No Line Left Behind: Assortative Matching Inside the Firm*

Achyuta Adhvaryu Vittorio Bassi Anant Nyshadham Jorge Tamayo†

Preliminary and Incomplete

July 24, 2019

Abstract

How do workers and managers sort within the firm? Do the most talented managers match with the most productive workers, or with those who are struggling to perform? We characterize this sorting pattern in six factories of a large garment manufacturer in India. Workers in this firm are organized into production lines, each supervised by a manager. We exploit the high degree of worker mobility across lines, together with worker-level productivity data, to estimate the sorting of workers to managers. We find negative assortative matching (NAM) – that is, better managers tend to match with worse workers, and vice versa. This stands in contrast to our estimates of the production technology, which reveal that if the firm were to positively sort, productivity would increase by between 1-4%. Coupling these findings with a survey of managers and with data on buyers and orders, we find that NAM is strongest for factories where management is most worried about falling behind and not meeting deadlines with important buyers. That is, NAM arises, at least in part, because the value of buyer relationships imposes minimum productivity constraints on each production line. Our results emphasize that suppliers to the global market, concentrated in developing countries, may be beholden to a small set of powerful buyers from developed countries, and as a result, may be driven to "misallocate" managerial skill in service of these relationships, but at the expense of productivity. *JEL Codes: L14, L23, M12*

*We would like to thank Oriana Bandiera, Tzuo Hann Law, Andrea Ichino, Philippe Kircher, Claudio Labanca, Monica Morlacco, Chris Moser, Tommaso Porzio, Andrea Prat, Geert Ridder and numerous seminar and conference participants for helpful comments and discussions. We also thank Smit Gade and Varun Jagannath for help in conducting manager interviews. All errors are our own.

†Adhvaryu: University of Michigan, NBER & BREAD, adhvaryu@umich.edu; Bassi: University of Southern California, vbassi@usc.edu; Nyshadham: University of Michigan & NBER, nyshadha@umich.edu; Tamayo: Harvard Business School, jtamayo@hbs.edu.

1 Introduction

How are teams determined within the firm? Which team composition maximizes productivity, and what other concerns or constraints might intervene to determine team composition in practice? These questions regarding how production is organized within the firm are at the core of organizational and personnel economics (Lazear and Oyer, 2007; Lazear and Shaw, 2007). Despite a great deal of theoretical work describing these types of firm decisions and their implications for firm productivity and growth (Holmstrom and Tirole, 1989; Kremer, 1993), still little empirical evidence exists regarding how teams are composed and why due to stringent data requirements.

This paper leverages high frequency data on changing team composition within a firm and resulting productivity to answer these types of questions. Specifically, we ask how do workers match to managers? That is, do the most talented managers match with the most productive workers? Or do good managers match with workers who are struggling to perform? These questions link topics of interest to personnel and organizational economics to recent empirical work in the study of management and productivity in economics. Managerial quality strongly determines firm productivity and growth (Bloom et al., 2013, 2018; Bloom and Van Reenen, 2007, 2011; Karlan et al., 2015; McKenzie and Woodruff, 2013, 2016), and differences in managerial quality of firms explain a substantial portion of the vast productivity dispersion across rich and poor countries (Caselli, 2005; Hall and Jones, 1999).

Middle managers like the production line supervisors we study are often emphasized as enablers or constrainers of worker productivity (Adhvaryu et al., 2019; Levitt et al., 2013), particularly in low income countries and labor-intensive manufacturing settings (Blattman and Dercon, 2016; Bloom and Van Reenen, 2007; McKenzie and Woodruff, 2016). Further, considering information frictions in these labor markets (Adhvaryu et al., 2018; Bassi and Nansamba, 2019) lead to substantial misallocation of skill across firms (Syverson, 2011), understanding how best to utilize the managerial talent firms already possess is crucially important. However, which assignment of workers to managers maximizes productivity, and whether firms seem to be leveraging these interactions between manager skill and worker skill are open questions.

We characterize the pattern of matching between workers and middle managers in the context of six factories of a large readymade garment manufacturer in India. Workers in this firm are organized into production lines producing independent orders of similar products, each of which is supervised by a line manager. Worker mobility across lines is very high, both monthly – as orders and resulting needs change – and daily – as absenteeism leads to critical manpower shortages on some lines. This high frequency shuffling of workers across lines, along with granular worker-level productivity data, allows us to document the pattern of sorting of workers to managers. We find evidence of negative assortative matching (NAM) – that is, better managers tend to match with worse workers, and vice versa. On average across the six factories in our data, we find that the correlation between the estimated worker and manager fixed effects from a two-way fixed effects model in the spirit of Abowd et al. (1999), but estimated using productivity data rather than wage data, is -16%. We further show that these results are robust to correcting for limited mobility bias (Abowd et al., 2004; Andrews et al., 2008, 2012).

These results are notable in that they contrast with most matching patterns obtained from studies of the sorting of workers across firms (Abowd et al., 1999; Card et al., 2018, 2013; Eeckhout, 2018) and are inconsistent with hypothesized complementarity or imperfect substitutability between worker and manager skill, which should generate *positive* assortative matching (PAM) (Bandiera et al., 2007, 2009; Lazear et al., 2015). This begs the question whether the pattern of NAM arises as the result of productivity maximization or if other intervening concerns might be driving this pattern in practice. In order to asses this, we investigate how the skill of managers and workers combine to determine team productivity.

We perform a series of tests to verify that the production function is in fact additively separable in logs, which is consistent with the underlying technology exhibiting a positive cross-partial between worker and manager skill in levels. This implies that a given worker productivity *increases* in the manager type, and so – absent other constraints in production – total productivity would be maximized by implementing positive assortative matching between workers and managers. This is confirmed by the results of a counterfactual exercise. Specifically, using our estimates of the worker and manager fixed effects, we simulate total productivity under the perfect positive assortative matching allocation between workers and managers. We find that indeed total productivity would increase by between 1-4% across the factories in our sample under this counterfactual allocation. This suggests that the shape of the underlying production technology cannot be an explanation for the observed pattern of *negative* sorting on average.

We hypothesize that negative matching arises because of strong incentives to avoid long delays in completing any particular order. That is, the firm is willing to forfeit some productivity to ensure that minimum productivity on least productive lines does not fall so low as to delay completion and delivery of an order. Delays of this sort can damage the relationship between the manufacturer and the brand buyer, leading to lower prices, less orders, and even termination of the relationship. These incentives of the firm to avoid delays are passed down to line managers as well.

We conduct a survey of managers to assess the importance of these considerations. We show that factories in which managers are most worried about falling behind on orders and not meeting deadlines with buyers are indeed the ones where negative sorting is strongest. In addition, using data on buyers and orders, we find that factories more beholden to large buyers, for whom the cost of damaging any one relationship is large, also exhibit the strongest negative sorting. Consistently with this, the results of our counterfactual simulations show that these are precisely the factories that would experience the highest increase in productivity under the perfect positive assortative matching allocation.

Our results suggest that the presence of underlying constraints related to the nature of supply chains (i.e., the risk of damaging valuable relationships with buyers) prevents firms from fully exploiting complementarities in production. These findings likely generalize to the global supply chain for many products, in which suppliers produce orders for multiple buyers, but an imperfectly competitive market might hold suppliers inside their production frontier as they strive to protect valuable relationships with buyers. These interactions between suppliers and buyers, and the reputational cost of non-delivery, have been the subject of some recent empirical studies in the trade and development literature (Cajal Grossi et al., 2019; Macchiavello and Morjaria, 2015).

Our study makes four main contributions. First, we complement the theoretical literature in organizational economics and firm productivity on how production is organized inside the firm (Holmstrom and Tirole, 1989; Kremer, 1993). Recent empirical studies in personnel economics have started to inform how co-workers impact each other's productivities (Amodio and Martinez-Carrasco, 2018; Boning et al., 2007; Hamilton et al., 2003), as well as how the interaction between workers and their supervisors determines firm productivity (Adhvaryu et al., 2019; Frederiksen et al., 2017; Hoffman and Tadelis, 2019; Lazear et al., 2015). We add to these related literatures by providing direct evidence on the sorting pattern of workers to managers *within* a firm. We also leverage granular team production data with high frequency, quasi-random changes in team composition to contribute estimates of how the skills of workers and managers combine to determine productivity. In doing so, we document that the realized pattern of negative sorting between workers and managers does not reflect the most productive possible match, and highlight competing considerations that affect team composition.

Second, we add to a rich empirical literature on management and productivity (Adhvaryu et al., 2019; Bloom and Van Reenen, 2007, 2011; McKenzie and Woodruff, 2016). Though recent experimental studies have convincingly proven that increasing managerial quality can increase the productivity of the firm (Bloom et al., 2013, 2018; Karlan et al., 2015; McKenzie and Woodruff, 2013), whether managerial skill complements or substitutes for worker skill and how the stock of managerial skill is distributed within the firm are open questions. We contribute two novel empirical facts from the labor-intensive manufacturing context: 1) manager skills and worker skills are imperfectly substitutable, which should provide incentives towards positive assortative matching, but 2) highly productive managers tend to be matched with low productivity workers. These two facts together emphasize another way in which managerial quality may contribute to low productivity in developing country settings (Bloom and Van Reenen, 2007; McKenzie and Woodruff, 2016). That is, not only is stock of managerial skill low, but the existing stock may not be properly allocated to maximize productivity. Suppliers to the global market, concentrated in developing countries, may be beholden to a small set of powerful buyers from developed countries. Accordingly, suppliers may be driven to "misallocate" managerial skill in service of these relationships, but at the expense of productivity.

Third, we add to a long-standing literature in labor on worker sorting (Abowd et al., 1999; Eeckhout, 2018). Empirical studies have focused on labor market level sorting of workers *across* firms using *wages* (Card et al., 2013; Lopes de Melo, 2018). We extend this literature to the estimation of sorting *within* the firm, leveraging granular data on *productivity* rather than wages. We note that the identification issues that arise when using wages (Eeckhout and Kircher, 2011; Hagedorn et al., 2017), do not apply when productivity is observable as is often the case in firm personnel data. We also find NAM, in contrast to most across firm patterns, and contrary to what is implied by the imperfect substitutability of managers and workers uncovered by the estimated production function. These results emphasize the distinction between empirical estimates of sorting and the shape of the underlying production function (Eeckhout and Kircher, 2011).

Finally, we document that the gap between the match of workers to managers that would maximize productivity and the match which prevails in the data is driven, at least in part, by concerns regarding maintenance of relationships with important buyers. This evidence complements recent empirical work in the literature on trade and development documenting the importance of relationships between suppliers and buyers in global supply chains. Some recent empirical work has focused on how prices and quantities reflect the importance of these relationships to suppliers (Macchiavello and Morjaria, 2015), and how the features of these relationships determine suppliers' markups and marginal cost (Cajal Grossi et al., 2019). Other studies have shown that these relationships can improve product quality and technical expertise of suppliers (Atkin et al., 2017). We add evidence that production and personnel decisions can reflect buyer relationship considerations as well, and that suppliers might even forfeit productivity to maintain relationships with buyers. This stands in contrast to recent evidence of learning-by-exporting, in that we document one way in which buyer relationships might *reduce* supplier productivity.

The rest of the paper is organized as follows. In Section 2 we describe the setting and data. In Section 3 we present estimates of the sorting pattern of managers to workers. Section 4 explores the potential drivers of the sorting pattern recovered in Section 3, focusing on the role of the underlying production technology and other constraints in production. In Section 5 we perform a counterfactual simulation to study the potential productivity gains from labor reallocation. Section 6 concludes.

2 Context, Data and Descriptives

In this Section, we describe the setting where our study takes place and the data used for estimation. We then present descriptive evidence on the degree of mobility of workers across production lines, as well as on productivity dispersion across workers and lines. As described later in the paper, a high degree of mobility and substantial dispersion in productivity are necessary to recover the pattern of sorting of managers to workers with the estimation procedure we follow.

2.1 Context and Data

We study the sorting of workers to production lines (and corresponding line managers) in six ready-made garment factories in Bengaluru, India. We leverage data on worker-level output and production team composition from one of the largest ready-made garment manufacturers in the world.

Production in these factories takes place as follows. Orders for export production from large international buyers are allocated by the marketing department of each production division (i.e., Ladies', Men's, Knits) to the factories based on capacity and regulatory compliance. Within each factory, the order is assigned to a production line based on first availability. The production line will then work on the entire order until it is ready to be prepared for shipment, usually in advance of the contracted delivery date. Each production line works on one order at a time, which usually takes between 20 to 30 days to complete.

