November 2019

Manufacturing Productivity with Worker Turnover

Ken Moon, Patrick Bergemann, Daniel Brown, Andrew Chen, James Chu, Ellen A. Eisen, Gregory M. Fischer, Prashant Loyalka, Sungmin Rho, Joshua Cohen

Abstract

We find that rapid worker turnover significantly disrupts the productivity of responsive manufacturers. Our study uses a uniquely rich dataset drawn from China-based FATP (final assembly, testing, and packaging) facilities that produce millions of units of consumer electronic goods weekly yet exhibit high worker turnover exceeding 300% annually. The data cover the firm's weekly production plans, 52,214 workers' compensations and assignments, and assembly station productivity. To study managerial prescriptions, we extend the classical production planning problem to include endogenous worker turnover as an Experience-Based Equilibrium and use advances in reinforcement learning and approximate dynamic programming to estimate and simulate our model. Our empirical analyses exploit instrumental variables, including the firm's demand forecasts as demand shifters". We find that turnover's impact on yield waste is conservatively $146-178M, and that a well-calibrated wage increase reduces the manufacturer's variable production costs (including wages) by up to 21%, or $594M for the product we study. The wage increase reduces the firm's reliance on a larger workforce and overtime to hedge against yield disruptions from turnover; it stabilizes a leaner workforce and improves both production reliability and exibility. In settings where performance depends on workers repeating known tasks in coordinated groups, our results suggest that firms responsively matching supply to demand can pay a steep price for a disruptively turnover-prone workforce.

Keywords: Data-driven workforce planning; Empiricle operations management; Employee turnover; Experience-Based Equilibrium; Productivity; Reinforcement learning; Structural estimation.

Working Paper

November 2019

reap.fsi.stanford.edu

REAP Rural Education Action Program

Manufacturing Productivity with Worker Turnover*

Ken Moon[a], Patrick Bergemann[b], Daniel Brown[c], Andrew Chen[d], James Chu[e], Ellen A. Eisen[c], Gregory M. Fischer[f], Prashant Loyalka[e], Sungmin Rho[g], Joshua Cohen[h]

a: The Wharton School; b: Booth School of Business; c: UC Berkeley; d: Apple Inc.; e: Stanford University; f: London School of Economics; g: IHEID; h: Apple University;

We find that rapid worker turnover significantly disrupts the productivity of responsive manufacturers. Our study uses a uniquely rich dataset drawn from China-based FATP (final assembly, testing, and packaging) facilities that produce millions of units of consumer electronic goods weekly yet exhibit high worker turnover exceeding 300% annually. The data cover the firm's weekly production plans, 52,214 workers' compensations and assignments, and assembly station productivity. To study managerial prescriptions, we extend the classical production planning problem to include endogenous worker turnover as an Experience-Based Equilibrium and use advances in reinforcement learning and approximate dynamic programming to estimate and simulate our model. Our empirical analyses exploit instrumental variables, including the firm's demand forecasts as "demand shifters". We find that turnover's impact on yield waste is conservatively \$146-178M, and that a well-calibrated wage increase reduces the manufacturer's variable production costs (including wages) by up to 21%, or \$594M for the product we study. The wage increase reduces the firm's reliance on a larger workforce and overtime to hedge against yield disruptions from turnover; it stabilizes a leaner workforce and improves both production reliability and flexibility. In settings where performance depends on workers repeating known tasks in coordinated groups, our results suggest that firms responsively matching supply to demand can pay a steep price for a disruptively turnover-prone workforce.

Key words: Data-driven workforce planning; Empirical operations management; Employee turnover; Experience-Based Equilibrium; Productivity; Reinforcement learning; Structural estimation.

1. Introduction

Rapid worker turnover, commonplace in early 20th century US manufacturing (Schlichter (1919), Brandes (1976), Jacoby (1983), Owen (1995)), has re-emerged as a compelling challenge in modern

* We gratefully acknowledge support from our data sponsor and from the Wharton Dean's Reseach Fund (Moon). We thank our editors, three referees, Hamsa Bastani, Rob Bray, Gérard Cachon, Morris Cohen, Cary Laird, and Serguei Netessine for comments and discussions. Participants at the Wharton/Penn Industrial Organization lunch, 2017 INFORMS Annual Meeting, 2018 MSOM Supply Chain Management SIG Meeting, 2018 Wharton Empirical OM Workshop, 2019 Management Science Workshop in Chile, and Berkeley Haas OITM seminar provided helpful feedback. Correspondence: kenmoon@wharton.upenn.edu

production processes.[1] Today's largest manufacturers employ hundreds of thousands to millions of workers annually. Yet at the China-based production facilities studied in this paper, the standing workforce is replaced three times per annum. Despite voluntary turnover rates exceeding 300%, these same facilities output tens of billions USD in complex consumer electronic devices.

Elevated turnover presently pervades all levels of the supply chain. A recent survey of Chinese firms underscores concerns about worker turnover (CEES (2017)). Similarly, Shih (2014) places significant turnover as the top-of-list concern for US manufacturers seeking to "reshore" production, citing as an example GE's US-based Appliance Park "hir[ing] 6,500 workers to yield the required 2,500 employees to begin volume production" and facing a further 23% turnover rate for the year. One of the largest US wholesale distributors of pharmaceuticals shared with us that annual voluntary turnover at its distribution centers had risen from historically single-digit percentages to upwards of 40% annually, even as extensive automation, data-driven feedback, and incentive-based compensation have boosted labor productivity.[2] More generally, managing turnover has remained or become critical for operations across the spectrum of industries, including retail services (65% annual turnover), transport (94% for truck operators), hospitality (74%), call centers (30-45%), warehousing and utilities (41%), and staffing-sensitive healthcare (19.2%).[3]

This paper's objective is to understand the effects of worker turnover on the manufacturer. Empirical studies have shown that human capital and its organization matter in manufacturing (e.g., Egelman et al. (2017)). Yet to our knowledge, limited attention has focused on characterizing and empirically assessing how worker turnover disrupts a manufacturer's production planning

[1] US manufacturing's annual separation rates (i.e., quits, layoffs, and terminations per 100 employees) fell from averaging 117.2 over the decade 1910-20 to between 47 to 64 across 1950-70, owing to reduced quits. Owen (1995) argues that labor specialization led manufacturers to promote employment relationships and benefits supported by centralized human resources.

[2] Conversation between the distributor, the first author, and Morris Cohen (April 2018).

[3] Korn Ferry, top retailer survey, Nov. 2016; American Trucking Associations, Jun. 2018; Bureau of Labor Statistics, "Job Openings and Labor Turnover - Jan. 2018," Table 16; Reynolds, P., "Exploring Call Center Turnover Numbers," Quality Assurance & Training Connection, Winter 2015; Compdata Surveys, "Compensation Data Healthcare 2015".

and performance. Moreover, we expect that the effect size depends on setting (see, e.g., Ton and Huckman (2008)). Relevantly for this paper, manufacturers producing innovative products face inherently uncertain demand, and worker turnover potentially undermines the production capacity and the volume flexibility that critically enable them to responsively match supply to demand. In contrast, manufacturers of functional products with predictable demands may find turnover less pernicious, and indeed well-cited studies find little evidence of sizeable productivity effects from worker turnover when orders are known or reliably secured in advance.[4]

Regarding the high turnover in the manufacturing facilities we study, perhaps the closest historical analogue comes from the Ford Motor Company. In 1913, turnover reached 370%, prompting Henry Ford to introduce his famous wage hike, the five-dollar day for 8 hours of work (Raff and Summers (1987)). Yet Ford's "[j]obs were sufficiently menial that it is unlikely that high turnover was extremely costly or that worker selection effects were important" (Raff and Summers (1987), p.S60).[5] Though Ford's turnover reduced massively to 54% by 1914, the program's efficacy remains debated. Merely 6% of the five-dollar day program's cost was offset by reduced training and recruitment from turnover (pp. S79-80), and empirical evidence finds that reductions in shirking and moral hazard inadequately justify the costs of compensating workers at supra-competitive (i.e., hiked) wages (e.g., Raff and Summers (1987), Cappelli and Chauvin (1991)).

Instead of mitigating shirking or avoiding training costs, could Ford's higher wages have primarily addressed turnover's disruptions to productivity? With different data from the present day, we explore the possibility that such operational disruptions ought to motivate innovative manufacturers like Ford to reduce and control worker turnover. We thus consider a distinctive case, rooted in operations management, for highly compensating low wage workers.

[4] See, e.g., Argote et al. (1990), p.151, citing "no evidence that turnover contributed to depreciation [of output performance] in the Liberty Ship environment [where] the jobs of production workers . . . were highly standardized and designed to be low in skill requirements." Arthur (1994) (Table 5 on p.682) finds no direct effect except when workers exercise significant discretion in how to complete tasks, and Shaw et al. (2005) find attenuated effects.

[5] Ford's assembly line built on Taylor (1911)'s scientific management in standardizing and engineering ("deskilling") tasks, while segregating planning from production. See Taneja et al. (2011) on Taylor; Hounshell (1984) on Fordism.

