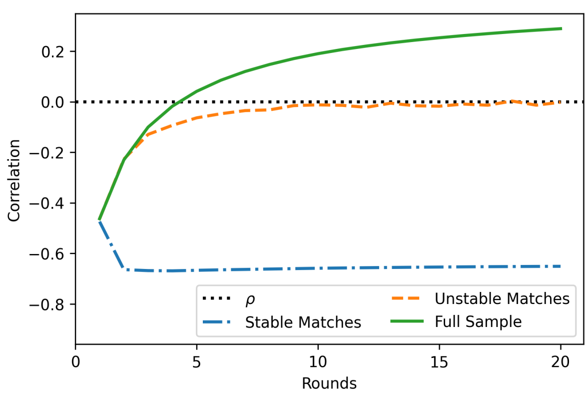
Figure D.3: Simulating Selection-Based NAM



(a) Share of Stores in Stable Matches



(b) Manager-Store Pairs After 20 Rounds



(c) Correlation Between Manager and Store FE

Panel (a) shows the share of stores that are in a stable match in each round. Panel (b) shows the manager-store pairs obtained after 20 rounds of the simulation. Panel (c) shows the correlation between manager and store quality within each round of matching. "stable" and "Unstable" matches are defined using the selection rule $\mu_s + \mu_m > 0$ for stable matches; ρ indicates the true correlation in the simulated data.

Lastly, we add that the selection-based NAM is also different from the endogeneity bias that would arise if managers choose to move to a store because of unobservable characteristics of that store. However, it possible that a selection rule *causes* endogeneity:

[24]Recall from section 2 that, in our data, about 90% of the stores that change their managers do so once or twice only over the course of three years.

for example, if companies switch managers whenever productivity dips below a certain threshold (after adjusting for a common time effect), it could be the case that negative shocks to productivity result in an otherwise stable match being broken. Furthermore, for a given store quality level, the match is more “resilient" to shocks if the manager’s quality is higher—causing a correlation between the manager fixed effect and the probability that a move happens. We argue that this is a lesser concern in practice. It is unlikely that a managerial change is triggered by a single bad shock, or else we would expect to see many more changes in the data. As mentioned in the previous subsection, many managerial movements are driven by staffing needs and other concerns uncorrelated with transitory shocks.

The discussion above uses deterministic threshold rule for ease of exposition, but the logic here applies equally well to cases where match stability is a stochastic, increasing function of both manager and store types. Moreover, the function need not be linear; any function where stability is increasing in both manager and store types poses the same selection-based downward biases in variances and covariance.

In conclusion, the selection-based NAM we describe here is fundamentally different from previous biases discussed in the literature. In particular, it is not resolved by increasing the size of the dataset. Moreover, it potentially applies in many situations that rely on movers to conduct such two-way fixed effect decompositions. Therefore the issues we discuss here likely transfer to other research settings past and present. For example, in a traditional AKM model where wages are a function of worker and firm effects (because unmeasured productivity increases in both), the match between a worker and an establishment is plausibly more stable when the pair’s joint productivity (and wage) is higher. Applying the same logic above, this will tend to create a negative correlation between worker and firm effects among movers, potentially creating biases in the measured variance and covariance of worker and firm effects. To the best of our knowledge, this issue has not been discussed elsewhere.

D.4 Selection Rule and Endogeneity

Finally, we discuss whether the selection rule we describe may act as mechanism that generates endogeneity in our estimates. Both the event-study and AKM estimation approaches require that managerial moves are uncorrelated with unobserved factors that may affect productivity. Strictly speaking, our simple selection rule has only three things that affect whether a manager moves (i.e., is in an unstable match): their own productivity, their store's productivity, and the value of the threshold. The threshold is fixed and therefore uncorrelated with the manager's FE. The manager and store FEs (productivities) are included as controls, so there should not be an issue with those either. Whatever causal effect the manager might have on the productivity of the stores they move to or from should be consistently estimated.

One way to summarize the bias we describe here is to recognize that the selection problem affects estimates of the second moments of the joint manager-store type distribution, but not its first moments. It is true that selection-driven truncation affects the average levels of manager and store types among movers, but because all types are measured relative to each other, this shift in the absolute value of their average does not affect measured values.

Stepping outside of our framework, store-level productivity might be affected by shocks in practice. Let us introduce a shock ε_{st} and suppose a manager moves to a different store as soon as $\mu_m + \mu_s + \varepsilon_{st} < k$. Endogeneity in the model happens if $\mathbb{E}[\mu_m \varepsilon_{st}] \neq 0$; this could happen if, for example, the probability of a manager moving away depends on the shock. Here, that probability is $P(move) = P(\varepsilon_{st} < k - [\mu_\mathbf{m} + \mu_\mathbf{s}])$. One important thing to note is that this is not a function of μ_m alone, but rather the *joint productivity* of the manager and store. Because of this, whether there is potential endogeneity depends on how managers and stores are distributed. We divide the discussion into cases.

Case 1: Suppose that manager and store productivity levels are uncorrelated. If a shock ε_{st} brings the pair $\mu_m + \mu_s$ below the threshold k, it also brings the pair $(\mu_m + \eta) + (\mu_s - \eta)$ below the same threshold. Because manager and store productivities are uncorrelated, these two pairs are just as likely to happen. This means managers with different productivities will both move (behave the same way) after the shock, as long as their joint productivity is the same distance of the threshold. It also means that the probability of moving is not a function of the manager's productivity.

Case 2: Suppose there is positive assortative matching (PAM) between managers and stores. Using the example above, we would now expect to see managers with lower productivity moving more frequently, as the pair $\mu_m + \mu_s$ is matched with a higher probability than the pair $(\mu_m + \eta) + (\mu_s - \eta)$. In this case, the manager with higher productivity is less likely to move because they are more likely to be paired to a higher productivity store.

Case 3: Suppose there is negative assortative matching (NAM) between managers and stores. Continuing with the same example, the probability that moves are a function of the manager's productivity depends on how strong the NAM is. Under NAM, the manager-store pair $\mu_m+\mu_s$ could be more or less likely than the pair $(\mu_m+\eta)+(\mu_s-\eta)$. If both pairs are just as likely, we are back to case 1. If $\mu_m+\mu_s$ is more likely, we are back to case 2. If $\mu_m+\mu_s$ is less likely, we can have endogeneity, but the bias goes in the opposite direction.

We should note that as we argued in section 3, managerial moves based on short-term shocks are unlikely. The decision to move/stay in a store is seldom made by manager alone. Staffing rules and company needs that are uncorrelated with productivity shocks also play an important, if not the most important, role in those decisions. And of course if managerial moves only happen after considering long term trends, the shock should not affect that decision, as $\mathbb{E}_t[\mu_m+\mu_s+\varepsilon_{st}]=\mu_m+\mu_s$.

E Correcting for Measurement Error

This section presents further details on our steps to address measurement error in the estimation of fixed effects. It first presents more details on grouped fixed effects and compares the results with out baseline results. Next, it discusses some potential issues and extensions of the Empirical Bayes adjustment and how it could affect our findings.