A typical production line has between 65-70 workers which usually corresponds to one worker per machine. Within each line, the production process is organized in a sequence, usually grouped by segments of the garment. Within these groups are feeding points at which bundles of material for a certain number of segments are provided. For example, a group of workers assigned to machines will complete x numbers of sequential operations to produce left sleeves, another similar group will do the same for right sleeves, and another for shirt fronts with pockets, and another group will work on the collar. Completed bundles of sections of garments then feed other segments of the line, until a completed garment results at the end of the line.

Each line is managed by a line manager, whose role is to motivate workers, assign them to tasks, and ensure that production remains on schedule by identifying and relieving bottlenecks. Importantly, workers on each line are only interacting with the line manager and *not* with the other workers on the line. So there is no team-work between workers on the line.[1] *Line* managers are supervised by a *production* or floor manager, who usually supervises four line managers. Production managers are in charge of ensuring that their lines run smoothly and meet the production targets.

There are three main stages of the production process for each garment: cutting, sewing and finishing. We focus on the sewing process for three reasons: first, it involves the majority of labor in the production process; second, it makes up the majority of the production timeline; finally, our data allow us to follow the daily composition of the team and the output of each worker/production line for the sewing process, which is needed for our analysis.

The data include daily worker-level data from six factories, and spans over two years, from July 2013 to July 2015. For each day of production during this period, we know the line a worker is assigned to, the task she is assigned to perform, how many garments she assembles, and the target quantity. The target quantity is higher for less complex garments (since workers can produce more simple garments in a given day), and therefore is an appropriate way to normalize productivity across lines producing garments of different complexity. Our measure of productivity is daily "efficiency", which equals the percentage of the target quantity of a particular garment per day that is achieved by the worker. This measure ranges from 0 (lowest efficiency) to 100 (highest efficiency).[2]

Line managers and workers are paid a fixed salary for their work, but are eligible to earn bonus pay each day that their line exceeds a minimum production threshold. The bonus is linear in productivity above this threshold and does not reflect the productivity of any other line in the factory.[3] Thus, managers are incentivized to utilize all available resources to maximize their output each day. Similarly, workers are also incentivized to exert effort. We have data on the daily wages of workers and managers (inclusive of the bonus) over the sample period. As described in more detail in the next section, this allows us to compare how

[1]The potential for spillover effects across workers is also limited by the fact that each worker has a buffer stock of material to work on, so that each worker productivity is not influenced by the productivity of the other workers on the line. We formally test for the presence of spillovers between workers in Appendix A.

[2]The target quantity for a given garment is calculated using a measure of garment complexity called the standard allowable minute (SAM). SAM is taken from a standardized global database of garment industrial engineering that includes information on the universe of garment styles. It measures the number of minutes that a particular garment should take to produce.

[3]The value of the bonus ranges between 8%-10%.

the results of the estimation procedure to recover the sorting pattern of workers to managers differ when using wages and when using productivity.

2.2 Descriptives on Worker Mobility and Productivity Dispersion

Worker mobility across lines is very common in these factories. One main reason for this is worker absenteeism. In our context, absenteeism shocks are frequent and large. On a typical day, 13-14% of workers are absent, and so it often happens that the available workers are reassigned across lines by production managers to make sure that each line has enough workers and is able to complete the critical operations required to finish the order in time. In addition, as discussed in more detail in Section 4, production managers often reallocate workers across lines depending on the specific manpower needs and the progress of the different lines with the order, regardless of absenteeism. For example, if a line is falling behind in the timeline to deliver an order, production managers can temporarily move there productive workers taken from other lines that are on track to meet their deadlines. This creates another reason for the observed mobility. It is important to note that any worker reassignment across lines is decided by managers, so that workers do *not* have the freedom to choose which lines to work at.

In Table 1, we present summary statistics at the factory (or unit) level. Over the sample period, we observe 23,608 workers distributed along the six factories and 120 production lines. Each factory has around 4,000 workers and 26 production lines. The average tenure of workers in these factories is around 9 months (with the median at around 5 months). The Table further shows that the share of *movers* (i.e., workers that are observed at more than one production line during their tenure) is 54%, confirming the high levels of mobility in our data.[4]

Figures 1A and 1B show the distribution of the average efficiency of workers and lines, pooled across days, so that in each graph there is a single observation for each worker and for each firm. These Figures reveal that there is substantial variation in the productivity of both workers and lines in our data.[5]

In the next Section, we describe the estimation approach to recover the sorting of workers to managers: this leverages the high degree of mobility of workers and the high frequency data on worker productivity and wages.

[4]Table A1 in the Appendix reports the distribution of lines workers are observed at, and shows that, conditional on moving, the median worker is observed at three lines.

[5]Figure A1 in the Appendix shows the distribution of the average efficiency of workers, by factory.

3 Estimating the Sorting Pattern within the Firm

We begin the empirical analysis by estimating the sorting pattern between production lines and workers within the six factories in our data. This relies on obtaining estimates of worker and line fixed effects, and then computing their correlation. To do so, we follow the approach in the seminal paper by Abowd et al. (1999), but using worker-level productivity data rather than wage data. This "AKM" approach is described in the next subsection, which discusses the main identifying assumptions and also highlights the advantages of using productivity data to overcome important identification concerns that the literature has raised about using the AKM approach with wage data. We then turn to presenting the results of the estimation.

3.1 Methodology

We estimate the following two-way fixed effects model:

$$ln(y_{it}) = \theta_i + \psi_{J(i,t)} + x'_{it}\beta + \nu_{it}, \tag{1}$$

where:

$$\nu_{it} = \eta_{i,J(i,t)} + \xi_{it} + \epsilon_{it}. \tag{2}$$

The dependent variable, $ln(y_{it})$, is log daily efficiency of worker i at time t; θ_i is a worker fixed effect; $\psi_{J(i,t)}$ is a fixed effect for the line (or manager) the worker was matched to at time t; x'_{it} are time-varying controls.[6]

Following Card et al. (2013), we assume that the error term in the main equation, ν_{it}, is the sum of a match-specific component, $\eta_{i,J(i,t)}$, a unit root component, ξ_{it}, and a transitory error, ϵ_{it}. The term $\eta_{i,J(i,t)}$ allows for the log-productivity of worker i to be inherently different across firms; the component ξ_{it} captures changes in the worker fixed effect over time, due for example to employer learning or human capital accumulation; the transitory error term ϵ_{it} reflects any additional unobserved worker-level variation, for example health shocks that affect the productivity of the worker at particular times.

As discussed in Abowd et al. (2002), the line and worker fixed effects in this model are separately identified only within "connected sets" of lines, linked by worker moves across

[6] The full set of time-varying controls includes style fixed effects to account for variation in productivity due to complexity of the style and size of the order, as well as year, month and day of the week fixed effects, to account for common seasonality and growth in productivity across lines. Finally, we include the experience that line $J(i,t)$ has in producing the current style at date t in the current production run, as measured by the number of consecutive days spent producing that style, to account for line-specific learning and productivity growth during the production run.

lines. Therefore, we estimate this equation on the largest connected set of workers and lines within each of the six factories in our sample.[7] As described in the previous section, the final estimation sample includes 23,608 workers observed at 120 lines for a period of 2 years. In total, this delivers 2,925,577 daily observations of the efficiency of a particular worker matched to a particular line within the connected set.

We now discuss the specific identification assumptions related to the estimation of equation (1), and present a number of tests to validate these assumptions.

3.1.1 Identification Tests

In order to consistently estimate the parameters in (1) by OLS, we again follow the literature (see, for instance, Card et al. (2013)) and make the following identifying assumptions: $E[\theta_i \nu_{it}] = 0$; $E[\psi_{J(i,t)} \nu_{it}] = 0$; and $E[x'_{it} \nu_{it}] = 0$, $\forall\ i,\ t$. In particular, identification of the line fixed effect requires a strong exogeneity assumption regarding the assignment of workers to lines with respect to ν_{it}: we need the assignment of workers to lines to be conditionally mean-independent of past, present and future values of ν_{it}. Note that this assumption allows for the possibility, for example, that better workers (i.e. workers with a higher fixed effect) are systematically more likely to move to more productive lines (i.e. to lines with a higher fixed effect), so that sorting on the fixed effects is allowed; what this assumption rules out is that workers and lines sort on the match-specific component, or on other transitory shocks to workers or lines. Any form of "endogenous mobility", whereby workers and lines sort on ν_{it}, would lead to biased and inconsistent estimates of the fixed effects.

We follow Card et al. (2013), and perform a series of tests for endogenous mobility. We begin by conducting an event study around moves to assess the extent to which moves might be systematically driven by productivity shocks or by sorting on the match-specific component of log-productivity. Specifically, we isolate movers in our data, and then rank them in terms of: (i) quartiles of the average co-worker residual efficiency at the line they moved away from; and (ii) quartiles of the average co-worker residual efficiency at the line they moved to.[8] Figure 2 then plots the average weekly residual efficiency of the mover on the y-axis: this is computed 6 to 10 days ($Period = -2$) and 1 to 5 days ($Period = -1$) before

[7] A group of lines and workers are connected when the group comprises all the workers that have ever matched with any of the lines in the group, and all of the lines at which any of the workers have been matched during the sample period. The largest connected set in our data includes about 98% of the observations.

[8] To calculate worker-level residual efficiency we run a regression of log daily efficiency of the worker on: factory fixed effects; year, month and day of the week fixed effects; style fixed effects; tenure (days) of the worker in the factory; tenure (days) of the worker in the line; finally, we include the experience of the line in producing the current style in the current production run, as measured by the number of consecutive days spent producing that style. Standard errors are clustered by line in this regression. We use this regression to calculate residual efficiency of each worker.

the move from the old line, and 1 to 5 days ($Period = 1$) and 6 to 10 days ($Period = 2$) after the move to the new line, on the x-axis.[9]. This is plotted by quartiles of the average residual efficiency of the origin line and destination line. To limit the amount of information on the graph, we only report moves away from either the top quartile in terms of co-worker residual efficiency (quartile 4) or the bottom quartile of co-worker residual efficiency (quartile 1).

If match-specific components, $\eta_{i,J(i,t)}$, are important in driving moves, so that workers are more likely to move to those lines where they are particularly productive, then we would expect workers to *gain* on average from moving to a new line. Instead, if moves are not driven by match-specific components, on average workers who move to higher productivity lines will become more productive, and workers who move to lower productivity lines will become less productive. Figure 2 shows that indeed workers moving to better lines tend to gain in terms of productivity, while workers moving to worse lines tend to lose in terms of productivity. In addition, workers who move from a line in the highest quartile to a line in the highest quartile experience close to zero change in productivity, and the same is true for workers who move between lines in the lowest productivity quartile. These results are consistent with the absence of an average "premium" for movers, which supports the identification assumptions.[10]

To further validate that the match-specific component of productivity is not important in driving moves, we compare the Adjusted R^2 from the estimation of equation (1) with the Adjusted R^2 from a fully saturated model with dummies for each worker-line combination. Table A3 in the Appendix shows that the improvement of fit from the fully saturated model is very limited (the Adjusted R^2 increases by only .017), so that any scope for sorting on this component is limited. Finally, in Figure A3 in the Appendix, we report the average residuals from the estimation of (1), by quartiles of the estimated worker and line fixed effects. The average residuals are very small for all groups (substantially below 1% in all cases). This is again consistent with match effects not being quantitatively important, and so provides

[9]The sample for the graph is restricted to the balanced sample of workers continuously employed at the origin line for at least 10 days prior to the move, and continuously employed at the destination line for at least 10 days after the move.

[10]As discussed in Card et al. (2013), if the moves are conditionally mean independent of the match-specific component, then the gains from moving from line X to line Y should be equal and opposite to the losses from moving from line Y to line X. That is, gains and losses for movers should be symmetric. A full symmetry test across all potential combinations of origin and destination lines (ranked by quartiles of co-worker residual efficiency) is reported in Figure A2 in the Appendix. The Figure again shows that moving to a higher productivity line results in a gain in productivity, while moving down results in a loss in productivity, with the exception of moves from the third to the second quartile, and moves from the fourth to the third quartile of lines, as these are associated with small increases in productivity despite being a move down. The Figure further reveals that the gains from moves up and the losses from moves down are relatively symmetric. While we do observe some deviations from the 45 degree line, these deviations do not appear to have a systematic direction, which is reassuring.

further support to the additive separability assumption of equation (1).