Our study is both empirical and analytical. We compile rich data regarding manufacturing worker turnover, drawn from FATP (final assembly, testing, and packaging) facilities that output millions of units of consumer electronic goods per week. Our panel dataset combines (A) human-resources data on the characteristics, line assignments, and compensations of 52,214 assembly workers from hire to turnover (509,498 worker-weeks), (B) volume and yield records tracing to stations on individual assembly lines, and (C) the firm's rolling-horizon production plans and forecasts. Weekly planning and forecast data are critical to our study of the responsive manufacturer handling variable product demand. Additionally, compensation data accounts for workers' pay-based incentives when investigating the operational drivers of worker turnover. We develop a structural model of production planning that incorporates endogenous worker turnover as an Experience-Based Equilibrium (Fershtman and Pakes (2012)), made tractable (i.e., handling a state transition matrix exceeding 1TB in sparse representation and estimated using over 12 billion simulations) by combining large-scale convex optimization with reinforcement learning (Osband et al. (2017)) and approximate dynamic programming (De Farias and Van Roy (2003, 2004), Farias et al. (2012)).

Our unique dataset, together with our econometric and structural empirical methods, enables us to consider four basic questions: (1) Given the simplified tasks and standardized practices associated with assembly manufacturing, what is the magnitude of productivity losses arising from worker turnover? (2) Through what mechanisms does worker turnover affect manufacturing productivity, and (3) what are the performance implications for the responsive manufacturer? (4) Lastly, what is the operational goal of managing the workforce as part of production planning, and how should managers control turnover's impact on performance?

Section 2 develops our key hypotheses concerning the magnitude and mechanisms of worker turnover's effects on assembly line productivity. Section 3 introduces our data, and Section 4 tests the empirical hypotheses. Sections 5 and 6 respectively estimate and simulate the responsive manufacturer's production planning problem, which we formulate as a stochastic control problem that embeds workers' endogenous turnover as an equilibrium response to their incentives.

2. Theory and Hypotheses

Managing disruptions and uncertainties in production is a major topic in supply chain operations. Manufacturers should align their production capabilities and decision-making with the demand uncertainty of their products (Fisher (1997), Fisher et al. (1997)). To better match product supply to uncertain demand, firms develop forecasts (e.g., Fisher and Raman (1996)), keep safety stocks and inventories (e.g., Graves (1988), Graves and Willems (2000), Gaur et al. (2005)), use financial instruments (e.g., Gaur and Seshadri (2005)), and both produce (e.g., Graves (1988), Fine and Freund (1990), Jordan and Graves (1995), Goyal and Netessine (2011)) and price (e.g., Moreno and Terwiesch (2015)) flexibly. Yet inexorably, supply-side uncertainties and disruptions affect the manufacturer's ability and costs to produce (e.g., Tang (1990), Tomlin (2006)). Despite the centrality of uncertainty in production planning and performance, the question of how productive uncertainties originate from workforce dynamics remains under-explored. This paper addresses this gap for disruptive worker turnover.

As a threshold matter, whether turnover significantly disrupts manufacturing productivity depends on how process-relevant knowledge is created and held within the firm. A manufacturing firm's production routines and technologies can directly embed substantial production know-how (e.g., Epple et al. (1996)). Indeed, "[i]f knowledge is embedded completely in an organization's technology (tooling, programming, assembly line layout and balancing, and so on), then transfer of knowledge [between employees] ... should be complete [when using] the same technology" (Epple et al. (1996), p.77), implying that turnover engenders no ill effects. Therefore, a key part of our hypothesis development (Section 2.2) focuses on the extent to which *tacit knowledge*, defined as know-how that is difficult to codify and transfer, is held by employees and lost upon turnover.

The final part of our paper (Sections 5 and 6) closes the loop by characterizing turnover's disruptive effects on production. Organizational disruptions to productivity cut into manufacturing yields and cause assembly workrates to vary *unpredictably*, which is especially detrimental for production (see, e.g., Fisher and Ittner (1999)). Diminished yields further impair the responsive manufacturer's ability "to increase output quickly and to respond rapidly to customer and market demand changes" (Macher and Mowery (2009), p.S47).

A key focus of this analysis will be on how firms hedge against such disruptions, using inventories and workforce planning. A labor-based hedge entails maintaining a relatively larger workforce and utilizing overtime. Through simulated counterfactuals,[6] we investigate the operational value of a stable workforce: Could stabilizing the workforce, even at the cost of spending more on higher wages, prove cost-effective (by reducing the need to hedge in the first place and making production more reliable)?[7] We address this question using our structural model developed in Sections 5 and 6, rather than formulating it as a hypothesis.[8]

2.1. Workgroup Productivity

An extensive, cross-disciplinary literature examines productivity in manufacturing.[9] Our study primarily analyzes worker turnover's effects on workgroup productivity at the level of assembly lines. Workgroup departures can result directly from turnover, when workers leave the firm, or indirectly when managers transfer personnel to rebalance lines. We use "attrition" to refer to a workgroup's departures, which include both turnover and transfers leaving the line.

By assessing worker attrition's proximate effect on workgroup productivity, we follow a considerable line of research in operations and related fields (e.g., Glebbeek and Bax (2004), Kacmar et al. (2006)), which is premised on the notion that broader performance and profits are derivatively impacted when workgroup "efficiency is poor for a sustained period of time."[10] With the view

[6] Fisher and Ittner (1999) similarly employ simulation to extend an empirical analysis of assembly plant data.

[7] The value of workforce stability has been examined in settings outside manufacturing; see, e.g., Kacmar et al. (2006), p.139. See also Emadi and Staats (2018), Kesavan et al. (2018), and Musalem et al. (2018).

[8] A growing literature, including Bray and Mendelson (2012), Akşin et al. (2013), Li et al. (2014), Kim et al. (2014), Yu et al. (2016), and Moon et al. (2018), uses structural estimation to explore questions of operational interest.

[9] E.g., Arthur (1994), Ichniowski et al. (1997), Krueger and Rouse (1998), Lazear (2000), Hossain and List (2012), Bloom et al. (2013), Egelman et al. (2017). See, e.g., Huselid (1995) and Holtom et al. (2008) regarding turnover generally, and Shaw (2011) for a survey of productivity effects.

[10] Kacmar et al. (2006), pp.134-5, "argue that the immediate impact of turnover is felt through reduced efficiency . . . eventually evident in performance."

that workgroup productivity and efficiency mediate broader performance consequences, a primary contribution of our work is to quantify the size of turnover's effects on manufacturing workgroups.

Our workgroup productivity metrics are based on precedents, which fall under two general categories: (1) measures of waste, or equivalently measures of yield, acting on the inputs of the work process (sometimes referred to as *process quality*); and (2) measures of the rate of work completed, supplanted at times by the associated waiting times (*process speed*). Important instances of (1) include defect rework rates in automotive manufacturing (MacDuffie et al. (1996), Fisher and Ittner (1999)) and line and die yields in semiconductor manufacturing (Macher and Mowery (2009)).[11] Examples of (2) include labor hours per automobile (MacDuffie et al. (1996), Fisher and Ittner (1999)), semiconductor manufacturing cycle times (Macher and Mowery (2009)), and production volumes by week, month, or shift (Argote et al. (1990), Epple et al. (1996), Egelman et al. (2017)).[12] For process quality, we define in Section 3.1 the assembly line yield metric which aggregates workstations' yields. In contrast, process speed proves problematic as a workgroup performance metric. In responsive manufacturing, workgroup speed adjusts to factors outside the workgroup, including the firm's anticipation of *future* demand and its active allocation of workloads between lines (i.e., *other* workgroups' performance conditions become influential).[13] Instead, we investigate turnover's effects on work output aggregated to the level of buildings, which are less affected by the firm actively allocating work between lines. For workgroups, we define in Section 3.1 the metric of first-pass yield, which is a measure of operator mishandling causing some process delays.

[11] In service settings such as fast food, examples of (1) include the costs of resource inputs over the quantity served (Darr et al. (1995)) and food waste (Kacmar et al. (2006)).

[12] Kacmar et al. (2006) study wait times in fast food service and Narayanan et al. (2009) task resolution times in software maintenance. Note that in certain settings where service time is the predominant, variable process input, (1) and (2) can overlap.

[13] Our manufacturer sets assembly lines' target workrates in weekly planning meetings. In contrast, many manufacturers fix line speeds for months or longer. For example, Epple et al. (1996), p.79, study an automotive assembly plant that changed its uniform line speed a single time over three years. ("Assembly plants such as the one we are studying typically operate with a fixed line speed for an extended period of time – several months or longer." (p.80))

Based on observing the manufacturer's practices and on the findings of the existing literature (e.g., Kacmar et al. (2006), Ton and Huckman (2008)), we expect that worker attrition negatively affects the productivity of the assembly line as a workgroup. Thus, we test the following hypothesis.

HYPOTHESIS 1. *Worker attrition negatively affects assembly line productivity.*

2.2. Role of Tacit Knowledge

Knowledge about the production process can reside in an organization's employees, and whether that knowledge is readily transferred separates explicit from tacit knowledge (Grant (1996)). Whereas explicit knowledge is easily communicated, tacit knowledge is difficult to codify and articulate.[14] Because it "can only be observed through its application and acquired through practice, its transfer between people is slow, costly, and uncertain" (Grant (1996), p.111). For collaborative workgroups, tacit knowledge can importantly include knowledge about how to coordinate activities with a particular group of coworkers. Because attrition dissipates such know-how, the degree of tacit knowledge held by manufacturing employees determines the impact of their attrition.[15]

[14] Kacmar et al. (2006), p.135, describes that "[e]xplicit knowledge is common in the food service industry, in which employees follow detailed instructions. . . . An example of tacit knowledge gained in the fast-food environment is the ability to anticipate when another batch of fries will be needed on the basis of the length of the customer line. . . . As tacit knowledge increases, so does the speed component of efficiency." In software maintenance, tacit knowledge permits engineers to "learn to look for and quickly identify important cues that point out the nature and potential source of the [maintenance task]" (Narayanan et al. (2009), p.1863). In semiconductor manufacturing, Macher and Mowery (2009), p.S44, describes "the common problem [of manufacturing problems following] the transfer of new process technologies from developmental (pilot) facilities to the full-volume commercial production facilities . . . not only because of differences in equipment, production volumes and worker skills, but also because of tacit know-how. . . . Intel utilizes a 'copy exactly' approach, whereby everything (e.g. methodologies, process flows, equipment sets, suppliers, clean rooms) in the development facility is exactly duplicated in the commercial production facility."