E.1 Grouped Fixed Effects

One approach to address 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, Lamadon and Manresa (2022). 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 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 G 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. We use 20 groups for Company A and 50 groups for Company B, so that each groups has, on average, about 50 stores. Varying the number of groups does not meaningfully affect our results.

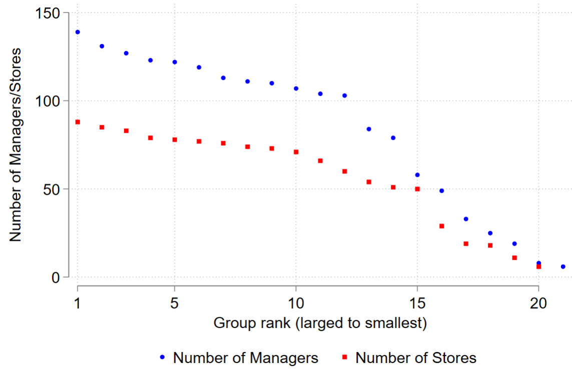
The key issue that arises in this step is the choice of variables used to cluster the units in our data. Ideally, all of the relevant heterogeneity across managers and stores would be captured by a few observable variables, so that there is little loss of information in the clustering process. We cluster 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 instead use their average log-tenure, along with the most common store format for each manager.

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.

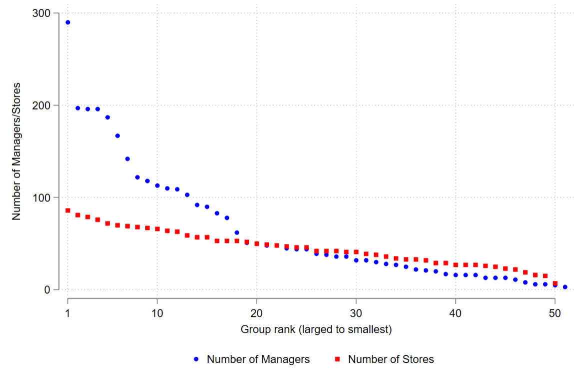
Step 2 Once groups are defined, we repeat the estimation of equation 1, with the exception that the effects on the right hand side no longer identify individual managers or stores, but rather the group to which they belong. Figure E.1 shows the number of managers and stores in each group, showing that the largest groups in this exercise can contain a significant number of managers in both companies.

Figure E.2 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. Looking at Figure E.2, we see that the distribution of the de-meaned grouped fixed effects has most of its mass within a smaller range when compared to the de-meaned "regular" fixed effects (figures A.2 and A.3). This indicates a lower heterogeneity across managers/stores, which could the result of the clustering in the first stage. The shape of the distributions all cases, however, is relatively similar.

Figure E.1: Number of Managers and Stores in Each Group

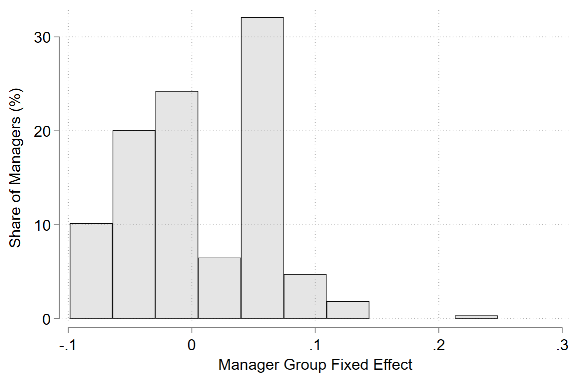


(a) Company A

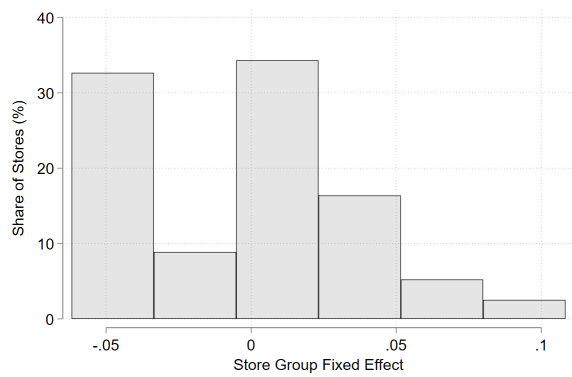


(b) Company B

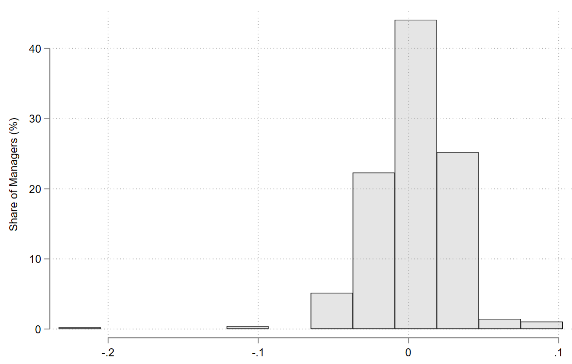
Figure E.2: Distribution of Grouped Fixed Effects



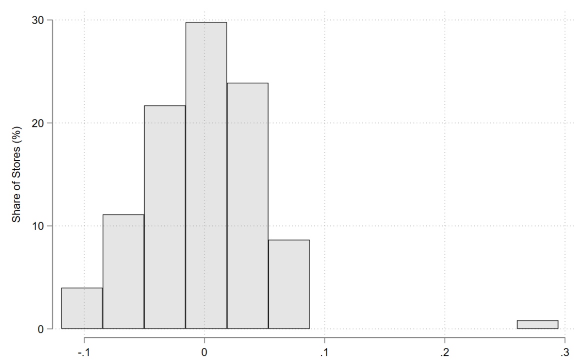
(a) Manager Fixed Effect (Company A)



(b) Store Fixed Effect (Company A)



(c) Manager Fixed Effect (Company B)



(d) Store Fixed Effect (Company B)

Lastly, Figure E.3 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 equation 1), EB-adjusted, and grouped. As seen in the figure, the unadjusted and EB-adjusted FEs tend to be very similar, though with a noticeable difference for managers in Company A. The slope of the relationship between manager and store fixed effects and productivity is a bit smaller after the EB-adjustment, as it entails some attenuation towards the connected set mean.

The grouped fixed effects also have a flatter relationship with productivity when compared to the unadjusted FEs, again indicating some regression to the mean (as the effects are constant within groups). In company A the group fixed effect estimates are relatively close to the EB-adjusted FEs, for both managers and stores. In company B, they have a much weaker relationship with productivity. This is likely an indication that the variables used to cluster manager/stores in the first stage for company B are not sufficiently rich to capture their heterogeneity, leading to potentially flawed results (recall that wages are not observed in company B, so we must rely on tenure to cluster managers). For this reason, we focus on the EB adjustment throughout the paper, as it produces fixed effects that are more reliably related to productivity.