A second concern about ξ_{it} arises if those workers who are performing particularly well at a given line are more likely to move up, and those who are performing particularly badly are more likely to move down, as this would lead to an overestimate of the line effect. Therefore, we check whether the productivity of movers at the origin line exhibits systematic trends in the days just before the move: while Figure 2 reveals that worker residual productivity does exhibit some movement in the periods before the move, these do not seem to be systematically related to whether the worker then moves to a higher productivity or a lower productivity line, which again supports the assumption of conditional exogenous mobility.[11]

Third, suppose that workers who experience a positive transitory productivity shock ϵ_{it} are more likely to move up: since the shock is transitory, this would lead to an underestimation of the line fixed effect (due to mean reversion). Again, as shown in Figure 2, the absence of systematic trends before (or after) the move takes place helps alleviate such concerns.[12]

Taken together, the evidence from this section provides support to the identification assumptions detailed above: while mobility in our data is high, such mobility does not seem to be driven by match-specific components, or other unobserved time-varying worker components.

3.1.2 Addressing Limited Mobility Bias

Another concern with the estimation of (1) is the so-called "limited mobility bias" (Abowd et al., 2004; Andrews et al., 2008, 2012). As discussed above, identification of the worker and line fixed effects requires: (i) observing a worker at multiple lines; (ii) observing a line with multiple movers. If in practice the number of moves is limited and the size of each line is small, then this can bias the correlation between the estimated worker and firm fixed effects, and the bias will be negative. Intuitively, if the presence of limited mobility does not allow to separately identify the worker and line fixed effects, then if the worker effect is overestimated, the line effect will be underestimated, and vice versa, creating negative bias in the correlation of the estimated worker and line fixed effects.

We address limited mobility bias in two ways. First, we note that in our data: (i) mobility is much higher than in typical matched employer-employee (MEE) datasets; and (ii) lines are much larger than the average firm in typical MEE datasets. As shown in Table 1, more

[11] We further validate this in Table A2 in the Appendix, which studies whether changes in average residual productivity in the two weeks before the move systematically predict the direction of the move. In line with Figure 2, we find no significant evidence of changes in productivity predicting where workers move.

[12] The results in Table A2 in the Appendix provide further support to this conclusion, as we fail to document a significant correlation between fluctuations in productivity in the two weeks before the move and the direction of the move.

than 50% of workers in our sample move at least once, and the average line has around 65-70 workers at each point in time. By contrast, the share of workers who move at least once is around 12% in the German data in Andrews et al. (2012), and about 35% in the Brazilian data in Alvarez et al. (2018). Also, the simulations in Andrews et al. (2008) are conducted only for firm sizes of at most 15 employees. These considerations already limit concerns related to limited mobility bias in our setting. To further alleviate concerns related to limited mobility bias, in Figure A4 in the Appendix we also perform the bias correction procedure suggested by Andrews et al. (2008), which is standard in the literature. In line with limited mobility bias not being substantial, our estimates are robust to such Andrews et al. (2008) correction.

3.1.3 The Advantage of Productivity Data

Before turning to the results, we note that a key advantage of our data is that it includes information on job-level productivity. This allows us to overcome a long-standing concern in the literature that wage data is inappropriate to recover the sign of sorting. In particular, Eeckhout and Kircher (2011) highlight that in the presence of a positive cross-partial in production between worker and manager types, wages (in levels) can be non-monotonic around the optimal allocation. This causes the firm fixed effect estimated from wage data to be uncorrelated with the underlying firm type, thus preventing identification of sorting with wage data. We are able to overcome their critique because even in the presence of a positive cross-partial, worker productivity will be monotonic in the firm type, which justifies using the AKM framework with productivity data. To further highlight the importance of using productivity data, we contrast our results using productivity data with those using wage data.

3.2 Results

We estimate equation (1) by OLS, clustering standard errors by line. Table 2 reports the results of the estimation, averaged across the six factories in our data. Column 1 focuses on the specification where we use worker productivity as outcome. The key parameter of interest is the correlation between the worker effect and the line effect, $Corr(\theta_i, \psi_j)$. This is *negative* and relatively large at around -16%. These results indicate that on average higher-productivity workers are more likely to be matched with lower-productivity lines (or managers).

By contrast, we note that the same estimation routine conducted with wages as outcomes returns a correlation very close to zero. This is in line with the nature of wage setting in these factories, where the largest component of worker wages is a fixed salary, with limited scope for wage variation, which again highlights the importance of using data on productivity.[13]

We further verify the results in Table 2 by studying the direction of moves in our data: the negative documented correlation between the worker and line fixed effects implies that the most common moves in our data should involve a high productivity worker moving to a low productivity line, and a low productivity worker moving to a high productivity line, as this is the pattern of moves that would yield negative assortative matching. On the other hand, moves of high productivity workers to high productivity lines, as well as low productivity workers to low productivity lines, should be less frequent.

Table 3 presents descriptive evidence on the pattern of worker moves across lines. To do so, we divide both workers and lines into those with an estimated fixed effect in the top quartile (i.e. "High" type) and those with a fixed effect in the bottom quartile (i.e. "Low" type) over the sample period. We then report the daily probability of observing a worker of a given type moving to a line of a given type. The results show that the likelihood of a high type worker moving to a low type line is indeed almost four times as large as the likelihood of a high type worker moving to a high type line. Also, low type workers tend to move more often to high type lines than to low type lines. Reassuringly, this pattern of moves is consistent with the negative sorting results in Table 2.

Next, we explore the robustness of this negative sorting result across different groups of workers. Table 4 shows that the negative sorting result is not driven by some particular groups of workers, and instead holds for both high and low skilled workers, and for high and low paid workers. Also, Table 4 shows that the sorting result holds both when workers are observed at their primary or "home" line, and when they are observed at other lines. In addition, the table shows that the results are very similar when they are estimated on the sample of movers only, which again reassures us that limited mobility bias is not substantial in our data.

Figure A5 in the Appendix reports the estimated correlations between worker and manager effects by factory. These are negative for five of six factories, and the other correlation is very close to zero, again suggesting that the negative sorting pattern estimated on average is reflective of the organization of production across most of the factories. As Figure A5 highlights however, there is substantial heterogeneity in the *magnitude* of sorting across factories. We return to explaining this heterogeneity in the next section. Finally, to further rule out concerns related to limited mobility bias, we show in Appendix Figure A4 that our

[13] The difference between the results with productivity and wage data is also in line with the discussion in Eeckhout and Kircher (2011) that the firm fixed effect estimated with wage data is usually uncorrelated with the true firm type, thus potentially leading to a correlation of zero between the worker and the firm fixed effect.

results by factory are robust to applying the bias correction procedure of Andrews et al. (2008).

4 Drivers of Negative Assortative Matching

The results from the previous section uncover the presence of negative assortative matching between workers and managers within the firm. This means that on average, high productivity workers are more likely to be observed at lines with low productivity during the sample period. The key question that emerges then is what is causing this sorting pattern: is this driven by the shape of the production function, so that higher ability workers are more productive when matched to lower productivity managers? Or does this reflect some other underlying constraint in production, so that even though it might otherwise be optimal to match high productivity managers with high productivity workers, some other "force" pushes the allocation towards NAM? In this section, we examine the importance of these two potential explanations in driving the observed sorting pattern. Understanding this is important to shed light on the constraints firms face in the production process, and how these affect firms' ability to exploit the underlying production technology between workers and managers in their production decisions.

4.1 The Role of the Underlying Production Technology

We begin by examining the potential role of the shape of the production function in driving the observed allocation: if the production function had a negative cross-partial, then this would explain the observed sorting pattern, as NAM is the allocation that maximizes output given that technology; on the other hand, if there is a positive cross-partial between manager and worker productivity, then this indicates that the optimal unconstrained allocation should exhibit PAM, and so the firm must face other constraints that push towards NAM.

We note that equation (1) is effectively a production function in logs, since the outcome is efficiency. In the previous section, we have assumed that this equation is additively separable in the worker and line fixed effects. This is equivalent to assuming that the production function, in levels, exhibits a positive cross-partial between worker and line type (i.e. that worker and line effects enter multiplicatively in levels). For instance, equation (1) is consistent with the underlying production function being Cobb-Douglas in levels. As long as the identification assumptions laid out in the previous section are satisfied, then this implies that productivity would be maximized by implementing a *positive* assortative matching allocation.

The identification tests conducted in the previous section support the assumption of additive separability in logs. In particular, the fact that the Adjusted R^2 from the estimation of equation (1) does not increase substantially once match effects are included (Appendix Table A3), and that the average residuals are small overall (Appendix Figure A3) suggests that match effects are not important in logs. This is in line with the cross-partial being zero in the log form of the production function. In turn, then this implies a positive cross-partial in levels.

To further explore the shape of the underlying production technology, Appendix A describes an alternative production function estimation procedure that we undertake. This does not rely on the estimation of worker and firm fixed effects, and so is complementary to the estimation of equation (1). In short, the procedure amounts to splitting the sample into two periods: in the first period (e.g. the first three months of data) we rank worker and lines by the decile of their raw average daily efficiency. This determines the worker and line "types". We then use these as inputs for estimating the production function in the second period. The results again confirm that we cannot reject that the production function is additively separable in logs, which implies a positive cross-partial in levels.

Taken together, this evidence shows that we cannot reject that the underlying production function is Cobb-Douglas. Since the cross-partial of the Cobb-Douglas is positive in levels, this means that the output of a given worker is increasing in the manager type. This is in line with much of the literature on managerial incentives, which tends to assume complementarities between managers and workers, and more generally between inputs in the production process.[14]

Given this underlying technology, we would then expect to find positive assortative matching as that would lead to output maximization. In line with this, as discussed in more detail in Section 5 below, we show that productivity would increase substantially in these factories under a counterfactual allocation that implements the positive assortative matching assignment. The shape of the underlying production function cannot be the driver of the observed allocation then, and so there must be other constraints that create an incentive for the firm to deviate from the positive assortative matching allocation. We turn to this in the next sub-section.

[14] For example, Bandiera et al. (2007) and Bandiera et al. (2009) consider complementarities between managerial and worker effort; Amodio and Martinez-Carrasco (2018) find evidence in favor of complementarities between worker quality and input quality.

4.2 The Role of other Constraints in Production

As discussed by the theoretical literature on sorting in the labor market (Eeckhout, 2018), the presence of a positive cross-partial in production typically leads to PAM in a competitive equilibrium in which production units compete with each other. On the other hand, even in the presence of complementarities, NAM can emerge if there is a social planner that internalizes the externalities imposed by the more productive teams to the less productive ones. If these externalities are large, then this can lead to NAM. In our context, this would be the case if the central management of the firm cares about *all* teams meeting a certain minimum level of productivity. That is, if there is a large penalty associated with having a production line fall behind on a given order, then this might result in the central management finding it optimal to pair the most productive workers with the least productive managers. This would then lead to NAM, even if the underlying production technology would push towards PAM.

To explore this possibility, we surveyed all production managers in the six factories in our sample.[15] Each production manager is in charge of multiple lines, and they are responsible for the productivity and progress of all the lines under their control. In addition, they are responsible for the assignment of workers to lines. The survey included all 80 such production managers across the six factories, and was designed to understand the main concerns of production managers. In particular, the survey focused on concerns related to lines falling behind with their orders, and what managers do to address such constraints.

Specifically, we asked managers to indicate the relative importance of four potential concerns in their operations: (i) lines not meeting their target/running slow; (ii) worker absenteeism; (ii) line manager absenteeism and (iv) customers not paying for their orders on time. Managers were asked to use a 1 to 5 scale to indicate the importance of each concern, with 1 meaning "not worried at all", and 5 meaning "very worried". Figure 3 shows that difficulties in meeting targets are reported as the most important concern, together with concerns about customers not paying on time for their orders. Interestingly, Figure 3 shows that concerns related to lines falling behind are more important than concerns related to worker or manager absenteeism, thus indicating that concerns related to slow progress of the lines are indeed important.