[15] Tacit knowledge has been studied by the literature concerned with organizational learning and knowledge acquisition in production (e.g., Macher and Mowery (2009)). Like Ton and Huckman (2008), we focus on attrition's effect on how effectively a firm exploits existing knowledge (successfully executing known activities), instead of on how it explores and generates knowledge that is entirely novel (March (1991)).

We test multiple hypotheses to identify tacit knowledge in manufacturing. First, when tacit knowledge is vested in the organization's employees, worker attrition negatively affects assembly productivity, and attrition's direct, knowledge-based effect is seen when the firm (perhaps hypothetically) does not intervene to mitigate it. Thus, Hypothesis 1 is informative, provided that we control for the firm's interventions. Second, in the presence of tacit knowledge, investments in transferring knowledge between employees become important. We develop the following two hypotheses, reflecting the distinct implications for member entry and exit in affecting workgroup performance (e.g., Argote et al. (1990), Narayanan et al. (2009)).

The direction of new worker entry's effect provides an additional test for tacit knowledge. Upon new member entry, "in knowledge-intensive contexts, . . . short-term productivity may drop because [existing] members may need to expend time and effort to bring the new members onboard" (Narayanan et al. (2009), p.1873-4), whereas "in less knowledge-intensive contexts the entry of a new member could reduce the burden of the responsibilities experienced by other workgroup members, allowing them to be more productive on their regular tasks" (p.1872).

Hypothesis 2. *Novice operators new to the factory negatively impact workgroup productivity.*

A next hypothesis investigates whether staffing interventions counteract attrition as member exit. Namely, suppose we as managers "reverse" workgroups' attrition by sourcing similarly experienced peers to replace departures. Would attrition still negatively affect workgroup productivity? In the presence of tacit knowledge, the answer can be affirmative and economically significant. Tacit knowledge sources from experience and direct, interpersonal interactions, thus raising the value of inter-employee helping and knowledge sharing (Borgatti and Cross (2003), Siemsen et al. (2007), Argote (2013)). Working relationships facilitate new members' learning about their tasks and about workgroup communication patterns and practices (Mincer (1962), Narayanan et al. (2009)), but these worker-to-worker relationships are severed and lost upon attrition regardless of whether the exiting worker's role is filled by an experienced peer.[16]

[16] Hypothesis 3 is consistent with tacit knowledge without ruling out social motivations like the benefit of working with familiar coworkers.

Hypothesis 3. *When similarly experienced workers replace the workgroup's exiting workers, workgroup productivity is still negatively affected.*

2.3. Interventions

Through detailed conversations with the data sponsor's operations specialists, we identified how managers intervene to mitigate the consequences of workgroup attrition.[17] As Figure 11's right panel later depicts, managers route work away from poorly performing lines, including by using inventories to shift work between weeks (e.g., pre-building ahead of high turnover). While their actions minimize the impact of poor yields, managers further assert that slowing assembly lines helps productivity in the presence of novice (i.e., newly hired) operators and ensures that trained operators handle critical workstations. The final hypothesis tests these views:

Hypothesis 4. *Modifying a workgroup's work speed (process speed) mediates workgroup attrition's impact on productivity (process quality).*

Hypothesis 4 is relevant in two ways. As Section 2.2 discusses, attrition's productivity impact without interventions reflects the tacit knowledge lost when employees leave. Second, Section 6 simulates production planning that incorporates effective managerial interventions, including the interventions tested in Hypotheses 3 and 4 if they are found effective.

3. Research Site and Data

We investigate the episodes of worker turnover and productivity taking place in the final assembly, testing, and packaging (FATP) of a smartphone model (i.e., one generation of a device series) that sold over 200 million units over its production life cycle. We study the product's forty-four assembly lines located in three buildings within a cluster of production facilities, which together exceed ninety football fields in production-floor area and can assemble millions of consumer electronic devices per week. Our data sponsor (the "firm") and one of its China-based contract manufacturers provided comprehensive production, planning, and human resources records spanning from September 2014 to June 2015.[18]

[17] Conversations and electronic correspondence (July and August 2018).

[18] We omit identifying details for confidentiality.

Workforce. Over the time of our study, 52,214 workers held active positions on the assembly lines,[19] after excluding 4,069 who left before completing their first weeks on the job (i.e., first 7 days including 2-3 of training). Each worker's staffing history is combined with her monthly compensation records. The data also provide her basic demographics, such as gender and age, and her starting and ending dates (hence overall tenure of employment) at the manufacturer.

An employee's compensation, disbursed monthly, is primarily determined by her hours worked. A base monthly salary of 1820 RMB attaches to the forty-hour workweek. In compliance with Chinese labor laws, multiples of her base hourly rate apply to overtime pay: Specifically, (A) 1.5 times for additional hours worked on business days; (B) 2 times for weekend hours; and (C) 3 times for hours worked on Chinese public holidays (Plevan et al. (2011)). Illustrating how significantly overtime shapes compensation, employees average 1080 RMB compensation *per week* in our sample. Because of overtime, compensation varies considerably by assembly-line workload in hours.

Employees collect retention bonuses along with their monthly salaries. At the time of our study, employees received retention bonuses in the amounts of: 100 RMB upon completing her first complete pay period (i.e., employment spanning from one calendar month's 26th to the following 25th); 200 RMB for two or three pay periods completed; and 300 RMB thereafter. Compensation details on shifts, skill bonuses, and meal stipends can be found in Appendix A.1.

Table 1 shows summary statistics for employee turnover. Overall, 6.4% of employee-week observations involve turnover, implying annual separation rates above 300%. Appendix A.2's descriptive evidence supports that turnover responds to compensation. Approximately a quarter of workers are female and they tend to stay longer, being represented more heavily among long-tenured employees than new recruits. The manufacturer hires some workers through government-licensed labor brokers called "dispatch agencies." These workers, commonly called "dispatch" employees, tend

[19] Overall, approximately 38% of the firm's overall supply-chain workforce supports the final assembly, testing, and packaging (FATP) of its products.

Table 1 Workforce Experience and Turnover

	All Person-weeks	Cross-section in Week Starting Oct. 6, 2014	Jan. 5, 2015	Apr. 6, 2015
Unique employees	52,214	12,630	15,655	11,597
Female	27%	27%	27%	26%
Dispatch hire	52%	58%	53%	49%
Experience (in days)				
25th quantile	33	31	78	25
Median	74	49	114	32
75th quantile	160	123	175	142
Mean	124	101	156	103
Female-only	147	125	186	120
New hire or transferred from non-FATP work	8.1%	5.5%	5.0%	3.2%
New and female	1.8%	1.1%	0.7%	0.8%
Weekly turnover	6.4%	10.1%	10.5%	9.4%
Turnover and female	1.4%	1.6%	3.3%	1.6%
Turnover and dispatch (at hire)	4.3%	8.8%	5.8%	6.5%

509,498 worker-weeks observed for 52,214 employees assigned to production-active line-weeks in Sept. 2014 – Jun. 2015. Summary statistics shown for worker cross-sections in selected weeks, plus for all person-weeks in data. Experienced employees are more heavily represented in all person-weeks. Experience is days from hire, which may pre-date our study period. Turnover is higher in the weekly cross-sections than for all person-weeks, because the selected weeks are all post-payweek so that they are comparable.

to originate farther away from the factory and turnover more rapidly on average.[20] Involuntary turnover (i.e., termination) is rare. Besides facing already rapid voluntary turnover, the contract manufacturer is subject to stringent Chinese labor regulations, which restrict employee termination (absent severance) to narrow reasons specified by the government (e.g., Plevan et al. (2011)). The overall stock of experience peaks in the middle of the product's life cycle, due in part to when employees are moved to the next generation product.

Production. The production process begins with the product's parts and components, including the device housing and coverglass. Parts are added to the device body as it passes through an assembly line containing over sixty types of workstations. The assembly line is divided into

[20] Owing mainly to regulatory developments, dispatch employees can apply to become regular employees at the end of their six-month dispatch labor contracts. While standard dispatch labor contracts are reported to last the legally permitted six months (Yu (2014)), some contracts may last four months with an option to convert to regular employee at three months. See Appendix A.1 for further background.

five stages: subassembly, pre-burn assembly, burn-in, post-burn assembly, and packout. At each workstation, an operator carries out her designated task within a cycle time of approximately 20-30 seconds. Unlike some other manufacturing settings, station-specific test fixtures assess the completion of each assembly task. Excepting a limited number of diagnostic test stations (most notably for packaging defects) that sample units selectively, no unit ships to a retailer or customer without having passed all test stations.[21] Over 6.8 million station-level data observations compile the numbers of unique device units (i.e., no double counting for multiple tests on a unit) that were input, passed, retested, and failed at two-hourly intervals at each station. We aggregate these into weekly workgroup productivity measures for the individual assembly lines.