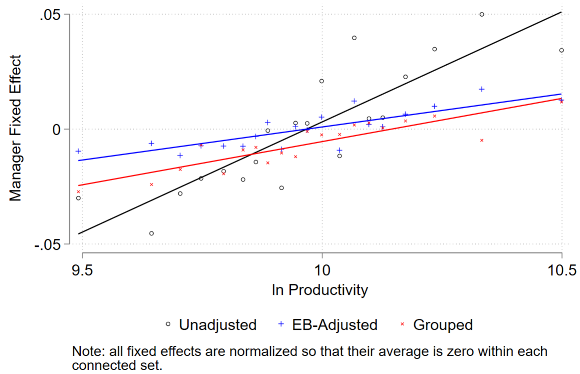
E.2 Robustness: Empirical Bayes Adjustment

This section discusses some remaining potential issues on the EB adjustment procedure, including the correlation between the adjusted and unadjusted estimates, and the potential impact of breaking some of the assumptions in the method.

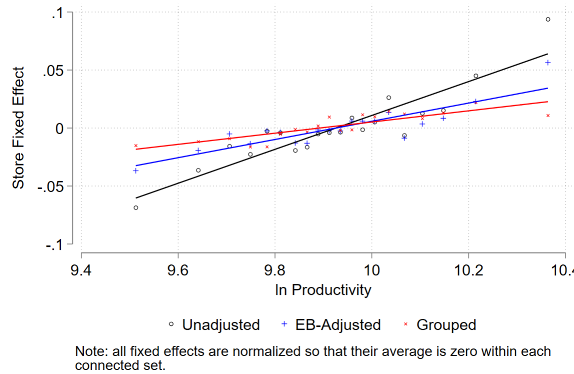
Adjusted vs Unadjusted Estimates. As shown in Figure E.3, the EB-adjustment does attenuate the relationship manager/store fixed effects and log productivity. However, the correlation between adjusted and unadjusted effects is still extremely high—close to 0.99 in all cases.

Mechanically, there are two ways that the EB-adjusted fixed effect ends up close to the initial estimates: either the weight $\hat{\kappa}_i$ is close to zero, or the averages $\hat{\lambda}_{c(i)}$ are close to $\hat{\mu}_i$. In our case, it is the latter. Almost all of the variation in $\hat{\mu}_i$ can be explained if one regresses it on a set of indicators for each connected set. What this means in words is that the variation in productivity of managers/stores *within* connected sets is significantly smaller than the variation in the productivity of managers/stores *across* connected sets.

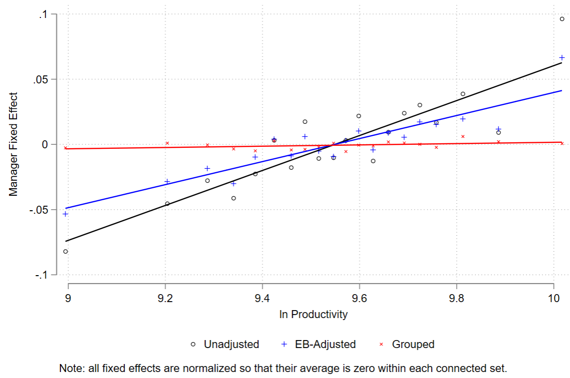
Figure E.3: Unadjusted, EB-Adjusted, and Grouped Fixed Effects



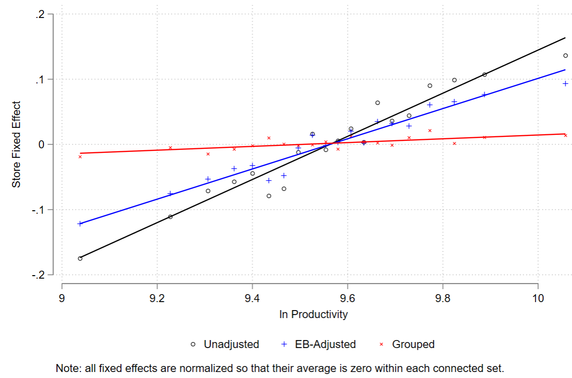
(a) Manager Fixed Effect (Company A)



(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.

This is to be expected, as each connected set is normalized by a different unit.[25] As a result, even if the EB adjustment attenuates $\hat{\mu}_i$ towards the connected set mean $\hat{\lambda}_{c(i)}$, the EB-adjusted FE will still be close to the original estimate.

In terms of the attenuation itself, it depends on the relative size of the variances $\hat{\tau}_i^2$ and $\hat{\sigma}^2$. The claim that the FEs are relatively close to the connected set mean implies that $\hat{\sigma}^2$ is relatively small, as it is identified by the variation of $\hat{\mu}_i$ around $\hat{\lambda}_{c(i)}$. However, the variance of the measurement error $\hat{\tau}_i^2$ (estimated using the variance of the FE estimator, corrected for limited mobility bias) is also small. This is identified by variation in the productivity within unit i over time. In our case, the average $\hat{\tau}_i^2$ is close to $\hat{\sigma}^2$, which means that the EB-adjusted effect lies about halfway between $\hat{\mu}_i$ and $\hat{\lambda}_{c(i)}$.

[25] Note that this does not necessarily mean that, say, stores in one connected set are radically different from stores in another connected set. It could simply mean that one set was normalized relative to a manager and another set was normalized relative to store, and the productivity effects of managers and stores may have a very different levels.

Another thing potentially limiting measurement error in our case is that our productivity metrics come from each company's own administrative data. They have a lot of incentive to measure revenue and employment well. Plus the same features that make our within-company comparisons useful for partially separating management practices from managers—e.g., common pricing and product mix across stores, etc.—might also limit the amount of noise in the estimation.

Correlated Measurement Error. One important assumption in the EB adjustment procedure is that the estimated fixed effects are is independently distributed. However, it is plausible that, for example, the measurement error of the fixed effects of managers who worked at the same store are correlated. Unfortunately, we are not aware of a straightforward way to do the EB adjustment procedure when the independence assumption is relaxed.[26] Nevertheless, to attenuate this concern, we calculate the full variance-covariance matrix of the fixed effects across all managers to show that the covariances of the FEs are significantly and consistently lower than the variance of the fixed effects ($\hat{\tau}_i^2$), and would thus have very little impact on the adjustment weights $\hat{\kappa}_i$.

Let V be the variance-covariance matrix of the sequence of estimated FEs $\{\hat{\mu}_m\}_m$. The diagonal entries in this matrix are $V_{ii} = \hat{\tau}_i^2$, and the off-diagonal terms are the covariances. To compare the size of the variance and covariances in V, we compute the matrix $U = V - (\hat{\tau}_1^2, \ldots, \hat{\tau}_M^2) \cdot I$ (where I is an identity matrix, so the resulting U is simply the matrix V after replacing its diagonal terms by zero). We would like to compare the size of the typical covariance of the measurement error across managers with the variance $\hat{\tau}_i^2$; but because the majority of the entries in the matrix U are zero (e.g., when managers are in different connected sets), simply taking averages may not yield the desired result.