To understand why production managers are concerned about lines falling behind, we asked them what are the consequences for the firm if a production line is slow and does not meet the deadline for a given order. We report answers to these questions in Panel A of Table 5, which shows that: 51% of production managers say that there would be substantial

[15] The survey took place in early 2019.

monetary losses for the firm from being late with an order, and 33% report that the firm might lose the customer altogether from being late with the order. In line with this being a problem for production managers, the survey further reveals that 19% of managers have experienced delays with a production line under their supervision.

We then asked managers about the strategies they adopt to try and avoid having lines fall behind. As shown in Panel B of Table 5, the data reveals that 91% of managers would consider moving workers across lines to help the low-performing lines catch up. Importantly, we asked which *types* of workers they would be more likely to shift from a highly productive line to a lower productivity line, and in 97% of cases managers would move the *most* productive workers to the *least* performing line.

Taken together, the evidence from this survey shows that constraints related to the structure of supply-chains push the firm towards implementing negative assortative matching: missing the deadline for an order is costly for the firm, and so to make sure that lines do not fall behind, production managers often reallocate high productivity workers to low productivity lines. Doing so is optimal for the firm, given the constraints related to the supply chain. However, the results of our production function estimation suggest that output and productivity would be higher in a counterfactual scenario in which the firm was not facing such supply-chain constraints, and could instead afford to keep some highly productive lines and some less productive lines. These results highlight that constraints related to the structure of supply-chains are significant and limit in important ways the productivity of the firm. We return to this point in Section 5, where we quantify the loss in productivity related to such supply-chain constraints.

4.3 Explaining the Heterogeneity across Factories

As discussed in Section 3, Figure A5 shows that there is significant heterogeneity in our estimates of the correlation between worker and line fixed effects across the six factories in our data. We now explore potential drivers of such heterogeneity.

4.3.1 Heterogeneity by Manager Concerns

The discussion in the previous sub-section suggests that we should expect the degree of NAM to be stronger in those factories where managers on average are more concerned about lines falling behind and not meeting deadlines for their orders with buyers. We provide evidence on this in Figure 4, which plots the estimated degree of NAM in each factory as a function of the relative reported concern of managers about lines not meeting deadlines.[16] The Figure reveals that, as expected, we find stronger NAM in those factories where managers are more concerned about lines not meeting their targets.

4.3.2 Heterogeneity by Buyer Characteristics

A related literature on supply-chains in developing countries highlights the value for suppliers of long-term relationships with buyers (Cajal Grossi et al., 2019; Macchiavello and Morjaria, 2015). We can then expect the firm to place more weight on safeguarding relationships with established and long-time buyers, as these are relationships that are particularly valuable to the firm. If this is the case, we would then expect to find a higher degree of NAM in those factories that are engaged with producing orders for many long-time and important buyers, as the firm wants to make sure that all deadlines with such buyers are met.

To explore this hypothesis, we exploit a dataset with information on buyers to the garment firm. The dataset covers the period from late 2011 to late 2015, and includes information on all orders placed to the firm, with a corresponding unique buyer identifier. Over this period, 125 different buyers placed an order to the firm, with the median buyer placing 105 orders. This confirms that this is a very large firm, with many active buyers. To identify "Long-time buyers" we create a dummy equal to one if the buyer is in the top tercile of the total number of independent orders placed to the firm.[17] For each factory in our data, we then calculate the number of Long-time buyers that placed orders at that factory over the sample period. Figure 5 then shows how the estimated NAM varies with the number of Long-time buyers. The Figure confirms that we find a stronger degree of NAM in those factories that are particularly involved with producing for important buyers.

While any conclusions from these Figures are obviously subject to the caveat of the small sample size of six factories, these results support the claim that constraints related to supply-chain relationships are important drivers of the sorting pattern of workers within the firm, and of overall firm productivity as a result.

5 Productivity Gains from Labor Reallocation

The analysis in the previous Sections shows that supply chain constraints push the firm towards implementing a negative assortative matching allocation, which is *not* the allocation

[16] For each factory, we divide the average reported concern about lines not meeting deadlines by the within-factory average of all concerns managers were asked about. We then plot this measure on the y-axis of Figure 4.

[17] Results using a different definition of Long-time buyer, e.g. the top quartile of the number orders, produce similar results.

that would maximize productivity in an unconstrained environment. In this section, we quantify the loss in productivity that is associated with implementing the current allocation. This sheds light on the size of the supply chain constraint, that is, on how much the firm is willing to sacrifice in terms of productivity in order to safeguard its relationship with powerful buyers.

To quantify the gains from labor reallocation, we simulate total firm efficiency under a perfect positive assortative matching allocation. The simulation is implemented as follows. We randomly extract one day from the sample period, and for that day we record: (i) the observed allocation of workers to lines and (ii) the fixed effects of the workers and the lines (estimated from equation (1)). We then artificially move workers across lines to implement the perfect positive assortative matching allocation. This corresponds to assigning the workers with the highest fixed effects to the lines with the highest fixed effects and so on, respecting the line sizes observed in the data. Individual worker efficiency is then predicted using the estimated equation (1), but with workers and lines matched following perfect positive assortative matching. The predicted log efficiency from (1) is then exponentiated to recover the counterfactual efficiency in levels, which is then summed across all workers and all lines. This procedure is then repeated on 1,000 randomly extracted days.

Figure 6 reports the estimated productivity gains from the counterfactual simulation, across the six factories in our data. We plot the mean increase in daily efficiency under positive assortative matching, together with confidence intervals, where bootstrap standard errors are used to construct the confidence interval. The Figure shows that the productivity gains from reallocation are in the range 1-4%. As expected, the gains are larger for those factories where NAM is also larger, as these are the factories that have more scope for gains from reallocation.[18]

The magnitudes in Figure 6 give us a sense of the “value” of this supply-chain constraint for the firm. That is, they show that the firm is willing to sacrifice between 1-4% of productivity in order to avoid delays in meeting deadlines with buyers. This is a sizable reduction in productivity, which is in line with such relationships being valuable for the firm. Recent literature has focused on the value of repeated relationships with large buyers (Cajal Grossi et al., 2019; Macchiavello and Morjaria, 2015). Our results contribute to this literature by showing that in order to safeguard such valuable relationships, the firm is willing to “misallocate” labor internally, thus giving up some productivity. This result provides new insights into how supply-chain constraints determine firm productivity.

[18]Note that the relationship between the gains from reallocation and the estimated NAM does not have to be monotonic necessarily. This is because the potential for each factory to gain from reallocation depends not only on the degree of NAM, but also on the distributions of the worker and line fixed effects, which are likely to vary by factory.

6 Conclusion

We characterize the pattern of sorting between workers and managers in the context of a large readymade garment manufacturer in India. We extend the across-firm sorting literature by estimating the sorting pattern *within* the firm, and by leveraging granular data on *productivity* rather than wages. We find evidence of negative assortative matching (NAM) – that is, better managers tend to match with the worse workers, and vice versa. However, our estimation results also uncover a positive cross-partial between worker and manager skill, which implies that the productivity of a given worker is increasing in the manager's type.

These two facts together emphasize another way in which managerial quality may contribute to low productivity in developing country settings (Bloom and Van Reenen, 2007; McKenzie and Woodruff, 2016). That is, not only is stock of managerial skill low, but the existing stock may not be properly allocated to maximize productivity. This discrepancy also emphasizes the distinction between empirical estimates of sorting and the shape of the underlying production function (Eeckhout and Kircher, 2011).

We note that if the firm were to instead positively sort, aggregate output would increase by between 1-4% across factories. We hypothesize that that negative matching arises, despite this loss in productivity, because of strong incentives to avoid long delays in completing any particular order. That is, the firm is willing to forfeit some productivity to ensure that minimum productivity on least productive lines does not fall so low as to delay completion and delivery of an order. We conduct a survey of managers to assess the importance of these considerations and find that factories in which managers are most worried about falling behind on orders are indeed the ones where negative sorting is strongest. In addition, factories more beholden to large buyers, for whom the cost of damaging any one relationship is large, also exhibit the strongest negative sorting.

Our results suggest that the presence of underlying constraints related to the nature of supply chains (i.e., the risk of damaging valuable relationships with buyers) prevents firms from fully exploiting the underlying production technology. These results contribute generalizable insights to the study of the large gap in productivity between developed and developing country firms (Caselli, 2005; Hall and Jones, 1999) and the impact of trade relationships on closing or widening this gap (Atkin et al., 2017). Our results indicate that suppliers to the global market, concentrated in developing countries, may be beholden to a small set of powerful buyers from developed countries, and may be driven, as a result, to "misallocate" managerial skill in service of these relationships, but at the expense of productivity. This stands in contrast to recent evidence of learning-by-exporting, in that we document one way in which buyer relationships might actually *reduce* supplier productivity.

References

Abowd, J. M., Creecy, R. H., and Kramarz, F. (2002). Computing person and firm effects using linked longitudinal employer-employee data. *Mimeo*.

Abowd, J. M., Kramarz, F., and Margolis, D. N. (1999). High wage workers and high wage firms. *Econometrica*, 67(2):251–334.

Abowd, J. M., Lengermann, P., and Pérez-Duarte, S. (2004). Are good workers employed by good firms? a simple test of positive assortative matching models. *Mimeo*.

Adhvaryu, A., Kala, N., and Nyshadham, A. (2018). The skills to pay the bills: Returns to on-the-job soft skills training. Technical report, National Bureau of Economic Research.

Adhvaryu, A., Nyshadham, A., and Tamayo, J. (2019). Managerial quality and productivity dynamics. *working paper*.

Alvarez, J., Benguria, F., Moser, C., and Engbom, N. (2018). Firms and the decline in earnings inequality in brazil. *American Economic Journal: Macroeconomics*, 10(1):149–189.

Amodio, F. and Martinez-Carrasco, M. A. (2018). Input allocation, workforce management and productivity spillovers: Evidence from personnel data. *The Review of Economic Studies*, 85(4):1937–1970.

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

Andrews, M., Gill, L., Schank, T., and Upward, R. (2012). High wage workers match with high wage firms: Clear evidence of the effects of limited mobility bias. *Economics Letters*, 117(3):824–827.

Atkin, D., Khandelwal, A. K., and Osman, A. (2017). Exporting and firm performance: Evidence from a randomized experiment. *The Quarterly Journal of Economics*, 132(2):551–615.

Bandiera, O., Barankay, I., and Rasul, I. (2007). Incentives for managers and inequality among workers: evidence from a firm-level experiment. *The Quarterly Journal of Economics*, 122(2):729–773.

Bandiera, O., Barankay, I., and Rasul, I. (2009). Social connections and incentives in the workplace: Evidence from personnel data. *Econometrica*, 77(4):1047–1094.

Bandiera, O., Barankay, I., and Rasul, I. (2010). Social incentives in the workplace. *The Review of Economic Studies*, 77(2):417–458.

Bassi, V. and Nansamba, A. (2019). Screening and signaling non-cognitive skills: Experimental evidence from uganda.

Blattman, C. and Dercon, S. (2016). Occupational choice in early industrializing societies: Experimental evidence on the income and health effects of industrial and entrepreneurial work. Technical report, National Bureau of Economic Research.

Bloom, N., Eifert, B., Mahajan, A., McKenzie, D., and Roberts, J. (2013). Does management matter? evidence from india. *The Quarterly Journal of Economics*, 1(51):51.

Bloom, N., Mahajan, A., McKenzie, D., and Roberts, J. (2018). *Do management interventions last? evidence from India*. The World Bank.

Bloom, N. and Van Reenen, J. (2007). Measuring and explaining management practices across firms and countries. *Quarterly Journal of Economics*, 122(4).

Bloom, N. and Van Reenen, J. (2011). Human resource management and productivity. *Handbook of labor economics*, 4:1697–1767.

Boning, B., Ichniowski, C., and Shaw, K. L. (2007). Opportunity counts: Teams and the effectiveness of production incentives. *Journal of Labor Economics*, 25(4):613–650.

Cajal Grossi, J., Macchiavello, R., and Noguera, G. (2019). International buyers' sourcing and suppliers' markups in bangladeshi garments. Technical report, working paper.

Card, D., Cardoso, A. R., Heining, J., and Kline, P. (2018). Firms and labor market inequality: Evidence and some theory. *Journal of Labor Economics*, 36(S1):S13–S70.

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

Caselli, F. (2005). Accounting for cross-country income differences. *Handbook of economic growth*, 1:679–741.

Eeckhout, J. (2018). Sorting in the labor market. *Annual Review of Economics*, 10:1–29.