Master Production Schedules. To coordinate responsive production, the firm's MPS (Master Production Schedule) is updated and shared twice-weekly with the contract manufacturer and the firm's procurement teams.[22] Each MPS contains a rolling horizon of anticipated order quantities based on the firm's evolving demand forecasts. Importantly for their role as instrumental variables in our empirical analysis, the MPS quantities are idealized, demand-based order quantities that ignore supply constraints, which are instead complied with by the accompanying CTB (clear-to-build) order quantities. See Appendix B for details.

3.1. Dependent Variables

At the workgroup level, our dependent variables are the assembly lines' yields (process quality) and first-pass yields (process speed). Both metrics are used in practice. Firm specialists track yields continuously in real-time, and yields bear directly (1) on the manufacturer's material costs of production and (2) on its ability to satisfy ever-shifting production plans on-schedule with quality-approved units. Because they fall when inexperienced operators mishandle the test fixtures, first-pass yields signal delays that prolong assembly flowtimes. Expanding and uncertain flowtimes

[21] Due to their discretionary sampling, we drop diagnostic test stations in constructing our productivity measures.

[22] These teams coordinate a supply chain involving over 700 suppliers. See Graves (1981) for an overview of the role and practice of master scheduling in final assembly production. Illustrating how responsively the firm produces, it is public knowledge that its distribution centers turn inventories in a matter of days.

strain the prodigious production spaces required under Little's Law and affect line balancing. We next define the yield and first-pass yield metrics.

At each workstation, we calculate the weekly station yield as the number of assembled units passing the station's quality check over the station's number of unique device units tested. For assembly line l in week t, yield is defined as the product of its station yields, reflecting that the devices being assembled pass sequentially through the line's workstations:

$$\text{Yield}_{lt} := \prod_{s\in S} \frac{\#\ \text{Pass}_{lts}}{\#\ \text{Input}_{lts}}, \tag{1}$$

where S is the set of stations s.[23] By nature, assembly line yields sensitively and noisily measure process quality: for example, halving the station yield of a single workstation (out of over sixty) halves yield. Nonetheless, even noisy and stochastic fluctuations in station yields carry operational consequences (Fisher and Ittner (1999)).

By measuring the share of an assembly line's units passing all test stations without incurring a single re-test,[24] its first-pass yield reflects the incidence of impediments delaying production flow. For assembly line l in week t, first-pass yield is defined as the product of its station first-pass yields, which are the workstations' respective proportions of units passed without re-testing:

$$\text{First-pass Yield}_{lt} := \prod_{s\in S} \frac{\#\ \text{Pass}_{lts} - \#\ \text{Retested}_{lts}}{\#\ \text{Input}_{lts}}. \tag{2}$$

As shown in Appendix D, our subsequent estimation results are virtually identical when using a version of first-pass yields that excludes failed units (i.e., $\#\ \text{Pass}_{lts}$ is the denominator).

Table 2 summarizes assembly lines' staffing, activity, and yields. Yields vary despite being closely monitored, exceeding 93% in over 70% of observations and 91% in over 95%. In the average line-week, 31.7% of units pass all station quality checks at first try, and first-pass yields dip below

[23] After consulting the firm's operations specialists, we exclude the burn-in stage from our yield calculations. These stations' tests hinge on the firmware installed on the device or on related component failure, rather than on correct device assembly. Procured parts and components are quality tested before being transported to the FATP lines, insulating our yield metrics from variation caused by component defect rates.

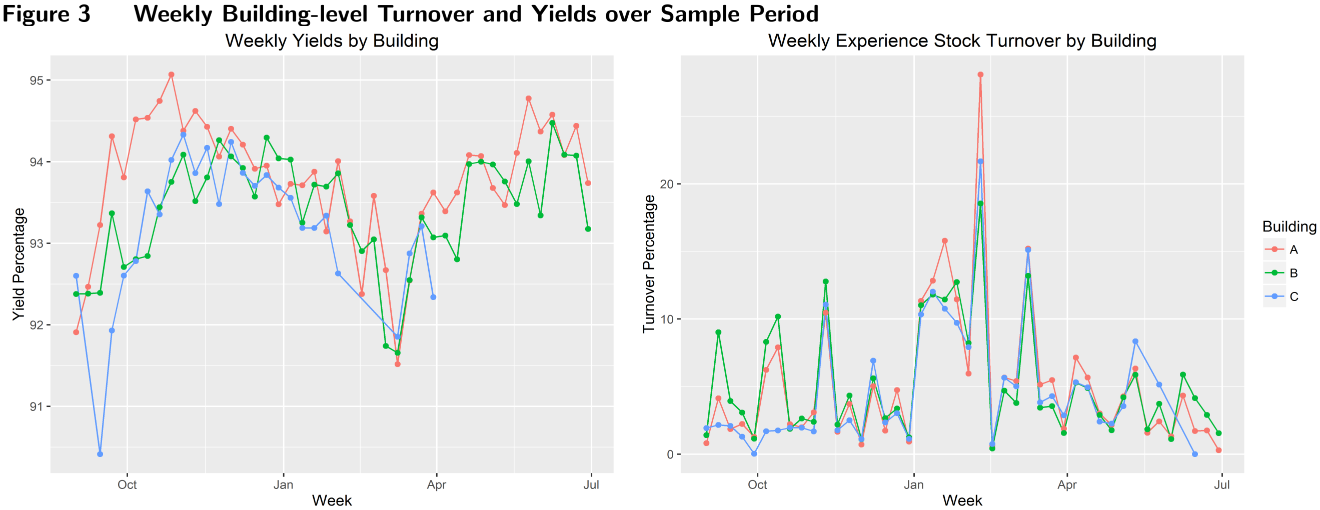
[24] A unit may be re-tested twice at a given test station before being deemed failed and removed from workflow.

Table 2 Weekly Assembly Line Activity

		Mean	Median	Std. Dev.
Staffing	Headcount	301	259	
	Female	30%	29%	
Turnover & Attrition	Turnover in headcount	5.1%	3.2%	
	Attrition in headcount	9.1%	5.1%	
	Turnover in experience stock	4.3%	2.4%	
	Attrition in experience stock	7.7%	4.0%	
Production	Line hours (all shifts)	134	141	22
	Units per hour	224	231	37
	Weekend utilization (% hours)	52%	54%	28%
	Yield	93.4%	93.7%	1.7%
	First-pass yield	31.7%	32.1%	5.5%

Unbalanced panel of 44 assembly lines observed in 1,516 production-active line-weeks in Sept. 2014 – Jun. 2015. An assembly line's attrition rate captures all workgroup exit, including from (A) turnover when workers exit the firm and (B) staffing adjustments transferring workers to other assembly lines.

Figure 3 Weekly Building-level Turnover and Yields over Sample Period



For clarity, low-yield outliers for Building C in Feb. and Apr. 2015 onward are omitted. By allocating work to productive assembly lines, building yields dampen the underlying variation of workgroup yields.

a quarter in fewer than 11% of line-week observations. Figure 3's left panel shows week-to-week variation in building-level yields.

In contrast to less responsive manufacturing facilities (see footnote [13]), the assembly lines' workrates vary considerably in units per hour. Table 2 shows that weekend utilization rates run high during our study but vary substantially, indicating that weekend work is far from certain.

3.2. Independent Variable

A workgroup's attrition is measured to reflect that its stock of tacit knowledge, gleaned from workers' assembly experience, diminishes when members leave the assembly line. Absent attrition, assembly line l would be staffed in week t by the set of workers, $J(l,t)$, that staffed the line in the prior week. Let $D(l,t)$ denote the subset that departs line l in week t. Each worker j has an experience level, Experience_{jt}, measured by her number of weeks spent in FATP assembly before week t. For week t, we define workgroup l's experience stock and experience loss as follows:

$$\text{ExperienceStock}_{lt} := \sum_{j\in J(l,t)} \text{Experience}_{jt};\ \textit{and}\ \text{ExperienceLoss}_{lt} := \sum_{j\in D(l,t)} \text{Experience}_{jt}. \qquad (3)$$

Finally, the workgroup attrition rate is the fraction of week t's experience stock lost:[25]

$$\text{WorkgroupAttrition}_{lt} := \frac{\text{ExperienceLoss}_{lt}}{\text{ExperienceStock}_{lt}}. \qquad (4)$$

4. Empirical Analysis

We introduce our strategy for investigating the empirical hypotheses. At a high level, because we examine how tacit knowledge underlies manufacturing productivity, we are interested in measuring attrition's knowledge-based effects on productivity without interventions that could mitigate them. The general empirical strategy is to control for the interventions, using the control variables defined in this section. However, as we additionally explain, we also require instrumental variables.

Effect on Productive Output. We first examine the magnitude of turnover's overall effect on the manufacturer's production. We measure weekly production output as the facility's number of units successfully assembled and passing all quality controls.

The analysis is carried out at the building level. This choice strikes a balance in aggregating production over manufacturing workgroups, i.e., assembly lines. Aggregation helps avoid measuring that high-performing lines are given busier workloads, which inflates turnover's effect on output. Building and time fixed effects focus the analysis on effects on output stemming from buildings'

[25] For robustness, we also ran regressions substituting the workgroup's attrition rate by headcount, $\frac{|D(l,t)|}{|J(l,t)|}$.