Thus, for each manager m, we select the 90th and 99th percentiles of the sequence $\{|U_{mi}|\}_i$ (note that absolute values are used because covariances can be negative) and calculate the ratios $r_m^X = pct_X(\{|U_{mi}|\}_i)/\hat{\tau}_m^2$, where pct_X indicates the Xth percentile in the sequence. Averaging these ratios, we find that the mean of r_m^{90} stands at about 0.005, with a standard deviation of 0.112. Increasing the percentile to 99 changes the average ratio r_m^{99} to about 0.021, with a standard deviation of 0.568. In short, the size of the covariance of the measurement error across managers is consistently and significantly smaller than the size of the variance of the measurement error. As a result, a generalized version

[26] The appendix in Chandra et al. (2016) discusses the possibility that different measures of quality may be correlated. In our case, this would mean that the measurement error of a manager and their matching store are related, but not that the error of managers that went through the same store are correlated.

of the EB-adjustment that takes into account an arbitrary correlation across managers is unlikely to have an important impact in our application, as the impact of the variances dominate that of the covariances.

F 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.

Our estimates are shown in Table F.1, and suggest that Company A could have increased their revenue by 0.65 percent per month during our sample (the equivalent of almost USD 700,000 each month!) simply by optimally allocating their managers within the observed connected sets—which means that the unconstrained reallocation could be even more beneficial. In Company B the gains are even larger: 1.83 percent gains per month, or about GBP 4 million. In each case, our results indicated that are substantial gains from reallocating the best managers to the best stores within the company.

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

	Company A		Company B	
	Estimate	95% CI	Estimate	95% CI
Percent	0.65	[0.51, 0.78]	1.83	[1.53, 2.13]
Value (millions)	0.68	[0.54, 0.83]	3.97	[3.32, 4.62]

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

G Event Study: Pre-Trends and Robustness Checks

Pre-trends. To test for the existence of pre-trends, Figures G.1 and G.2 present the coefficients from model (3) in the 12 to 7 months before the managerial change (given that our event study *begins* 6 months before the change in manager). We find that none of the coefficients are statistically different from zero at the usual confidence levels, nor that there is a clear trend in any of the outcomes we analyze, suggesting that stores were in a common trend before the change in manager became apparent. One potential exception is log energy consumption for company B, which does appear to be in a slightly increasing path. However, none of the coefficients are statistically significant, and we do not see any meaningful effects on energy consumption as the store manager changes, which makes this issue less important.

Alternative Event Study Specification. To test whether the specification of the event study is relevant for the results found in the paper, we run the alternative model

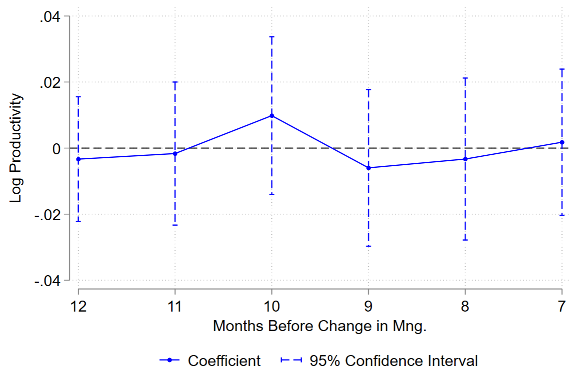
$$y_{st} = \alpha_s + \alpha_t + \mathbf{1}\{s \in \mathbb{T}(t)\} \sum_{k=0}^{k=4} \beta_k \mathbf{1}\{K_{st} = k\} + \varepsilon_{st}, \tag{4}$$

which is identical to equation (3), but where the estimated treatment effects start when the new manager comes into the store. As explained in the main text, this can be important because it changes the set of untreated stores, which now includes all stores that haven't switched managers yet (regardless of when the change happens). The results are shown in Figures G.3 and G.4; for comparison with the findings in the main text, we also plot the "pre-trends" in this alternative specification, which now consist of the 6 months that lead up to the managerial switch (note that in this case we don't expect the 'no pretrend' condition to hold). While not identical to the findings in out baseline model (due to the issues described above), the results here are very similar.

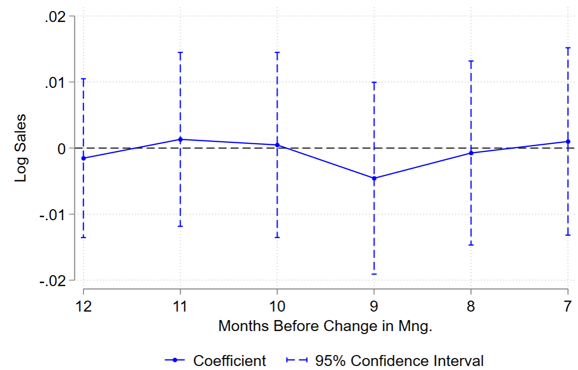
Managers who Stay, and Managers who Leave. As mentioned in section 4, there are many reasons why a manager could leave a store. Among those, managers who leave a store but remain within the company (that is, move to another store) might be very different from those that leave the Company altogether (are fired, move to a different job). To test this assertion, we run event study (3) again after splitting the treated sample into stores whose manager moves to another store within the company and stores whose manager leaves altogether (we disregard managers that leave the company after a tenure longer than 20 years to account for retirees).

Figures G.5 and G.6 present the results, with red lines indicating stores whose manager left altogether, and blue lines indicating stores whose manager simple switched. All in all, we find some patterns that are similar to the baseline specification, though the negative effects we find before (lower sales and productivity before the manager leaves) are attenuated for stores whose manager simply moves within the company. In company A the sharp drop on wages almost disappears for managers that stay within the company (although $\hat{\beta}_0$ is still negative and smaller than $\hat{\beta}_{-1}$ for log salary). These effects are sensible, as managers that stay within the company continue to have an incentive to perform even if they know they might leave a store; managers that know they are leaving altogether may not have the same incentives.