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

Frederiksen, A., Kahn, L. B., and Lange, F. (2017). Supervisors and performance management systems. Technical report, National Bureau of Economic Research.

Graham, B., Imbens, G., and Ridder, G. (2014). Complementarity and aggregate implications of assortative matching: A nonparametric analysis. *Quantitative Economics*, 5(1):29–66.

Graham, B., Imbens, G., and Ridder, G. (2018). Identification and efficiency bounds for the average match function under conditionally exogenous matching. *Journal of Business & Economic Statistics*, pages 1–14.

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

Hall, R. E. and Jones, C. I. (1999). Why do some countries produce so much more output per worker than others? *The Quarterly Journal of Economics*, 114(1):83–116.

Hamilton, B. H., Nickerson, J. A., and Owan, H. (2003). Team incentives and worker heterogeneity: An empirical analysis of the impact of teams on productivity and participation. *Journal of Political Economy*, 111(3).

Hoffman, M. and Tadelis, S. (2019). People management skills, employee attrition, and manager rewards: An empirical analysis. *working paper*.

Holmstrom, B. R. and Tirole, J. (1989). The theory of the firm. *Handbook of industrial organization*, 1:61–133.

Ichino, A. and Maggi, G. (2000). Work environment and individual background: Explaining regional shirking differentials in a large italian firm. *Quarterly Journal of Economics*, 115(3):1057–1090.

Karlan, D., Knight, R., and Udry, C. (2015). Consulting and capital experiments with microenterprise tailors in ghana. *Journal of Economic Behavior & Organization*, 118:281–302.

Kremer, M. (1993). The o-ring theory of economic development. *The Quarterly Journal of Economics*, 108(3):551–575.

Lazear, E. P. and Oyer, P. (2007). Personnel economics. Technical report, National Bureau of economic research.

Lazear, E. P. and Shaw, K. L. (2007). Personnel economics: The economist's view of human resources. *Journal of economic perspectives*, 21(4):91–114.

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

Levitt, S. D., List, J. A., and Syverson, C. (2013). Toward an understanding of learning by doing: Evidence from an automobile assembly plant. *Journal of Political Economy*, 121(4):643–681.

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

Macchiavello, R. and Morjaria, A. (2015). The value of relationships: evidence from a supply shock to kenyan rose exports. *American Economic Review*, 105(9):2911–45.

McKenzie, D. and Woodruff, C. (2013). What are we learning from business training and entrepreneurship evaluations around the developing world? *The World Bank Research Observer*, 29(1):48–82.

McKenzie, D. and Woodruff, C. (2016). Business practices in small firms in developing countries. *Management Science*.

Moretti, E. and Mas, A. (2009). Peers at work. *American Economic Review*, 99(1):112–45.

Syverson, C. (2011). What determines productivity? *Journal of Economic Literature*, 49(2):326–365.

Appendix A: Production Function Estimation

In this Section we perform a simple alternative production function estimation procedure, which sheds light on the role of the underlying production technology.

Our procedure exploits the fact that we have a long panel of workers and firms, observed daily for two years. In particular, to get a measure of underlying worker and line productivity or "type", for each worker and each line we calculate their average productivity over the first three months in which they are observed in the data. To do this, we use all workers and firms in our data. We maintain a strict exogeneity assumption in the allocation of workers to lines, which in the context of production function estimation is in line with the approach of Graham et al. (2014, 2018). Under this assumption, and as long as there is enough mobility of workers across lines, this approach allows to recover an underlying measure of worker and firm productivity in the first three months of data. As shown in Table 1, the share of movers is high at more than 50% in our data, which is reassuring. We then rank workers and lines into quartiles of this baseline measure of productivity, and estimate a production function using *only* the later time periods, i.e. excluding the first three months of data. So effectively we calculate the worker and line types in the first three months of data, and then use these to estimate the production function in the later time periods.[19] For robustness, we also repeat this procedure calculating worker and line underlying productivity in the first four months of data (rather than the first three months). Specifically, we estimate the following linear in logs model:

$$ln(y_{it}) = \alpha workerQ_i + \beta lineQ_{J(i,t)} + \gamma workerQ_i \times lineQ_{J(i,t)} + x'_{it}\delta + \epsilon_{it} \tag{3}$$

where $ln(y_{it})$ is the log daily efficiency of worker i on day t; $workerQ_i$ is the quartile of average productivity of worker i as measured in the first three months of data (so note that this is not time-varying in the second period); $lineQ_{J(i,t)}$ is the quartile of the average productivity of line j where worker i is matched at time t (again estimated in the first three months only); finally, x'_{it} are time-varying controls.[20]

Equation (3) is an approximation to a Constant Elasticity of Substitution (CES) production function in logs. We expect positive estimates of α and β as we expect the output to increase in both worker and firm types. The coefficient γ on the interaction instead cor-

[19] Our choice to compute the firm and worker average productivity using the first three months of data reflects the fact that average tenure for workers in the sample is around nine months. So by using the first three months to estimate the worker type, we still have available about six months for the actual production function estimation.

[20] These include style fixed effects to account for variation in productivity due to complexity of the style and size of the order, as well as year, month and day of the week fixed effects, to account for common seasonality and growth in productivity across lines.

responds to the cross-partial in the log-form of the CES. Therefore, a negative coefficient on this interaction term would imply a negative cross-partial in levels; a coefficient of zero would be consistent with the underlying production function being Cobb-Douglas, and so with a positive cross-partial in levels; and a positive coefficient would again imply a positive cross-partial in levels and complementarity in production. We estimate regressions like these by OLS, clustering standard errors at the line level.

Appendix Table A4 reports the results: as expected, we find positive and significant estimates of both α and β. For instance, looking at columns 1 and 2, we see that an increase of one quartile in line productivity is associated with a 1.7-1.8% increase in worker productivity. In addition, we cannot reject that the interaction term is *zero*: the estimates of the interaction terms are very small in magnitudes, and not significant. Therefore, we cannot reject that the underlying production function is Cobb-Douglas. Since the cross-partial of the Cobb-Douglas is positive in levels, this means that the output of a given worker is increasing in the manager type. Table A4 further shows that the results are very similar whether we use the first three months or the first four months of data to calculate the worker and line types in the initial period.[21]

In sum, the results of this simple production function estimation procedure again confirm that we find a positive cross-partial in levels between managers and workers, so that the productivity of a given worker increases in the manager type.

[21] As discussed in Section 2, the nature of work in these production lines is such that there is no direct interaction between workers in the production process. Also, each worker has a buffer stock of material to work at. This limits the potential for complementarities or substitutabilities across workers. The literature however has pointed out the potential importance of peer pressure or social motives in creating spillover effects on productivity across workers even when the production technology does not feature teamwork (Bandiera et al., 2010; Ichino and Maggi, 2000; Moretti and Mas, 2009). We formally check for the presence of spillover effects across workers in Table A5 in the Appendix, where we add to regressions like (3) the share of workers in the highest quartile working at line j at the same time as worker i. Reassuringly, we find that the coefficient on this variable is not significant, which again suggests that production is separable across workers and any spillover effects across productivity groups are limited in our context. We note however that the standard error on the estimates of the effect of the share of workers in the highest quartile is sizeable.

Tables and Figures

Table 1: Summary Statistics

Factory	N. Observations	N. Workers	N. Movers	N. Lines
1	742,221	5,001	2,464	29
2	573,855	4,743	2,660	16
3	595,409	4,504	2,239	26
4	311,961	2,702	1,673	8
5	292,077	2,699	1,659	17
6	410,054	3,959	2,598	24
Total	2,925,577	23,608	13,293	120

Note: The data is from six factories included in the study. The data spans from July 2013 to July 2015. A "Mover" is defined as a worker who is observed at more than one production line during the sample period.

Table 2: Estimates of Sorting Pattern from AKM Regressions

Statistic	*Productivity*	*Wages*
$Var(\theta)$	0.0150	0.0105
$Var(\psi)$	0.0176	0.0001
$Var(y)$	0.2729	0.0168
$Var(\psi)/Var(\psi+\theta)$	0.6629	0.0140
$Corr(\psi,\theta)$	-0.1604	-0.0221

Note: Table 2 reports the estimates from the two-way fixed effects estimation procedure in Abowd et al. (1999) using productivity and wages as an outcome. The regressions are estimated by OLS, clustering standard errors by line. The data includes daily worker-level data from six garment factories. The data spans over two years, from July 2013 to July 2015. Our sample consists of 120 production lines and 23,608 workers. We keep the largest connected set between lines and workers for each factory, which includes around 98% of the observations. The statistics reported in the table are first computed by estimating the econometric model separately for each factory, and then a weighted average across factories is calculated, weighted by the sample share of each factory.

Table 3: Daily Probability of Moves Between Worker and Line Types

		Worker	
		High	Low
Line	High	0.0019	0.0088
	Low	0.0075	0.0056

Note: The table reports the daily probability of observing the following events: (i) a H-type worker moving to a H-type line; (ii) a H-type worker moving to a L-type line; (iii) a L-type worker moving to a H-type line; (iv) a L-type worker moving to a L-type line. Workers and lines are assigned to the H- and L-types based on quartiles of the fixed effects estimated from the AKM procedure. See Table 2 for details of the estimation. Workers and lines in the top quartile of the fixed effects are assigned to the H-type; workers and lines in the bottom quartile of the fixed effects are assigned to the L-type.

Table 4: Heterogeneity by Different Groups of Workers

	$Corr(\theta, \psi)$	
	High	*Low*
Grade	-0.240	-0.175
Salary	-0.169	-0.168
All Workers		
Outside Homeline	-0.165	
Homeline	-0.126	
Movers		
Outside Homeline	-0.169	
Homeline	-0.169	

Note: The Table reports the estimates from the two-way fixed effects estimation procedure in Abowd et al. (1999). The regressions are estimated by OLS, clustering standard errors by line. The data includes daily worker-level data from six garment factories. The data spans from July 2013 to July 2015. Our sample consists of 120 production lines and 23,608 workers. We keep the largest connected set between lines and workers for each factory, which includes around 98% of the observations. High grade workers are those in the top skill level, called “A+++”, and correspond to roughly 30% of the sample. Workers are split into high and low salary groups based on median salary. The statistics reported in the table are first computed by estimating the econometric model separately for each factory, and then a weighted average across factories is calculated, weighted by the sample share of each factory.

Table 5: Supply-Chain Constraints Reported by Production Managers

Panel A: Concerns Related to Meeting Targets	Mean
Monetary loss to firm from falling behind with order	51%
Firm risks losing customer from falling behind	33%
Own line has fallen behind with order	19%
Panel B: Steps Taken to Meet Targets	Mean
Would move workers to avoid falling behind	91%
Would move good performers to slow lines	97%
Would move poor performers to slow lines	3%

Note: data is from the survey of production managers conducted in the six factories in the study.