Table 4 Turnover's Effects on Building-level (Log) Final Outputs

Regressors	
Building turnover rate	-3.864^{**}
	(1.258)
Log pre-turnover staffing level	0.364^{***}
	(0.052)
R^2	0.80
Significance levels → *** - 0.001 ** - 0.01 * - 0.05	

Unbalanced panel of 111 building-weeks (31 for building C) in Sept. 2014 – Jun. 2015. Building-level log final outputs are regressed on building-level turnover rates and log pre-turnover staffing headcount, with two-way fixed effects for 3 buildings and 40 weeks. Results are similar when turnover rates are in headcount instead of experience stock.

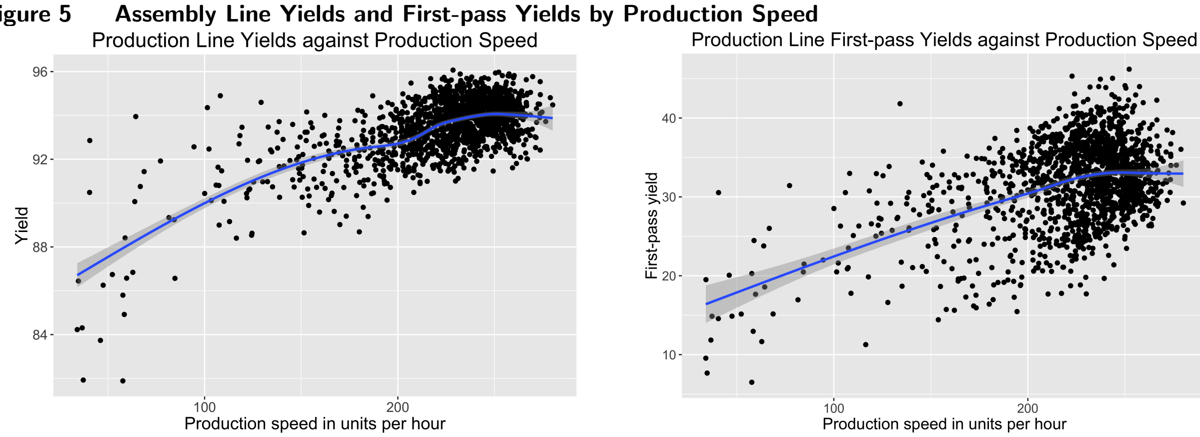
variation in turnover experienced in the same week. We thus control for time fixed effects including for weekly updates in demand forecasts and seasonal effects from the pay cycle.

We regress log building output on building-level turnover, with building and time fixed effects and controlling for the building's (log) pre-turnover staffing level which scales with its amount of production activity. The results are shown in Table 4. We find that each additional percentage point in building turnover translates into a 3.8% drop in final output. For comparison, each percentage point increase in final output associates with a 0.36% increase in initial staffing.

Workgroup Attrition Effects. We next examine how workgroup attrition affects productivity after controlling for mitigating interventions. In analyzing yields, we control primarily for managers intervening to maintain productivity by slowing affected lines. Specifically, we control for assembly lines' weekly production speeds in units per hour, UPH_{lt}. We source assembly lines' production speeds and hours from benchmark stations that the firm's operations specialists use to track assembly line activity.

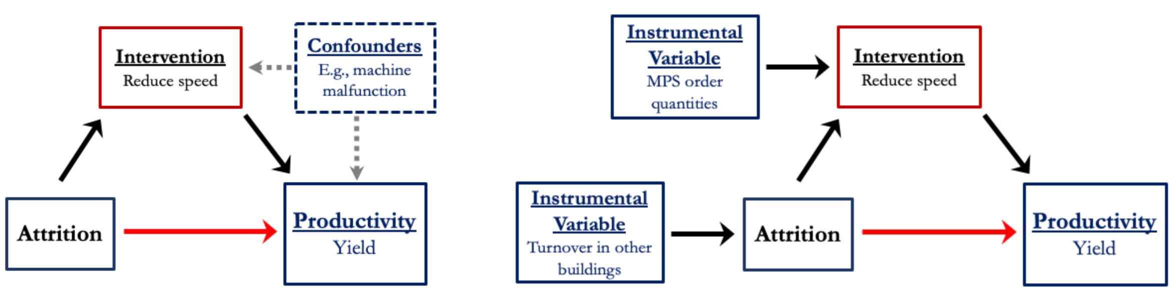
Figure 5 plots, without any controls, the assembly lines' weekly yields and first-pass yields against their production speeds. Surprisingly, although managers claim that they slow assembly lines in order to *raise or maintain* productivity, slower speeds are empirically associated with *falling* yields and first-pass yields. This can be explained by the fact that assembly lines are slowed in response to myriad productivity issues unrelated to turnover. For example, when an assembly line's workflow is impaired by a machine malfunction or a logistical snafu (a negative productivity shock), managers slow its pace to salvage its productivity. By inducing interventions, such unobserved productivity shocks cause the controls to endogenously associate with reduced yields. As shown in Figure 6, the productivity shocks are unobserved confounders, i.e., omitted variables. Regressing the productivity variables on workgroup attrition and controls would misattribute the unobserved productivity shock to the speed accommodation made to counter it.

Figure 5 Assembly Line Yields and First-pass Yields by Production Speed



Slower production speeds empirically associate with lower yields (left panel) and first-pass yields (right panel), but these associations are potentially endogenous as problematic lines are often slowed. 1,516 production-active line-weeks are plotted with fitted conditional expectations.

Figure 6 Identifying the Effect of Workgroup Attrition on Yields



(a) As observed (b) With instruments

The workgroup-level analysis controls for the manufacturer's mitigating intervention(s), but the effect of intervening is confounded by unobserved productivity shocks. Attrition can likewise respond to productivity shocks, e.g., if managers are reluctant to transfer workers from poorly performing lines. We employ the instrumental variables of MPS order quantities, which influence production speeds, and turnover in other buildings, which affects whether the firm transfers workers. Neither variable should be related to an assembly line's idiosyncratic productivity shock.

Workgroups' attrition rates potentially suffer a similar sort of endogeneity. In transferring employees between lines, reasonable managers would source personnel from poorly performing lines only reluctantly, to avoid worsening their performance. Then low attrition rates would attach to negative productivity shocks, causing the data to mask low worker attrition's productivity benefit.

We introduce two instrumental variables that induce variation in the control or attrition variable (*relevance*) but which are uncorrelated with the assembly lines' unobserved productivity shocks (*exclusion restriction*). Responsive manufacturers tailor their production plans to fit the freshest information about demand, and our manufacturer's weekly MPS reflects updates from its most recent demand forecasts. (Recall that CTB separately tracks supply constraints.) The demand-driven MPS clearly influences production (relevance), but is independent of lines' idiosyncratic productivity shocks (exclusion restriction).[26] As Figure 7 shows, when the MPS quantities forecast that production ramps up, assembly lines slow while hours worked rise. Appendix B defines MPS-based variables capturing the ramping of quantities in each of the next three weeks' orders.

The second instrumental variable exploits that managers' incentives for transferring personnel away from a line depend on the urgency of the manufacturer's staffing needs elsewhere. Consequently, as Figure 8 shows, the turnover rates in the two buildings not housing the assembly line predict its attrition rate (relevance). However, other buildings' turnover rates are independent of an assembly line's own idiosyncratic productivity shocks, e.g., machine malfunction (exclusion restriction).

Regression Model. We estimate the following model of yield productivity:

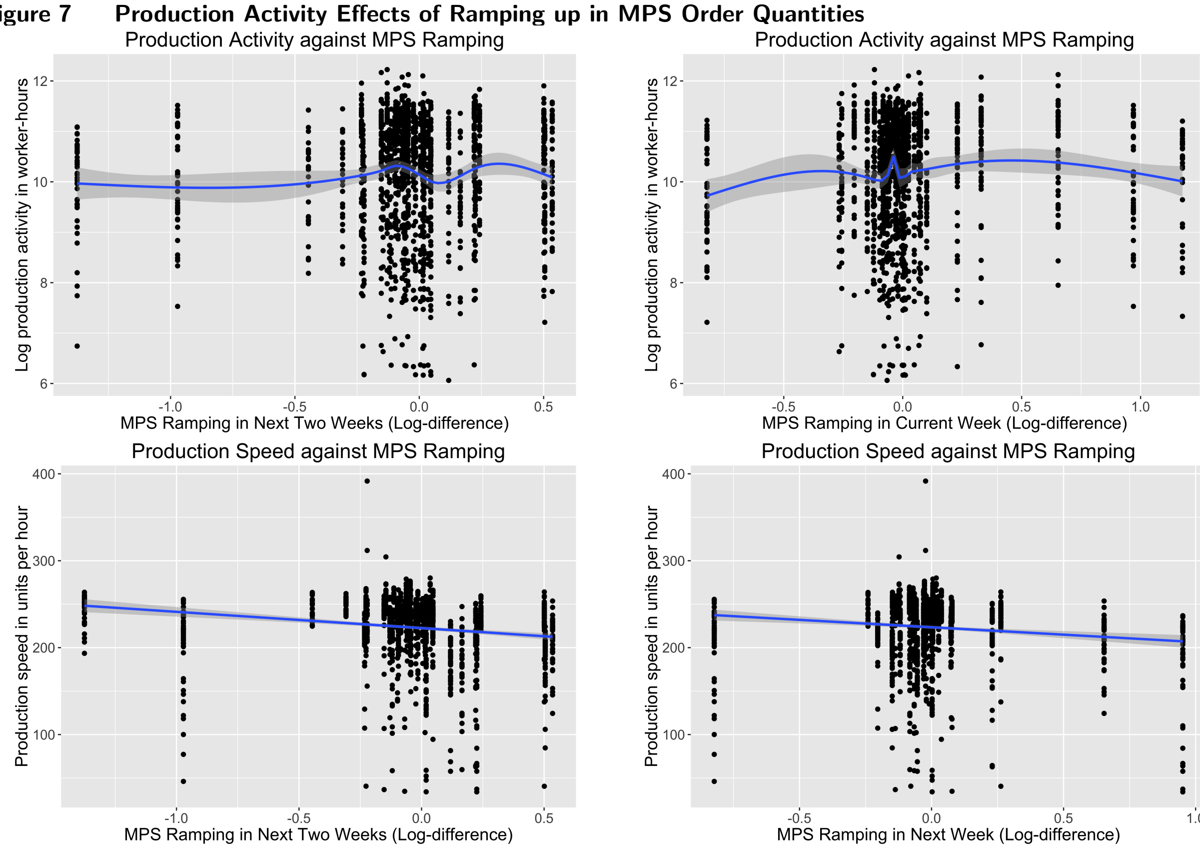
$$\log(\text{Yield}_{lt}) = k_l + \beta \cdot \text{WorkgroupAttrition}_{lt} + \gamma \cdot X_{lt} + \epsilon_{lt}, \tag{5}$$