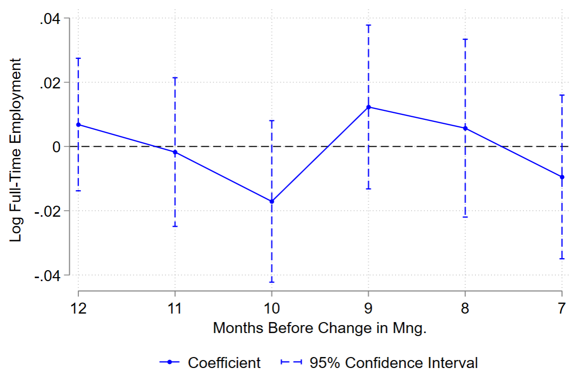
Figure G.1: Event Studies: Pre-Trends
Company A



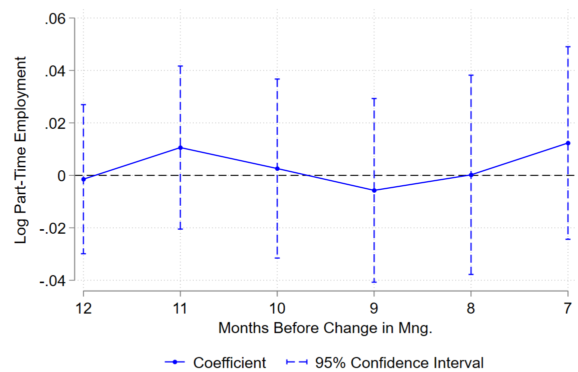
(a) Productivity (log)



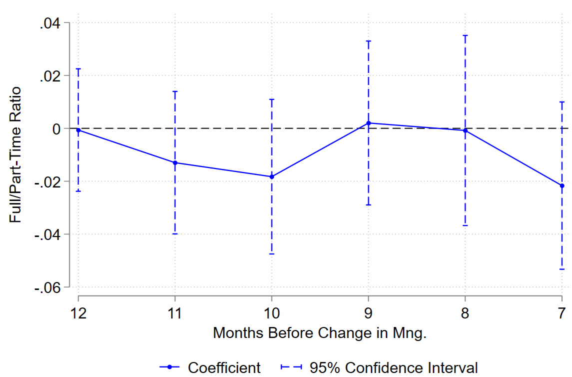
(b) Sales (log)



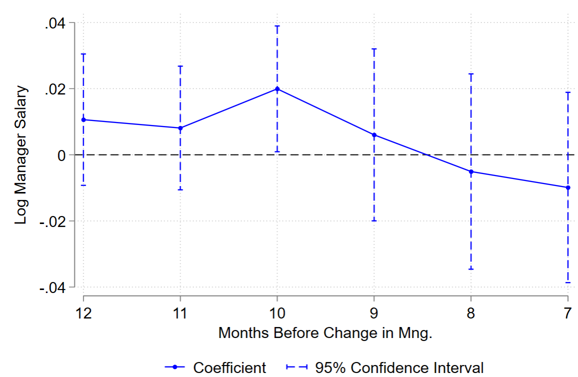
(c) Full-Time Employment (log)



(d) Part-Time Employment (log)



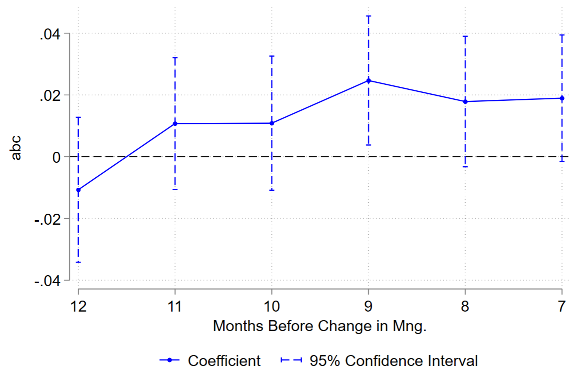
(e) Full/Part-Time Employment (ratio)



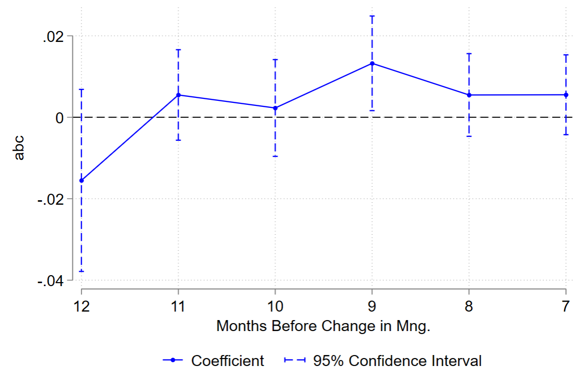
(f) Manager Salary (log)

Note: Dots show the value of each coefficient β_k while whiskers indicate the 95% confidence interval constructed with standard errors that are clustered at the store level (estimated following Borusyak, Jaravel and Spiess, 2024). Sample includes all managerial changes in the data.

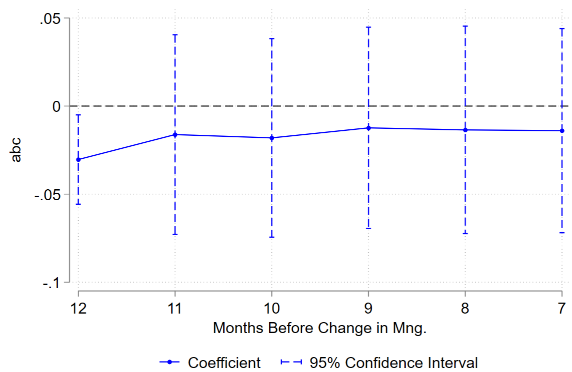
Figure G.2: Event Studies: Pre-Trends
Company B



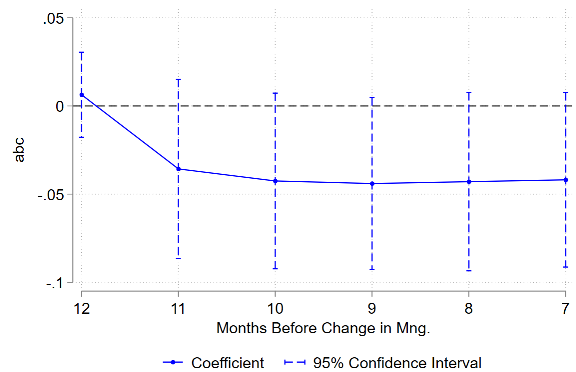
(a) Productivity (log)



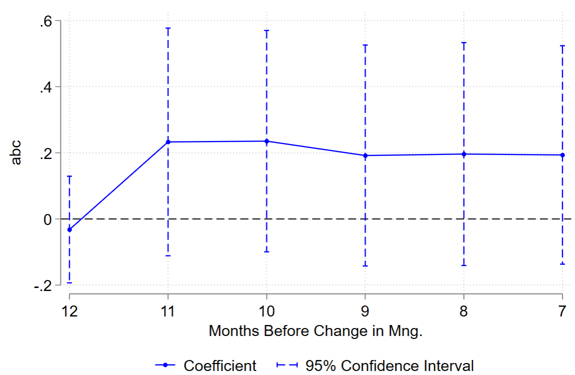
(b) Sales (log)



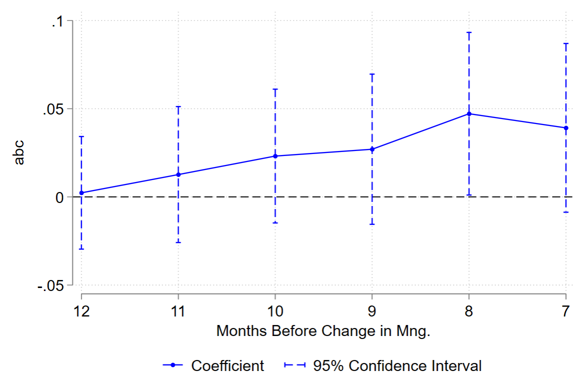
(c) Full-Time Employment (log)