Figure 1A: PDF of Worker Productivity

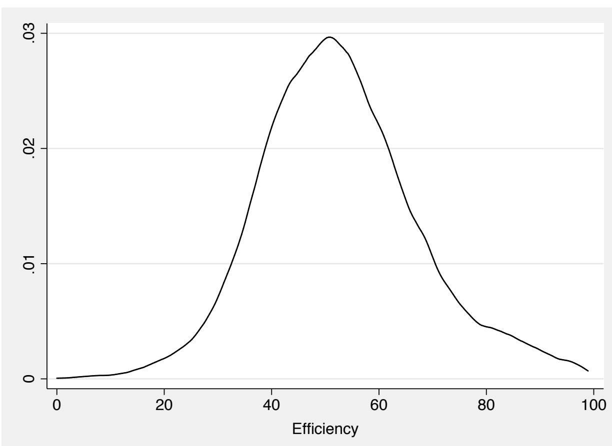
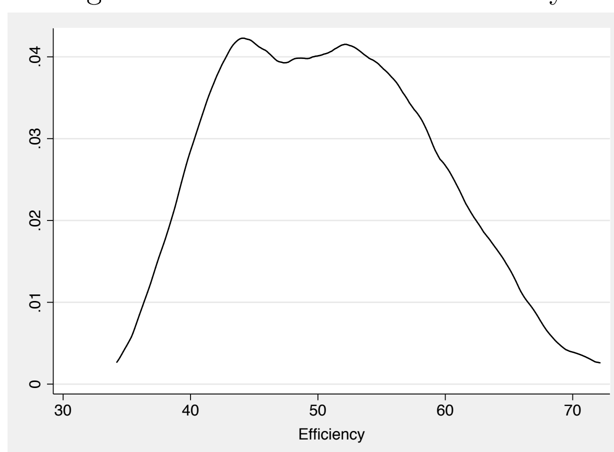
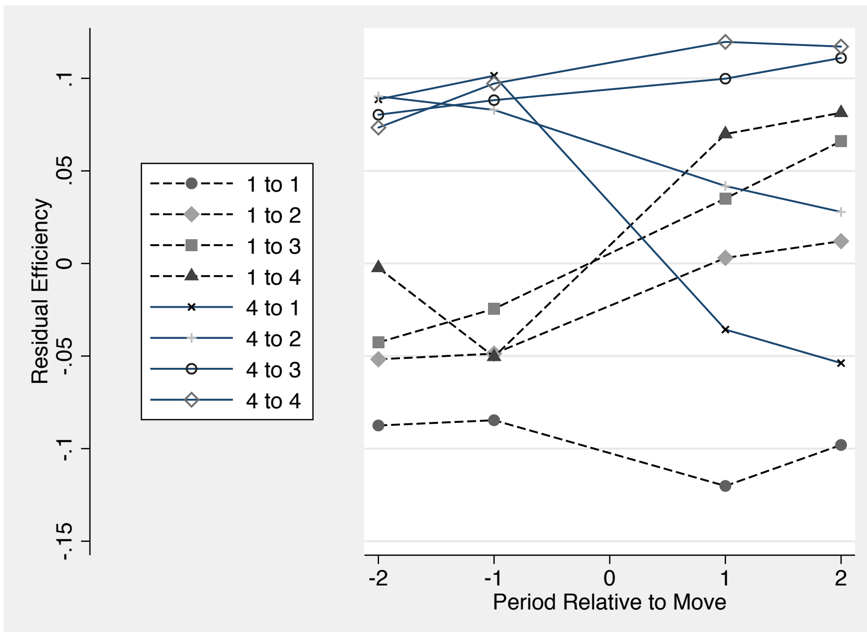


Figure 1B: PDF of Line Productivity



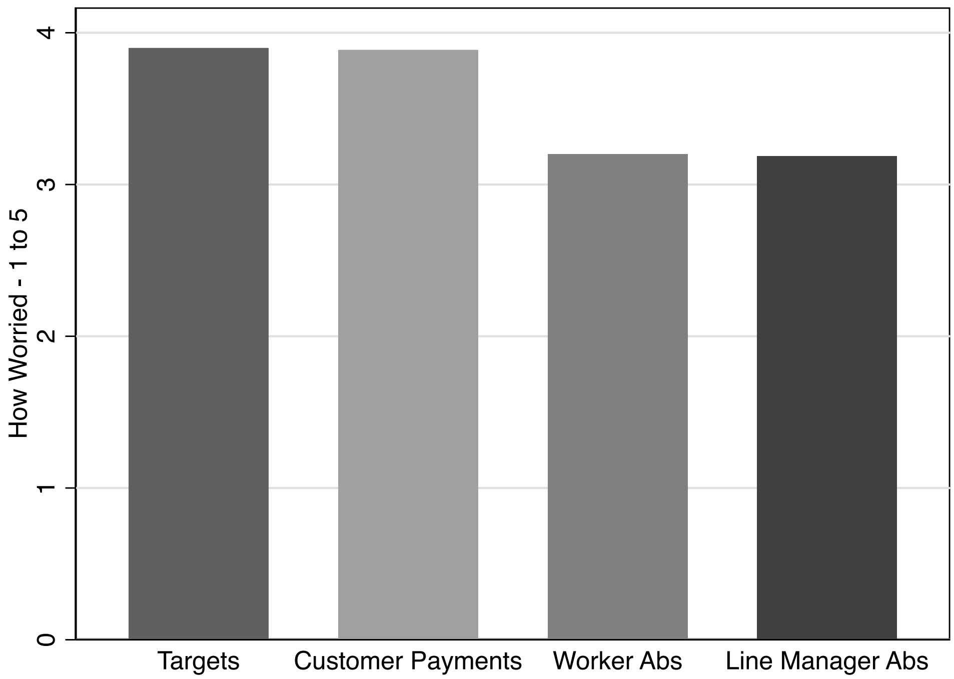
Note: Figure 1A and 1B show the distribution of the average efficiency of workers and lines, across the six factories included in our study. The data spans from July 2013 to July 2015. The sample is defined in Table 1. Our measure of productivity is daily Efficiency, which equals the percentage of the target quantity of a particular garment that is achieved per day. The target quantity is calculated using a measure of garment complexity called the standard allowable minute (SAM), which is equal to the number of minutes that a particular garment should take to produce.

Figure 2: Event Study Around Moves



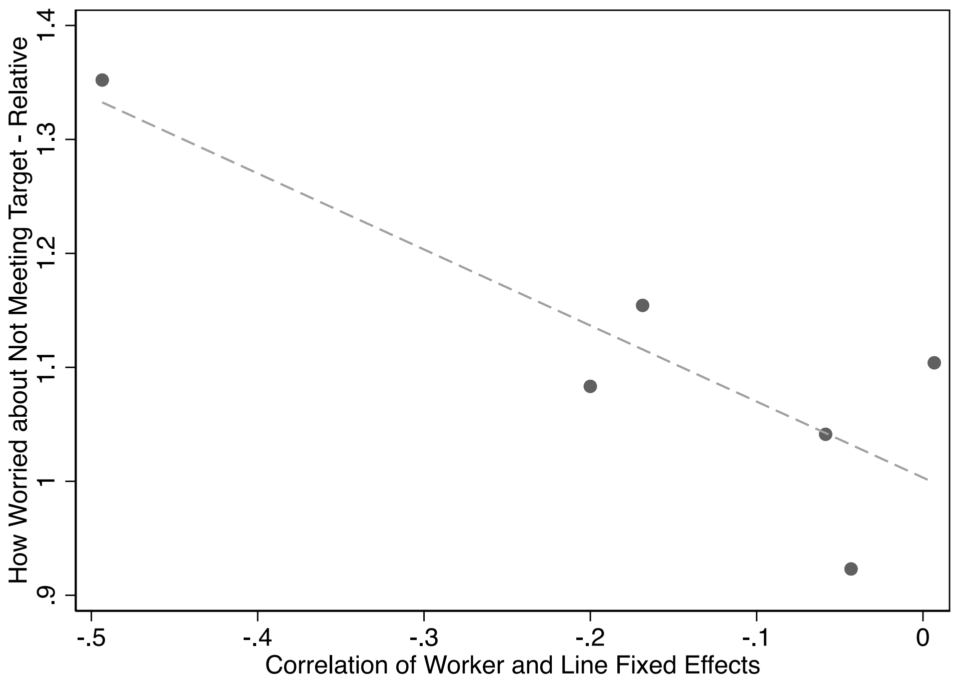
Note: we rank movers in terms of: (i) quartiles for the average residual productivity of the line they moved away from; and (ii) quartiles for the average residual productivity of the line they moved to. The Figure then plots the average residual efficiency of the mover on the y-axis, computed 6 to 10 days ($Period = -2$) and 1 to 5 days ($Period = -1$) before the move from the old line, and 1 to 5 days ($Period = 1$) and 6 to 10 days ($Period = 2$) after the move to the new line, on the x-axis. The graph only considers moves away from either lines in the top quartile of co-worker residual efficiency (i.e. lines in quartile 4) or lines in the bottom quartile of co-worker residual efficiency (i.e. lines in quartile 1). The sample for the graph is restricted to the balanced sample of workers continuously employed at the origin line for at least 10 days prior to the move, and continuously employed at the destination line for at least 10 days after the move. To calculate worker-level residual efficiency we run a regression of log daily efficiency of the worker on: factory fixed effects; year, month and day of the week fixed effects; style fixed effects; tenure (days) of the worker in the factory; tenure (days) of the worker at the line; finally, we include the experience of the line in producing the current style in the current production run, as measured by the number of consecutive days spent producing that style. Standard errors are clustered by line in this regression. We use this regression to calculate residual efficiency of each worker.

Figure 3: Concerns of Production Managers



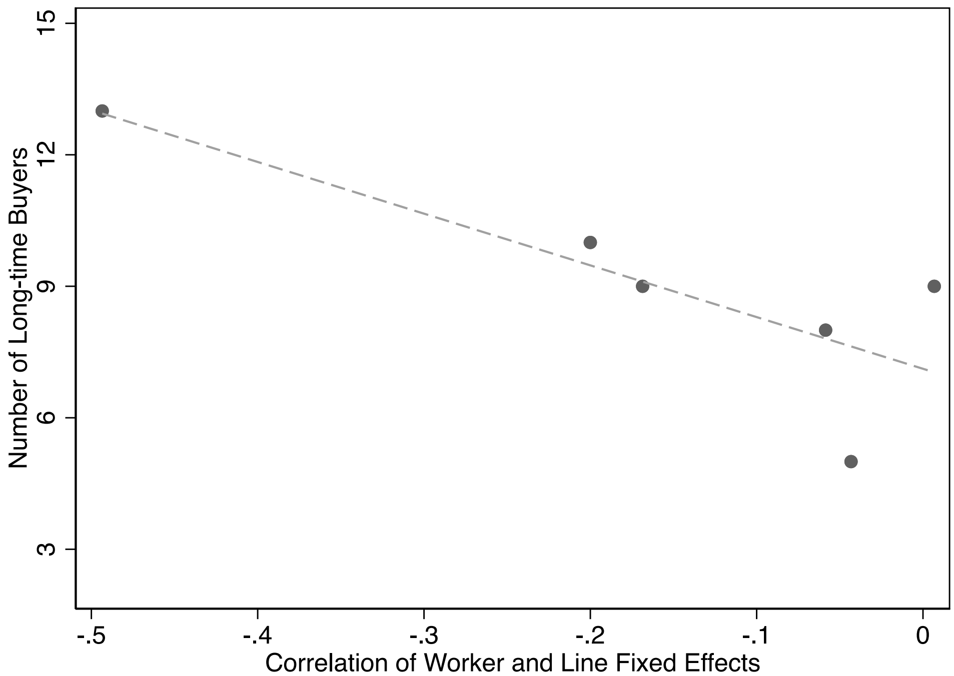
Note: data is from the survey of production managers conducted in the six factories in the study. We asked managers to indicate the relative importance of four potential concerns in their operations: (i) lines not meeting their target/running slow; (ii) worker absenteeism; (ii) line manager absenteeism and (iv) customers not paying for their order on time. Managers were asked to use a 1 to 5 scale to indicate the importance of each concerns, with 1 meaning "not worried at all", and 5 meaning "very worried". The Figure reports the average for each question.

Figure 4: Heterogeneity in NAM by Manager Concerns



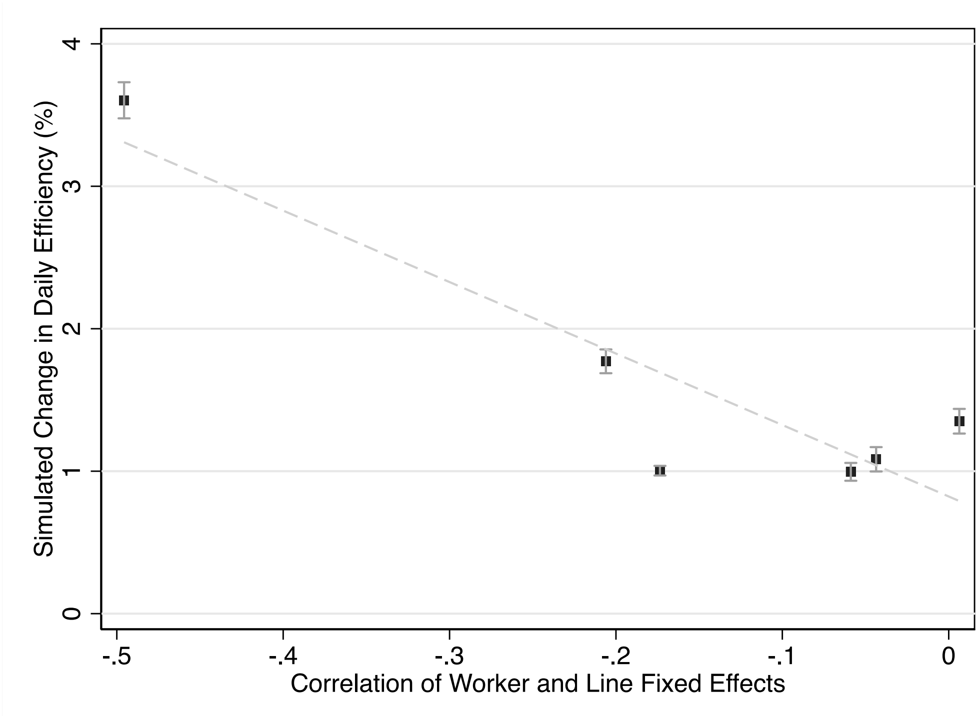
Note: The Figure plots on the x-axis the estimated correlations between the worker and firm fixed effects from the two-way fixed effects estimation procedure described in Section 3; and on the y-axis the factory-level relative average importance of concerns related to falling behind with orders and not meeting targets. Specifically, for each factory, we divide the average reported concern about lines not meeting deadlines by the within-factory average of all concerns managers were asked about. We then plot this measure on the y-axis. The Figure also shows the line of best fit from an OLS regression of the relative average importance of concerns related to falling behind with orders and not meeting targets, on the degree of NAM in the factory.