where X_{lt} includes the control variable(s) for assembly line l's production in week t, and k_l is a line-level fixed effect capturing supervisor effects[27] and installed fixed capital. The line-week's

[26] The MPS can be viewed as a "demand shifter" tracing out the supply-side response to exogenous shifts in demand.

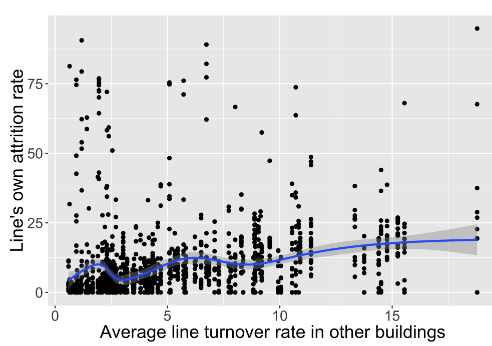
[27] We believe there may be some performance-based movement of supervisors. If such movements allay productive issues stemming from turnover, our findings may be conservative.

Figure 7 Production Activity Effects of Ramping up in MPS Order Quantities



Ramping in MPS order quantities positively predicts assembly line workloads in (log-scaled) worker-hours and slower workrates to protect yields. 1,516 production-active line-weeks are plotted with fitted conditional expectations.

Figure 8 Attrition Effect of Worker Turnover in Other Buildings



Assembly line attrition rates positively associate with average turnover (i.e., firm exit) rates in the buildings not housing the line. 1,516 production-active line-weeks are plotted with the fitted conditional expectation.

idiosyncratic productivity shock, ϵ_{lt}, endogenously correlates with the firm's potentially mitigating adjustments to the assembly line's speed.

Table 9 Workgroup Yield Analysis

(a) Log Yield Regression Results with IV

Workgroup Regressors	(1)	(OLS)	(2)	(OLS)
Attrition rate	-0.067^{***}	-0.001	-0.117^{*}	0.000
	(0.019)	(0.003)	(0.045)	(0.003)
$\log(\text{UPH}_{lt})$	0.739^{*}	0.245^{***}	0.541	0.252^{***}
	(0.296)	(0.023)	(0.358)	(0.023)
$\log(\text{UPH}_{lt})^2$	-0.065^{*}	-0.020^{***}	-0.041	-0.020^{***}
	(0.029)	(0.002)	(0.034)	(0.002)
Dummies for weeks to payweek			Y	Y
R^2	0.41	0.45	0.38	0.46
Wu-Hausman p-value	<0.001		<0.001	
Sargan p-value	0.995		0.484	
Significance levels → *** - 0.001	** - 0.01	* - 0.05		

Unbalanced panel of 1,516 production-active line-weeks in Sept. 2014 – Jun. 2015. Fixed effects for 44 assembly lines. IV regressions (1) and (2) use instrumental variables: average line turnover in other two buildings and MPS-based variables. Weak-instrument F-test p-values all fall below 0.001 to reject null.

(b) Projected Impacts on Workgroup Yields

Attrition rate	*(Quantile)*	Yield
0% (Set median yield)	*(0-17%)*	93.7%
4.0%	*(50%)*	93.2%
9.5%	*(75%)*	92.6%
18.7%	*(90%)*	91.6%
45.3%	*(97.5%)*	88.8%

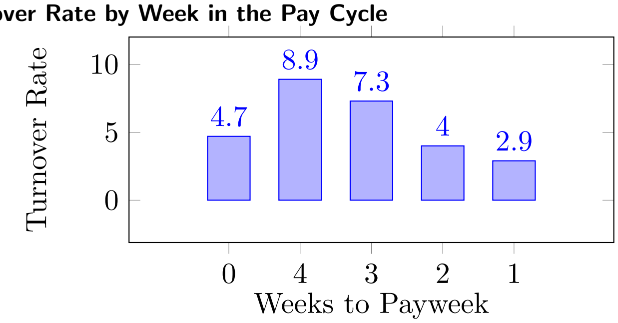
Line-week median yield of 93.7% used as normalizing benchmark. Projections apply IV estimates (2) to sample quantiles. Interpret, e.g., 75% sample quantile in attrition rate (9.5%) projects to drop yield by 1.1% to 92.6%.

The instrumental variables are further detailed in Appendix C, which provides all first-stage regressions. Figure 6's panel (b) provides an illustration of how instrumental variables are used to identify the causal effects of attrition and line workrate on productivity.

4.1. Yield Regression Results

Table 9a reports our results from regressing workgroups' log yields on their attrition rates, controlling for production speeds. We report results with and without the instrumental variables and also report results that remove the seasonal variation of the pay cycle, in which turnover rates decline before payweek and rise after workers are paid. (As Figure 10 shows, assembly line turnover averages 2.9% weekly immediately before payweek and 8.9% immediately following.) All analyses include assembly line fixed effects. We test for evidence of endogeneity (Wu-Hausman) and for violation of the exclusion restriction using over-identified moments from multiple MPS variables (Sargan).

Figure 10 **Average Turnover Rate by Week in the Pay Cycle**



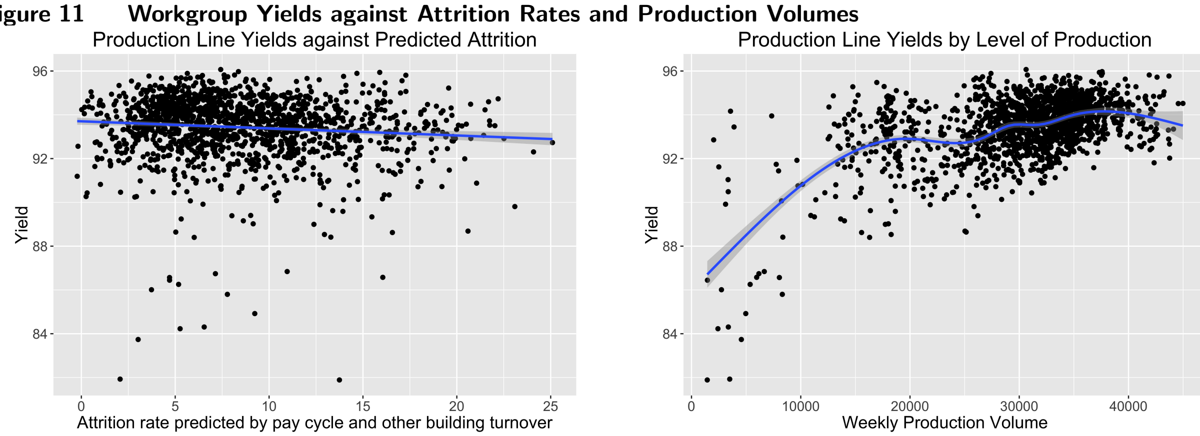
Columns (1) and (2) display the primary regression results, and we focus on regression (2) which controls for the pay cycle. Supporting Hypothesis 1, we find that assembly line worker attrition negatively impacts yields. Table 9b projects attrition's yield impact. Workgroup attrition causes yields to decline in excess of 1.1 percentage points (0.6 standard deviation) in a quarter of our sample's line-weeks and over 2.1 percentage points (1.2 s.d.) in one-tenth. By conservative calculation, these declines amount to $178 million in material waste accumulated over the life cycle of the studied device.[28] Yet, as column (OLS) shows, these costs are typically masked because attrition occurs when managers endogenously source transfers from relatively well-performing lines.

Regarding Hypothesis 4, regression (2) finds no evidence that reducing production speeds protects product yields. Instead, descriptive evidence shows that the manufacturer routes heavier workloads to the best-performing lines. See Figure 11's right panel.