(d) Part-Time Employment (log)



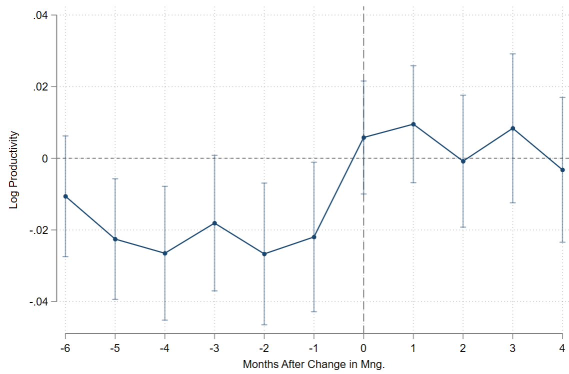
(e) Full/Part-Time Employment (ratio)



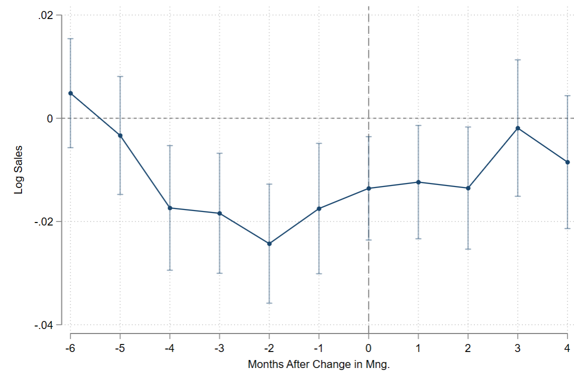
(f) Energy Consumption (log)

Note: Dots show the value of each coefficient β_k while whiskers indicate the 95% confidence interval constructed with standard errors that are clustered at the store level (estimated following Borusyak, Jaravel and Spiess, 2024). Sample includes all managerial changes in the data.

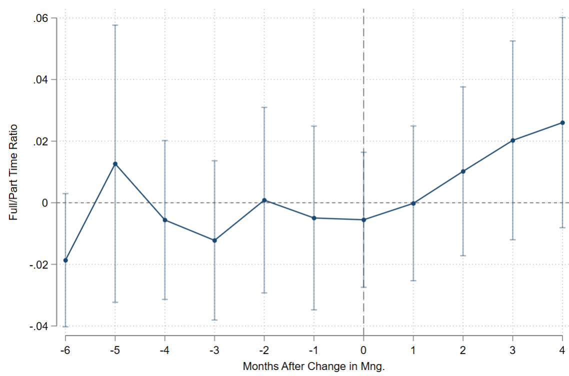
Figure G.3: Event Studies: Alternative Specification
Company A



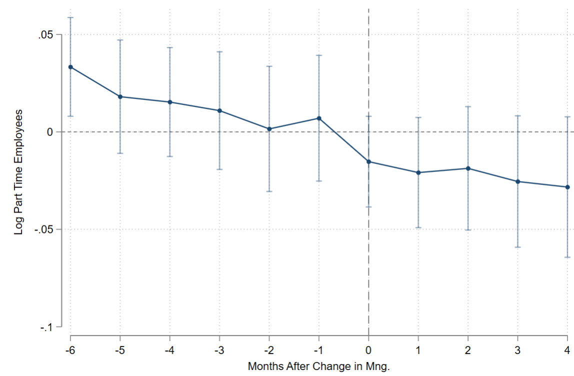
(a) Productivity (log)



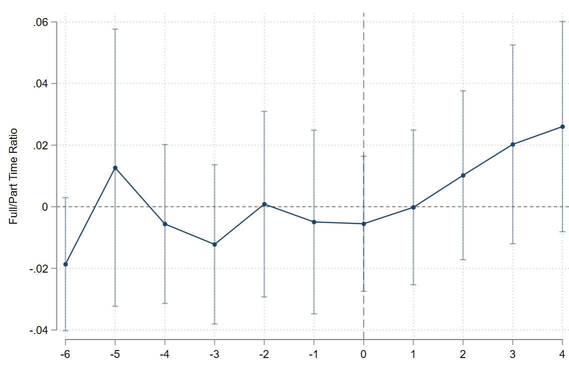
(b) Sales (log)



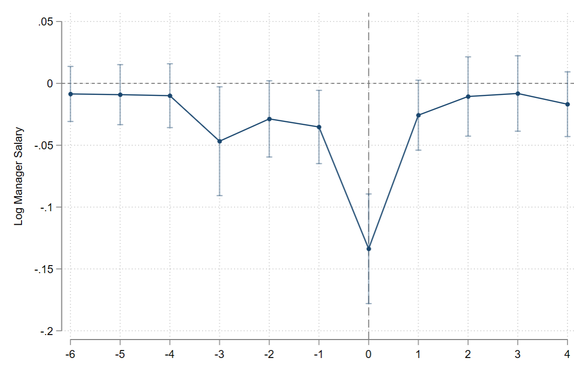
(c) Full-Time Employment (log)



(d) Part-Time Employment (log)



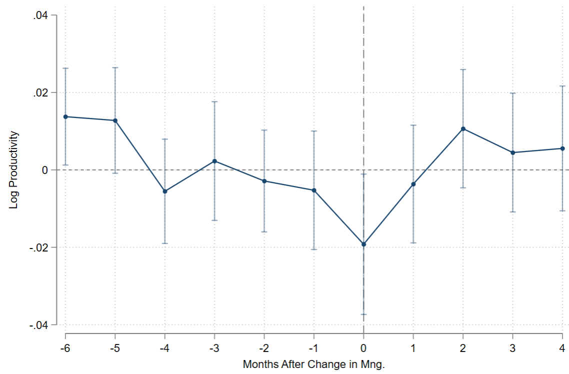
(e) Full/Part-Time Employment (ratio)



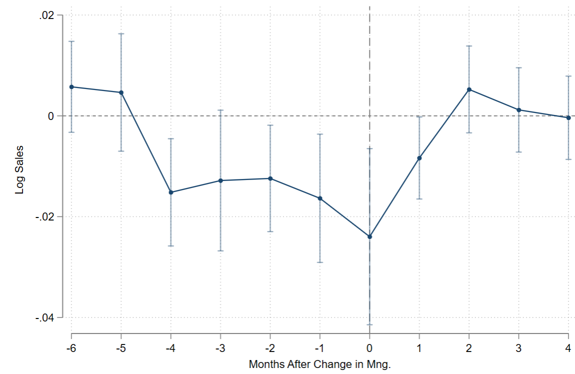
(f) Manager Salary (log)

Note: Dots show the value of each coefficient β_k while whiskers indicate the 95% confidence interval constructed with standard errors that are clustered at the store level (estimated following Borusyak, Jaravel and Spiess, 2024). In this specification, the "event" is defined as the month when the manager moves (unlike in Figure 5, when the event was define as the 6th month before the move happened). Because of this, the 'pre-trends' on the event study capture the impact that the managerial move has on stores on the months leading to the move.