Figure 5: Heterogeneity in NAM by Prevalence of Long-time Buyers



Note: The Figure plots on the x-axis the estimated correlations between the worker and firm fixed effects from the two-way fixed effects estimation procedure described in Section 3; and on the y-axis the number of long-time buyers that placed orders at the factory over the July 2013-July2015 sample period. "Long-time buyers" are those who placed a number of orders to the firm in the top tercile. Long-time buyers are identified from a buyer dataset which includes all orders placed to the firm from late 2011 to late 2015. The Figure also shows the line of best fit from an OLS regression of the number of long-time buyers, on the degree of NAM in the factory.

Figure 6: Simulated Productivity Gains from Labor Reallocation



Note: The Figure plots the simulated productivity gains from implementing the perfect positive sorting allocation, across the six factories in our data, against the estimated correlation between the manager and worker fixed effects from the AKM model. The simulation was conducted as follows: a day is randomly extracted from the sample; on that day, the allocation of workers to lines is observed, together with the fixed effects of the workers and the lines; the perfect positive assortative matching allocation is then implementied by assigning the workers with the highest fixed effects to the lines with the highest fixed effects and so on, respecting the line sizes observed in the data. Worker productivity is then predicted using the estimated equation (1), but with workers and lines matched following perfect positive assortative matching. Log efficiency is then exponentiated to recover the counterfactual productivity in levels. This is then summed across all workers and all lines. This procedure is repeated on 1,000 randomly extracted days. We report the mean increase in daily efficiency across the simulation, together with the 95% confidence intervals, where bootstrap standard errors are used to construct the confidence intervals. The Figure also shows the line of best fit from an OLS regression of the average efficiency gain from reallocation, on the degree of NAM in the factory.

Appendix Tables and Figures

Table A1: Distribution of Number of Lines Workers are Observed at

N. Lines seen at	Freq.	Percent
1	10,315	43.69
2	5,781	24.49
3	3,434	14.55
4	1,826	7.73
5	1,088	4.61
6	654	2.77
7	299	1.27
8	112	0.47
9	57	0.24
10	26	0.11
11	7	0.03
12	5	0.02
13	4	0.02
Total	23,608	100

Note: The table reports the distribution of number of lines that workers in the sample are observed at, during the sample period. The data is from six factories included in the study. The data spans from July 2013 to July 2015.

Table A2: Do Changes in Productivity Predict the Direction of Moves?

Panel A	D[1 to 1]	D[1 to 2]	D[1 to 3]	D[1 to 4]
Δ Efficiency	0.0218	-0.0112	-0.00589	-0.00472
	(0.0246)	(0.0249)	(0.00905)	(0.00786)
Observations	9,156	9,156	9,156	9,156

Panel B	D[2 to 1]	D[2 to 2]	D[2 to 3]	D[2 to 4]
Δ Efficiency	0.0262	-0.0215	0.00482	-0.00959
	(0.0386)	(0.0468)	(0.0175)	(0.0324)
Observations	6,683	6,683	6,683	6,683

Panel C	D[3 to 1]	D[3 to 2]	D[3 to 3]	D[3 to 4]
Δ Efficiency	0.00617	0.00145	0.00316	-0.0108
	(0.0251)	(0.0414)	(0.0335)	(0.0168)
Observations	6,616	6,616	6,616	6,616

Panel D	D[4 to 1]	D[4 to 2]	D[4 to 3]	D[4 to 4]
Δ Efficiency	-0.0151	0.0215	0.0108	-0.0173
	(0.0119)	(0.0330)	(0.00941)	(0.0368)
Observations	5,919	5,919	5,919	5,919

Note: we rank movers in terms of: (i) quartiles for the average residual productivity of the line they moved away from; and (ii) quartiles for the average residual productivity of the line they moved to. The ranks are based on average efficiency in the two weeks before the move, and the two weeks after the move. The Table reports the results of OLS regressions where the dependent variables are the conditional probabilities of moving from a line in the X quartile to a line in the Y quartile. Panel A only considers moves away from lines in quartile 1; Panel B only considers moves away from lines in quartile 2; Panel C only considers moves away from lines in quartile 3; and Panel D only considers moves away from lines in quartile 4. For example, the variable D[1 to 1] takes value one if the worker moves from a line in quartile 1 to another line in quartile 1, and zero otherwise. We regress such dummy variables on the change in average worker-level log efficiency between the second week and the first week before the move. The sample for the graph is restricted to the sample of workers continuously employed at the origin line for at least two weeks prior to the move. All regressions further control for the following covariates: factory fixed effects; style of the origin and destination line fixed effects; tenure (days) in the factory; tenure in the factory squared and cubic; tenure (days) on the current line; tenure on the current line squared and cubic; number of days the line has been working on a specific style; days on a specific style squared and cubic. The Table reports OLS regression coefficients, with standard errors clustered at the line level in parentheses.

Table A3: Contribution of Worker, Line and Match Effects to Explaining Productivity

N	2,925,577	2,925,577	2,925,577	2,925,577	2,925,577
R^2	0.2703	0.2949	0.3281	0.3324	0.3557
R^2 Adjusted	0.2699	0.2945	0.3223	0.3266	0.3433
Time FE	Yes	Yes	Yes	Yes	Yes
Manager FE	No	Yes	No	Yes	Yes
Worker FE	No	No	Yes	Yes	Yes
Match Effects	No	No	No	No	Yes

Note: this table reports estimates from the two-way fixed effects estimation procedure in Abowd et al. (1999) using productivity as outcome. The regressions are estimated by OLS, clustering standard errors by line. The data includes daily worker-level data from six garment factories. The data spans over two years, from July 2013 to July 2015. Our sample consists of 120 production lines and 23,608 workers. We keep the largest connected set between lines and workers for each factory, which includes around 98% of the observations. The statistics reported in the table are first computed by estimating the econometric model separately for each factory, and then a weighted average across factories is calculated, weighted by the sample share of each factory.

Table A4: Production Function Estimates

	(1) *3 Months*	(2) *3 Months*	(3) *4 Months*	(4) *4 Months*
	ln(efficiency)	ln(efficiency)	ln(efficiency)	ln(efficiency)
Line Type	0.0173***	0.0182***	0.0134**	0.0161***
	(0.00582)	(0.00667)	(0.00544)	(0.00613)
Worker Type	0.0102***	0.00942**	0.0116***	0.0124***
	(0.00382)	(0.00426)	(0.00366)	(0.00410)
Line Type X Worker Type	-0.000890	-0.000699	-0.000596	-0.000877
	(0.000695)	(0.000692)	(0.000598)	(0.000679)
Observations	1,396,661	919,677	1,123,444	916,741
Mean of Dv	3.853	3.853	3.858	3.858
Sample of Workers	All	Movers	All	Movers

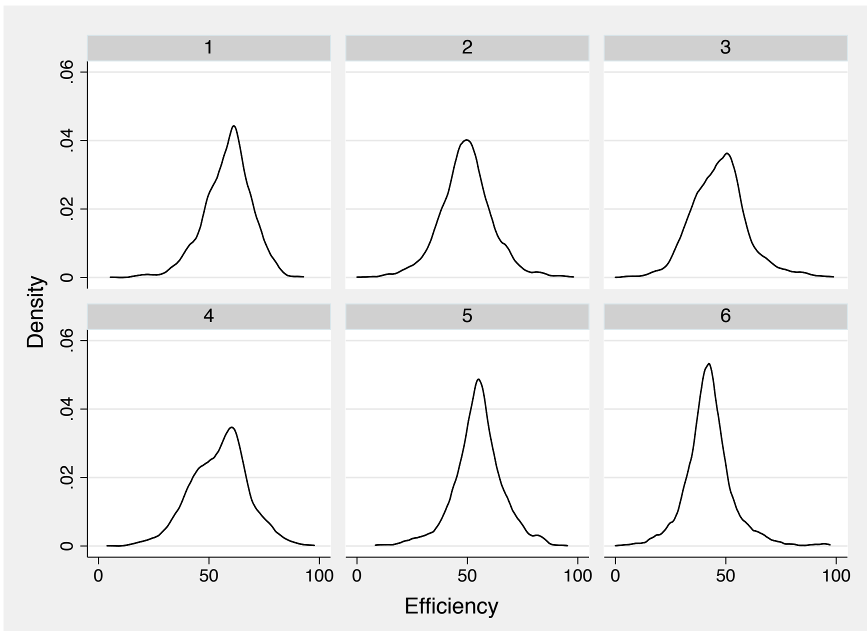
Note: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. OLS regression coefficients. Standard errors are clustered at the line level. The dependent variable is worker log daily efficiency. We include style fixed effects to account for variation in productivity due to complexity of the style and size of the order, as well as year, month and day of the week fixed effects, to account for common seasonality and growth in productivity across lines. The Worker Type and Line Type are estimated by taking worker-level and line-level averages in the first three months (columns 1-2) or four months (columns 3-4) of data. The estimation of the production function is then performed on the later time periods, and so excluding the first three months of data (columns 1-2) or the first four months of data (columns 3-4).

Table A5: Production Function Estimates: Controlling for Share of High Types

	(1) *3 Months*	(2) *3 Months*	(3) *4 Months*	(4) *4 Months*
	ln(efficiency)	ln(efficiency)	ln(efficiency)	ln(efficiency)
Line Type	0.0197**	0.0194*	0.0217***	0.0206**
	(0.00803)	(0.0104)	(0.00788)	(0.00985)
Worker Type	0.0103***	0.0108**	0.0121***	0.0139***
	(0.00382)	(0.00464)	(0.00362)	(0.00454)
Line Type X Worker Type	-0.000918	-0.000887	-0.000700	-0.00111
	(0.000690)	(0.000712)	(0.000588)	(0.000697)
Share of High Type	-0.0137	-0.0108	-0.0483	-0.0216
	(0.0382)	(0.0473)	(0.0410)	(0.0486)
Observations	1,396,661	919,677	1,123,444	916,741
Mean of Dv	3.853	3.853	3.858	3.858
Sample	All	Movers	All	Movers

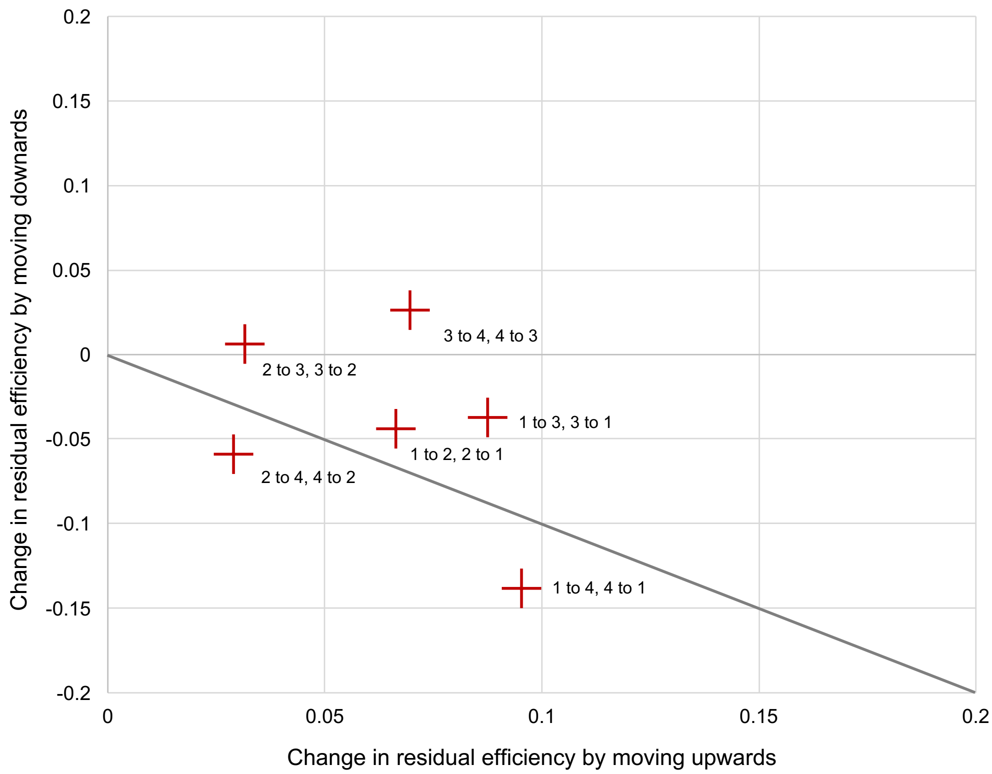
Note: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. OLS regression coefficients. Standard errors are clustered at the line level. The dependent variable is worker log daily efficiency. We include style fixed effects to account for variation in productivity due to complexity of the style and size of the order, as well as year, month and day of the week fixed effects, to account for common seasonality and growth in productivity across lines. The Worker Type and Line Type are estimated by taking worker-level and line-level averages in the first three months (columns 1-2) or four months (columns 3-4) of data. The estimation of the production function is then performed on the later time periods, and so excluding the first three months of data (columns 1-2) or the first six months of data (columns 3-4).