Finally, in Table 12, we re-estimate the yield regressions after adding an additional independent variable for testing Hypothesis 3. Recall that Hypothesis 3 hinges on whether the negative effect on workgroup productivity persists when similarly experienced workers replace those exiting. We define the workgroup's *net* attrition rate as its fraction of experience lost (in weeks of experience) *after* allowing the experience of newly staffed members to compensate for that of exiting members. The estimated coefficients of attrition rate and net attrition rate – both added to the same regression specification – can help us determine whether tacit knowledge or worker experience best explain yield loss (columns 1 and 2 of Table 12). Should net attrition rates explain yields, then this

[28] Using public estimates of the manufacturer margin and device cost to build, the conservative lower bound attributes to each line-week the highest productivity drop in Table 9b that is *less than* its own projected drop.

Figure 11 Workgroup Yields against Attrition Rates and Production Volumes



The left panel plots product yields against assembly line attrition rates predicted by other buildings' turnover (the instrumental variable). The right panel shows production activity routing to high-yield line-weeks. 1,516 production-active line-weeks are plotted along with fitted conditional expectations.

Table 12 Log Yield Regressions Testing Hypothesis 3

Workgroup Regressors	(1)	(OLS)	(2)	(OLS)
Attrition rate	-0.069^{***}	-0.001	-0.127^{**}	0.000
	(0.020)	(0.003)	(0.045)	(0.003)
Net attrition rate	-0.003	0.000	0.004	0.000
	(0.005)	(0.000)	(0.006)	(0.000)
$\log(\text{UPH}_{lt})$	0.775^{*}	0.245^{***}	0.474	0.252^{***}
	(0.373)	(0.023)	(0.377)	(0.023)
$\log(\text{UPH}_{lt})^2$	-0.068	-0.020^{***}	-0.035	-0.020^{***}
	(0.020)	(0.002)	(0.036)	(0.002)
Dummies for weeks to payweek			Y	Y
R^2	0.40	0.45	0.35	0.46
Wu-Hausman p-value	<0.001		<0.001	
Sargan p-value	0.561		0.784	
Significance levels → *** - 0.001	** - 0.01	* - 0.05		

Unbalanced panel of 1,516 production-active line-weeks in Sept. 2014 – Jun. 2015. Fixed effects for 44 assembly lines. IV regressions (1) and (2) use instrumental variables: lagged own-line turnover, average line turnover in other two buildings, and MPS-based variables.

would support that attrition's productivity impact can be avoided by replacing departing workers with similarly seasoned peers. Should attrition rates explain yields, the empirical evidence supports Hypothesis 3 and the presence of tacit knowledge.

However, to carry out this analysis, we require an instrumental variable to separately identify net attrition from attrition. The chosen instrumental variable should predict whether managers replace departures with peers, yet remain exogenous to the assembly lines' unobserved productivity shocks in the current week. An assembly line's lagged turnover is such an instrument. When the line weathers two consecutive weeks of turnover, continuing to assimilate incoming personnel becomes challenging, making net attrition more likely in the second week. The first-stage regressions in Appendix C show that net attrition rates are significantly predicted by lagged assembly line turnover and not by whether turnover occurs in the other buildings. By using this instrument, intuitively, we compare an assembly line's productivity in a second week of attrition to its response in the first, when attrition was countered by adding experienced personnel. We would only find empirical support for Hypothesis 3 if productivity were just as negatively impacted in the initial week of attrition as it is in the second.

As Table 12 shows, we find support for Hypothesis 3, since only attrition rates significantly explain yields. The regression results are economically similar to Table 9a, and the Sargan tests support the joint validity of the instruments.

4.2. First-pass Yield Regression Results

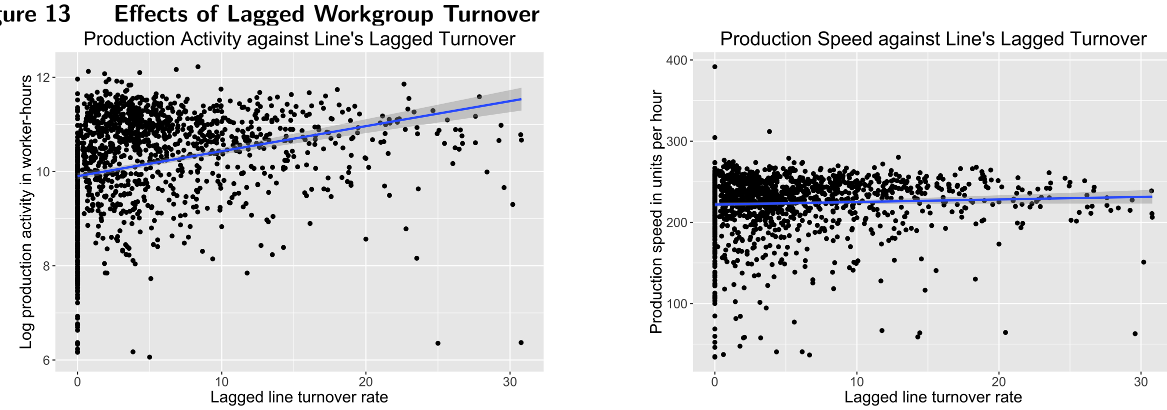
The study of manufacturing first-pass yields is empirically more complex. First-pass yields are more than three times as variable as yields and influenced by a greater number of workgroup factors. In the ensuing analysis, we additionally control for assembly lines' staffing of new hires and for their assigned workloads in worker-hours. For week t, we control for assembly line l's staffing of new hires, as a percentage, and its workload in worker-hours, L_{lt}.[29]

The analysis requires additional instrumental variables and assumptions. Just as managers avoid sourcing transfers from assembly lines with productivity issues, they may avoid staffing new hires onto lines suffering operational delays. If we do not account for the biased placement of new hires onto well-performing lines, our findings would falsely mute their impact on production, and estimating this impact is critical for testing Hypothesis 2. Managers clearly route work to highly performing lines (Figure 11, right panel), biasing the use of workloads as a productivity control.

The first instrumental variable we add, MPSInfo_t, measures the proportional "surprise" (i.e., revision from the last report) baked into the MPS order quantity reported in week t. See Appendix B

[29] With these controls, the regression model derives from re-arranging a Cobb-Douglas production function, where total productivity can depend on a workgroup's attrition rate and control variables.

Figure 13 Effects of Lagged Workgroup Turnover



Lagged assembly line turnover positively predicts the line's workload in worker-hours. However, no significant change in production speed is observed, which helps us separate the effect of speed from worker-hours. Fitted conditional expectations (vertical on horizontal) are plotted for 1,516 production-active line-weeks.

for a definition. As the first-stage regressions show, managers avoid staffing new hires as intensively when MPS ramps up by surprise (relevance). Surprising changes to the firm's demand forecasts are exogenous to assembly lines' idiosyncratic productivity shocks (exclusion restriction). Intuitively, applying the instrument strengthens the estimated impact from new hire staffing only if productivity is lower when MPS ramps *without* a surprise than with one. Thus, such a finding is conservative if one imagines that surprises disrupt productivity in ways we may not capture.

The second additional instrumental variable is lagged assembly line turnover. Workers are clearly incentivized by their hours-based compensation, and indeed their exit rates depend strongly on how much they are owed for the upcoming payday (see Section 6 and Appendix A.2). In turn, when a workgroup's hours fall behind, managers are incentivized to help it "catch up" for purposes of protecting compensation and retaining workers. Figure 13 illustrates this effect. Workers on last week's high-turnover lines have accrued fewer hours, because managers route workloads to highly performing lines instead (Figure 11, right panel). The situation shifts towards re-balancing workers' accumulated hours in the current week. As shown by Figure 13, managers intensify the workloads of these assembly lines without affecting their production speeds (relevance). Figure 13's flat production speed curve (right panel) corroborates the Sargan tests in suggesting that no residual productivity effect carries over from the prior week of high turnover (exclusion restriction).

Table 14 Workgroup First-pass Yield Analysis

(a) Log Workgroup First-pass Yields Regressions with IV

Workgroup Regressors	(1)	(2)	(OLS)	(3)	(4)	(OLS)
New hire staffing %	-3.065^{***}	-2.672^{***}	-0.930^{***}	-3.016^{***}	-3.026^{***}	-0.921^{***}
	(0.312)	(0.452)	(0.055)	(0.309)	(0.445)	(0.055)
Attrition rate		-0.347	-0.024		-0.080	-0.022
		(0.309)	(0.033)		(0.305)	(0.033)
$\log(\mathrm{UPH}_{lt})$	8.537^{*}	6.348	1.141^{***}	8.408^{*}	9.466^{*}	1.153^{***}
	(3.765)	(3.969)	(0.240)	(3.638)	(3.853)	(0.238)
$\log(\mathrm{UPH}_{lt})^2$	-0.857^{*}	-0.637	-0.080^{**}	-0.843^{*}	-0.936^{*}	-0.081^{***}
	(0.307)	(0.388)	(0.024)	(0.354)	(0.380)	(0.024)
$\log(L_{lt})$	-0.107^{*}	-0.101^{*}	0.011^{*}	-0.099^{*}	-0.110^{*}	0.012^{*}
	(0.045)	(0.041)	(0.005)	(0.044)	(0.045)	(0.005)
Dummies for weeks to payweek				Y	Y	Y
R^2	0.19	0.18	0.42	0.21	0.21	0.42
Wu-Hausman p-value	<0.001	<0.001		<0.001	<0.001	
Sargan p-value	0.578	1		0.461	0.212	

Significance levels → *** - 0.001 ** - 0.01 * - 0.05

Unbalanced panel of 1,516 production-active line-weeks in Sept. 2014 – Jun. 2015. Fixed effects for 44 assembly lines. IV regressions use instrumental variables: lagged own-line turnover, average line turnover in other two buildings, and MPS-based variables. Weak-instrument F-test p-values all below 0.001 to reject null.