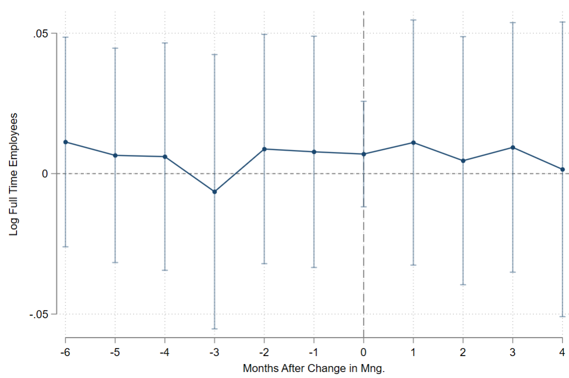
Figure G.4: Event Studies: Alternative Specification
Company B



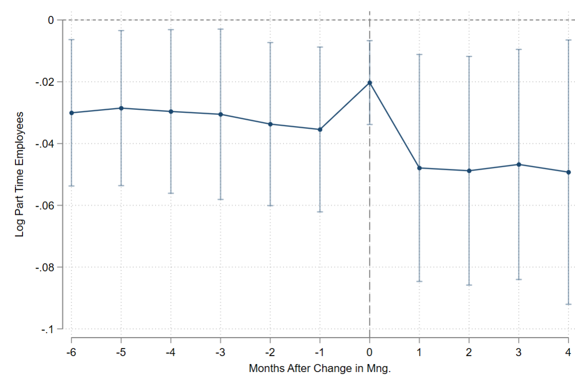
(a) Productivity (log)



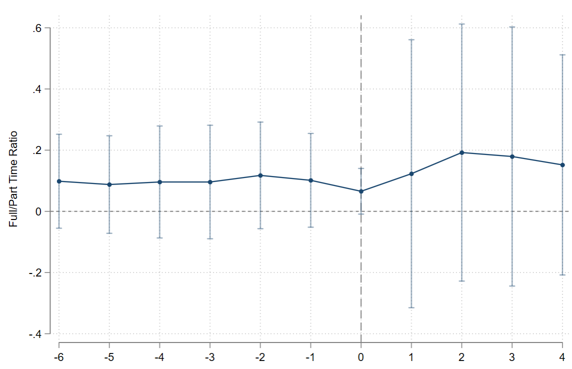
(b) Sales (log)



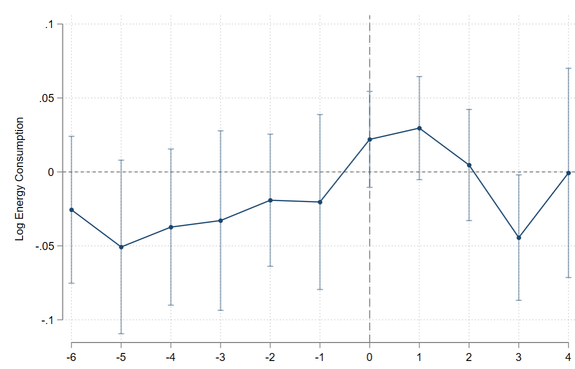
(c) Full-Time Employment (log)



(d) Part-Time Employment (log)



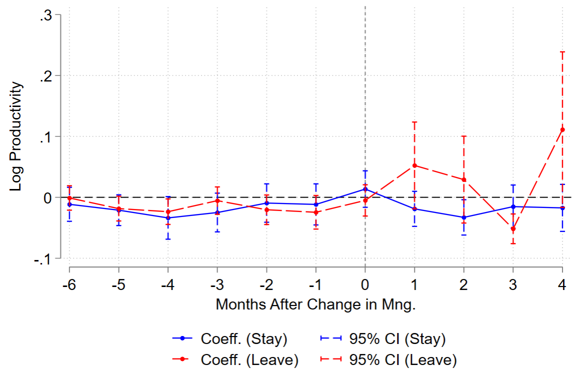
(e) Full/Part-Time Employment (ratio)



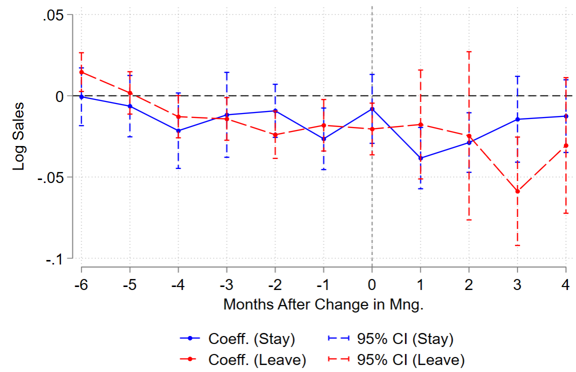
(f) Energy Consumption (log)

Note: Dots show the value of each coefficient β_k while whiskers indicate the 95% confidence interval constructed with standard errors that are clustered at the store level (estimated following Borusyak, Jaravel and Spiess, 2024). In this specification, the "event" is defined as the month when the manager moves (unlike in Figure 6, when the event was define as the 6th month before the move happened). Because of this, the ’pre-trends’ on the event study capture the impact that the managerial move has on stores on the months leading to the move.

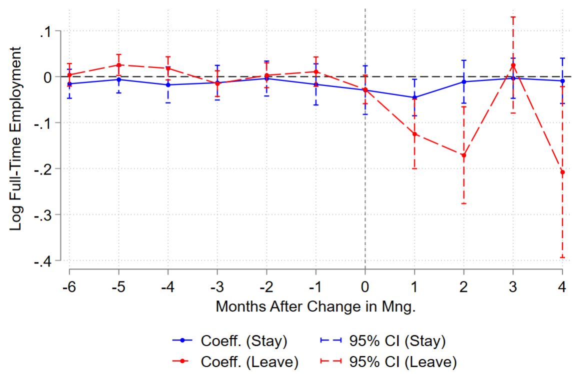
Figure G.5: Event Study: Managers that Stay vs Managers that Leave (Company A)



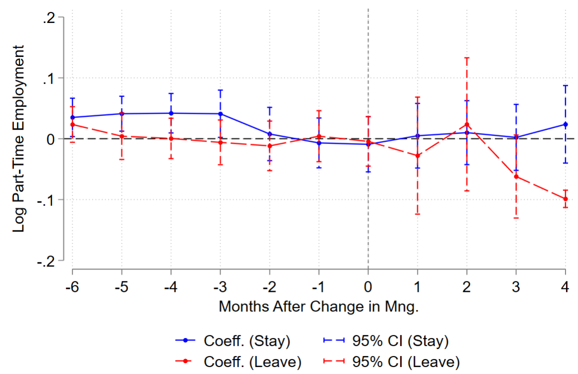
(a) Productivity (log)