Figure A1: Dispersion in Worker Productivity, by Factory



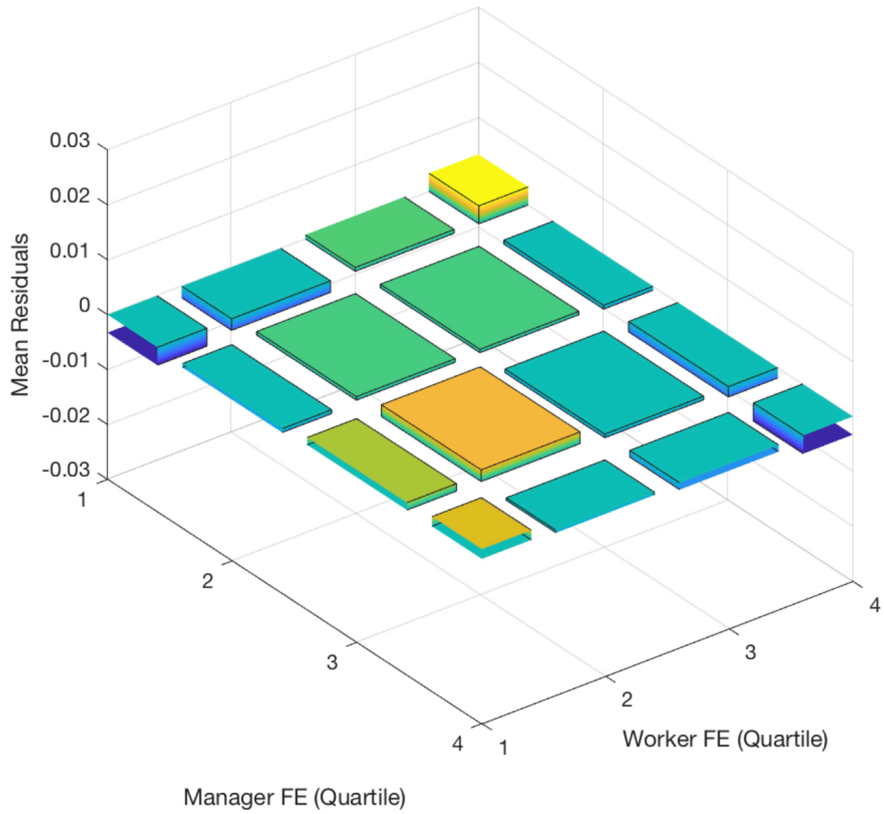
Note: Figure A1 shows the distribution of the average efficiency of the worker by factory. The data includes daily worker-level data from 6 garment factories. The data spans over two and a half years, from July 2013 to July 2015. Our sample consists of 120 production lines. A typical production line has between 65-70 workers which usually corresponds to one worker per machine. Our measure of productivity is Efficiency, which is equal to the garments produced divided by the target quantity of that particular garment. The target quantity is calculated using a measure of garment complexity called the standard allowable minute (SAM), which is equal to the number of minutes that a particular garment should take to produce.

Figure A2: Symmetry Test for Endogenous Mobility



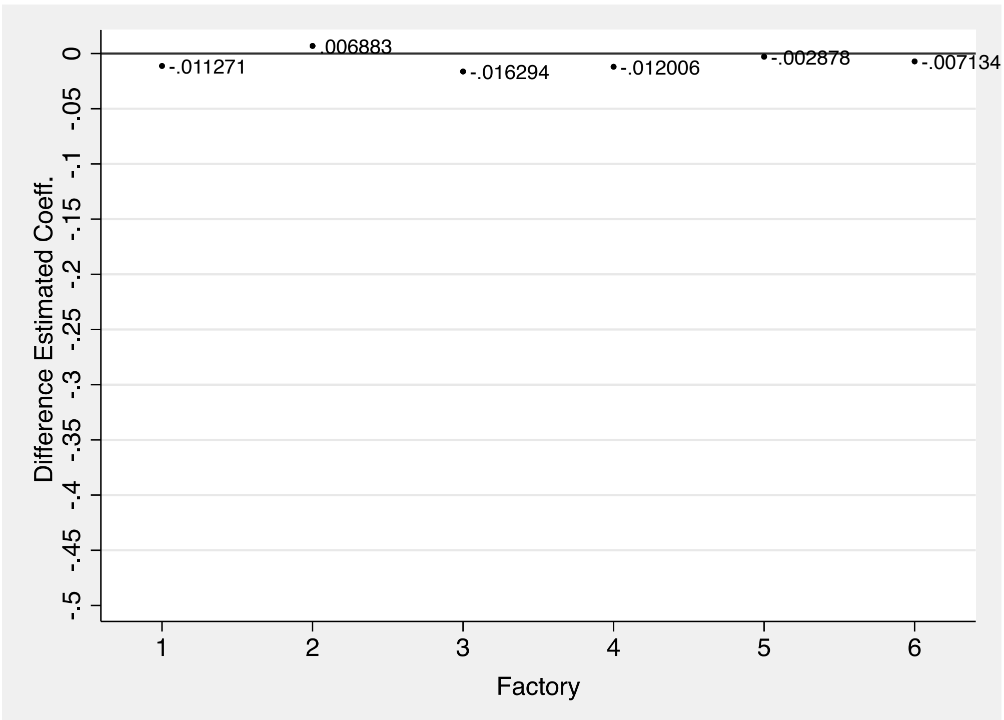
Note: we rank movers in terms of: (i) quartiles for the average residual productivity of the line they moved away from; and (ii) quartiles for the average residual productivity of the line they moved to. The ranks are based on average efficiency in the two weeks before the move, and the two weeks after the move. The Figure then plots the average change in residual efficiency for movers from lines in quartile X to quartile Y, against the change in residual efficiency for movers in the opposite direction. So for example, the point labelled "2 to 3, 3 to 2" corresponds to the average change for movers from lines in quartile 2 to quartile 3, plotted against the change for movers from lines in quartile 3 to quartile 2. The sample for the graph is restricted to the balanced sample of workers continuously employed at the origin line for at least two weeks prior to the move, and continuously employed at the destination line for at least two weeks after the move. To calculate worker-level residual efficiency we run an OLS regression of log daily efficiency of the worker on: factory fixed effect; year, month and day of the week fixed effects; tenure (days) in the factory; tenure (days) at the line; number of days the line has been working on a particular style. We use this regression to calculate residual efficiency of each worker.

Figure A3: Mean Residuals by Quartile of Worker and Line Fixed Effects



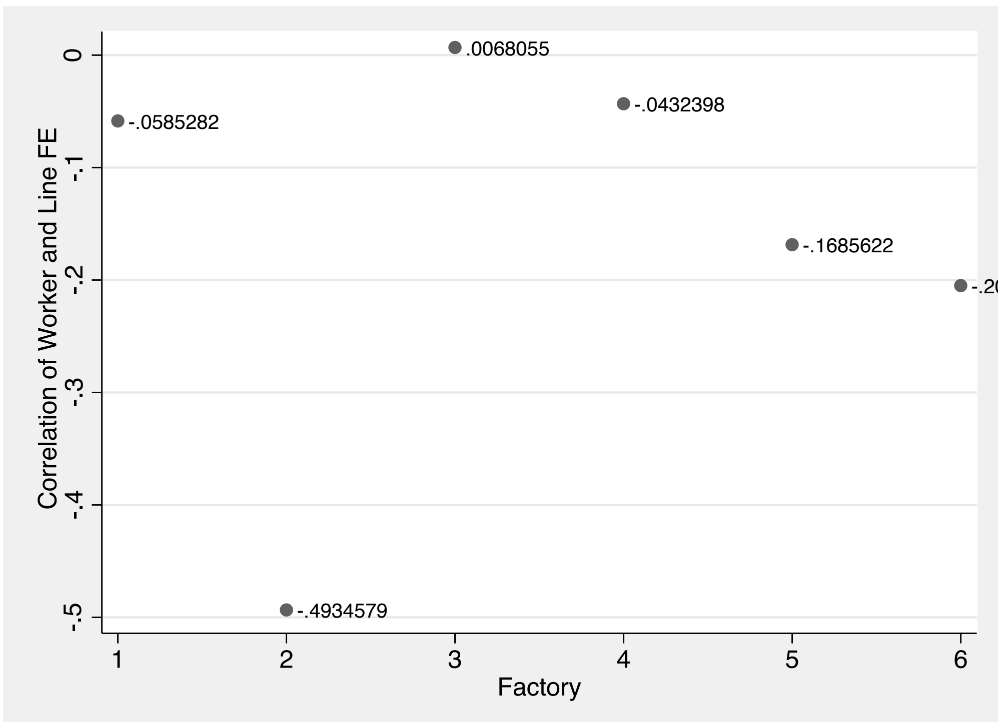
Note: this figure reports estimates from the two-way fixed effects estimation procedure in Abowd et al. (1999) using productivity as outcome. Specifically, the figure reports mean residuals by quartile of the estimated worker and line fixed effects. The regressions are estimated by OLS, clustering standard errors by line. The data includes daily worker-level data from six garment factories. The data spans over two years, from July 2013 to July 2015. Our sample consists of 120 production lines and 23,608 workers. We keep the largest connected set between lines and workers for each factory, which includes around 98% of the observations. The statistics reported in the table are first computed by estimating the econometric model separately for each factory, and then a weighted average across factories is calculated, weighted by the sample share of each factory.

Figure A4: Correlation of Worker and Line FE by Factory: Difference between the Uncorrected and Corrected Estimates (Andrews et al., 2008)



Note: Figure A4 reports the difference of the estimated correlation between Worker and Line FE following the procedure suggested by Andrews et al. (2008) and OLS (two-way fixed effects estimation procedure in Abowd et al., 1999), for each factory. The data includes daily worker-level data from 6 garment factories. The data spans over two and a half years, from July 2013 to July 2015. Our sample consists of 120 production lines. A typical production line has between 65-70 workers which usually corresponds to one worker per machine. We keep the largest connected set between lines and workers for each factory.

Figure A5: AKM Estimates: Correlation of Worker and Line FE by Factory



Note: Figure A5 reports the estimates from the two-way fixed effects estimation procedure in Abowd et al. (1999) for each factory. The data includes daily worker-level data from 6 garment factories. The data spans over two and a half years, from July 2013 to July 2015. Our sample consists of 120 production lines. A typical production line has between 65-70 workers which usually corresponds to one worker per machine. We keep the largest connected set between lines and workers for each factory. We include style fixed effects to account for variation in productivity due to complexity of the style and size of the order, as well as year, month and day of the week fixed effects, to account for common seasonality and growth in productivity across lines.