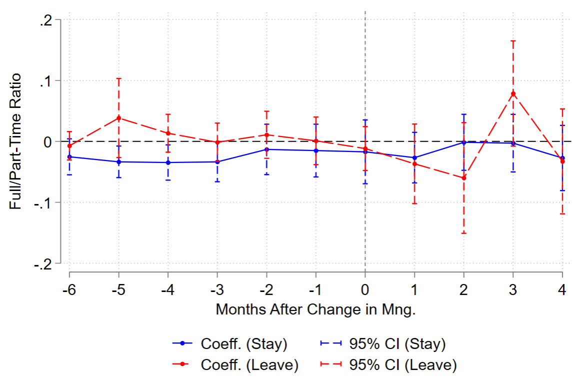
(b) Sales (log)



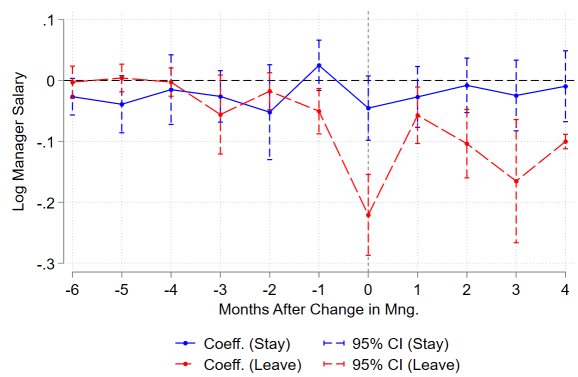
(c) Full-Time Employment (log)



(d) Part-Time Employment (log)



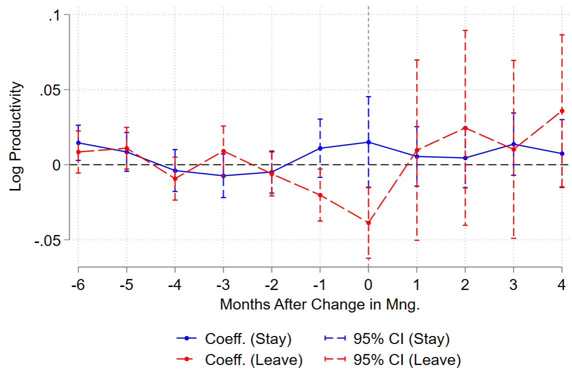
(e) Full/Part-Time Employment (ratio)



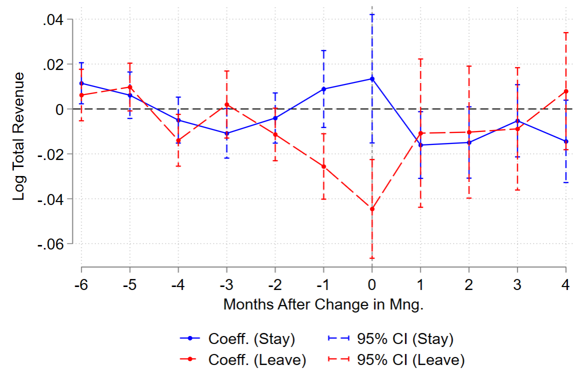
(f) Manager Salary (log)

Note: dots show the value of each coefficient β_k, while whiskers indicate the 95% confidence interval (both estimated following Borusyak, Jaravel and Spiess, 2024). The red line represents coefficients for managers that leave the company after the event; the blue line represents managers that move store but remain within the company.

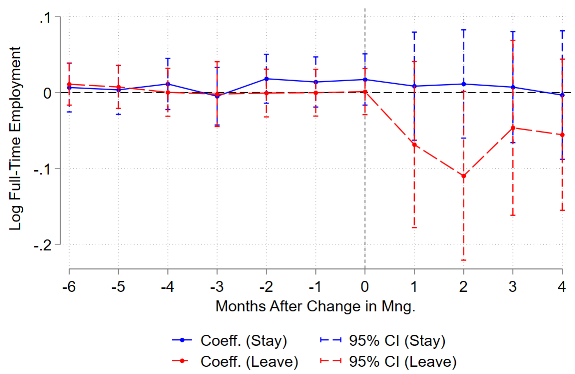
Figure G.6: Event Study: Managers that Stay vs Managers that Leave (Company B)



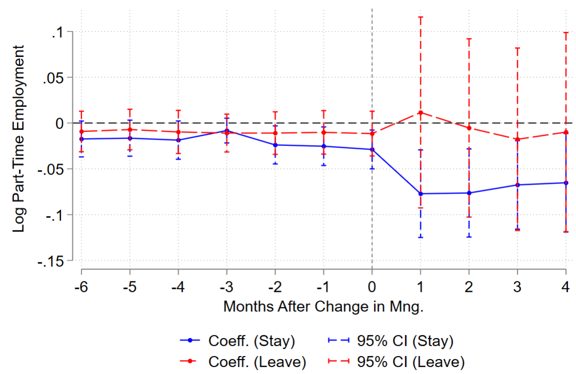
(a) Productivity (log)



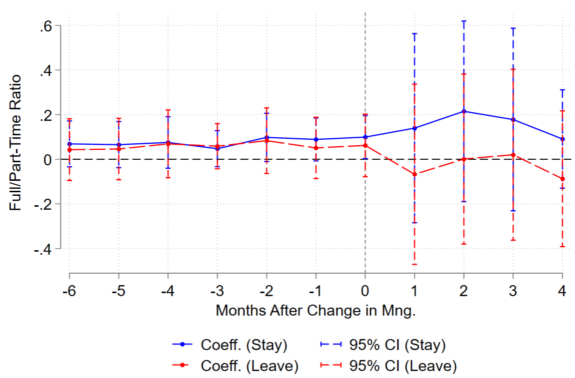
(b) Sales (log)



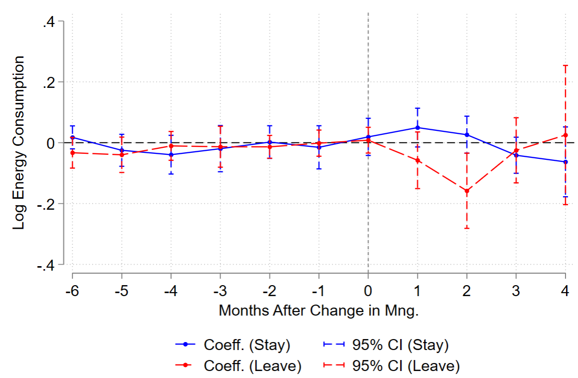
(c) Full-Time Employment (log)



(d) Part-Time Employment (log)



(e) Full/Part-Time Employment (ratio)



(f) Energy Consumption (log)

Note: dots show the value of each coefficient β_k, while whiskers indicate the 95% confidence interval (both estimated following Borusyak, Jaravel and Spiess, 2024). The red line represents coefficients for managers that leave the company after the event; the blue line represents managers that move store but remain within the company.