(b) Projected Impacts on Workgroup First-pass Yields

New hire staffing	*(Quantile)*	First-pass Yield	Speed in UPH	*(Quantile)*	First-pass Yield
0% (Set median yield)	*(0-37%)*	32.1%	182	*(10%)*	36.6%
0.5%	*(50%)*	31.7%	212	*(25%)*	34.1%
2.6%	*(75%)*	29.6%	231 (Set to median yield)	*(50%)*	32.1%
9.0%	*(90%)*	24.3%	246	*(75%)*	30.5%
25.1%	*(97.5%)*	14.7%	256	*(90%)*	29.4%

Line-week median first-pass yield of 32.1% used as normalizing benchmark. Projections apply IV estimates (3) to sample quantiles. Interpret, e.g., 75% sample quantile in new hire staffing (2.6%) projects to drop first-pass yield by 2.5% to 29.6%.

Table 14a reports findings for first-pass yields. Regression (3) shows our main estimates regarding the effects of assembly lines' new hire staffing, production speeds, and workloads. Unlike for yields, we do not find an effect from workgroup attrition (Hypothesis 1) in regressions (2) and (4). We do find significant support for Hypothesis 4, whereby reducing speed by 21% from 231 UPH (sample median) to 182 UPH (10% quantile) raises first-pass yields by 4.5 percentage points.

Addressing Hypothesis 2, we find an economically and statistically significant negative impact on first-pass yields from staffing new hires. As Table 14b shows, in one-tenth of line-weeks, new hires reduce first-pass yields by at least 7.8 percentage points. In a quarter of these, first-pass yields

Table 15 Productivity Effects of Pay Cycle Turnover

(a) Fixed-effect IV Regression Results: Log Workgroup Yields Regressed on Turnover

Dependent variable	Regressor: (Pay cycle turnover)
log(Yield)	-0.168^{**}
	(0.051)
R^2	0.002

Significance levels → *** - 0.001 ** - 0.01 * - 0.05

Unbalanced panel of 1,516 production-active line-weeks in Sept. 2014 – Jun. 2015. Fixed effects for 44 assembly lines.

(b) Projected Impacts on Workgroup Yields

Workgroup turnover rate	*(Turnover Quantile)*	Yield
0% (Set median yields)	*(0-23%)*	93.7%
2.4%	*(50%)*	93.3%
6.0%	*(75%)*	92.7%
11.7%	*(90%)*	91.8%
18.6%	*(97.5%)*	90.8%

Sample's yield of 93.7% is used as the normalizing benchmark. Projections apply the pay cycle regression estimate to sample quantiles of turnover.

diminish by more than 17.4 percentage points, more than halving the number of units that would have passed all station tests at the first instance. The instrumented estimates in (1) and (3) are substantially larger than the (OLS) estimate, which is consistent with managers tending not to place new hires onto lines already hampered by delays.

Using instruments, we estimate that heavier workloads diminish first-pass yields. In contrast, in the underlying data, (OLS) demonstrates how elevated worker-hours significantly and positively associate with higher first-pass yields. Such positive association defies easy explanation other than by managers routing work to smoother-running lines, and the lagged turnover instrument appears to correct for the resulting endogeneity bias.

4.3. Robustness Checks

We conduct three robustness checks. Our sample of line-weeks covers all assembly lines during normal levels of production activity. Of the three buildings covered by the study, Building C operates assembly lines for 31 of 40 weeks, while Buildings A and B operate throughout the study. Table EC.9 in Appendix D replicates the analyses of Tables 9a, 12, and 14a after excluding Building

C. We find overall similar results, although by excluding nearly a third of the assembly lines in our sample, we lose statistical power particularly for discerning the effects of production speeds on first-pass yields.

Second, we re-run, and find nearly identical results for, the analysis of Table 14a using a version of the first-pass yield measure that excludes failed units (see p.14). See Table EC.10.

Lastly, despite our extensive conversations with the firm's operations specialists to document productive interventions, managers may mitigate worker turnover in additional ways not captured by the control variables we use. Such unobserved mitigations would be omitted variables in our regressions. Notably, the Sargan tests do not detect any statistical evidence of such omitted variables, which if present would violate the exclusion restriction.

We provide some additional assurance, by considering a special situation where we expect managers *should* be applying such interventions. The measured strength of turnover's effect on productivity in these situations offers a lower bound on the size of the effect without mitigation (Hypothesis 1).

Such a situation is offered by the seasonal influence of the manufacturer's pay cycle on worker turnover, which Figure 10 shows. Managers expect that turnover elevates following payweek, enabling them to fully intervene. To measure the effect of pay cycle turnover, we regress log yields on pay cycle turnover (worker turnover predicted by weeks to payweek), while including line fixed effects. Notably, we choose *not* to include any controls for known interventions. Instead, we *assume* that the manufacturer *does* intervene to mitigate pay cycle turnover's effects in ways that we both can and cannot observe. We estimate the effect of pay cycle turnover net of all mitigations, which conservatively lower bounds the size of the unmitigated effect.

Table 15a reports pay cycle turnover's statistically significant impact on yields, projecting economic costs of conservatively \$146 million in waste, which is comparable with previous results. The robustness check suggests that our previous analyses do not omit significant mitigations as controls. Regarding why the observed controls appear to do little for pay cycle turnover, we suggest that intervening becomes challenging when pay cycle turnover occurs facility-wide.

4.4. Aggregating Productivity Effects

We find empirical support for all four hypotheses. By eroding tacit know-how held by employees, turnover diminishes yields by [0.5-0.7] percentage points on average, and its unmitigated effects can cost the firm $178 million ($146 million under robustness check) in material waste accumulated over the life cycle of the studied device. In the remaining sections, we close the loop by incorporating turnover's effects into the responsive manufacturer's production planning and performance.

5. Structural Model and Estimation

We structurally model the operational management of turnover's disruptive effects. Disruptive turnover harms manufacturing yields, introduces yield volatility, and ultimately impairs the manufacturer's ability to increase its output rate in response to arriving orders.[30]

Our production planning model permits the firm to handle disruptive worker turnover using operational levers, after fixing its wage schedule. First, the firm uses inventories. Inventories prevent the responsive manufacturer from incurring backorders and (through pre-building) allow it to flexibly decide when production takes place. However, productivity-preserving flexibility comes at a cost. For example, pre-building to avoid disruptive pay cycle turnover means that production fluctuates cyclically, increasing labor costs because either capacity matches the production peaks (and its utilization necessarily falls) or overtime is used. Supply chain costs may increase as orders placed to suppliers become more volatile. More generally, the responsive manufacturer *hedges* through the workforce, by maintaining a relatively larger workforce and utilizing overtime.

We accordingly model the manufacturer's production planning problem to flexibly control when production occurs and by whom (i.e., deciding weekly production quantities, production speeds, and staffing). To optimize how work and individual employees are allocated to specific assembly lines, we combine reinforcement learning (Osband et al. (2017)) and large-scale convex optimization. Specifically, reinforcement learning decides the firm's weekly production quantities while learning

[30] For the remainder of the paper, we abstract away from the distinction between the manufacturer and firm and the setting-specific manner in which they apportion costs and risks.

a weekly cost (reward) function that itself derives from solving a convex optimization problem over hundreds of thousands of decision variables (e.g., who sits where).

After estimating the model, we carry out counterfactual analyses to investigate the operational value of stabilizing the workforce, including through a wage hike like Henry Ford's. By reducing the need to hedge in the first place and making production more reliable, could higher wage rates actually reduce costs? Because the answer depends on workers' incentives, our structural model of production planning incorporates the factory worker's perspective in considering turnover.

Section 5.1 overviews the model. Sections 5.2 and 5.3 model workers' decisions to either continue in their jobs or to exit the firm. We model workers' incentive-compatible voluntary turnover, as involuntary turnover is legally curtailed and rare. (See Section 3.) Section 5.5 estimates workers' preferences. Building on these sections, Section 5.6 models dynamic production planning that includes weekly decisions on production quantities, workloads, and staffing, while accounting for their effects, alongside compensation, on worker turnover under equilibrium incentives.

Our study has limitations. As with any structural estimation study, we make certain modeling assumptions about workers' preferences based on the setting and their rationality in expectations. We lack data regarding workers' outside options for employment, especially competitor's wage responses. However, Appendices A.1 and A.2 describe why wage hikes appear effective in practice.

5.1. Model Overview

Each week t's events are organized into the following sequence: (1) The Master Production Schedule and clear-to-build order quantities are revised. (2) Workers $i \in W_t$ (set of workers active during the prior week) learn their private exit values, ϕ_{it}, and independently and simultaneously decide whether to exit the firm to collect ϕ_{it}. (3) The manufacturer decides assembly line workloads and staffing, including placing new workers. (4) Production occurs. The firm realizes variable production costs, and workers receive utility flows based on individual compensation and workplace states.

5.2. Equilibrium Model of Worker Turnover

In the model's weekly stage (2), workers make incentive-compatible decisions to either continue in their jobs or to exit the firm. Workers act on preferences over work hours, line speed, and their workgroup attrition rates as well as income, which we refer to collectively as the *workplace state*, and preferences can vary with experience as well as by gender and worker type (i.e., status as regular or dispatch employee).

Each week, a worker stays with the firm if and only if her expected, present-valued utility of staying, VC_i, exceeds her exit value, ϕ_{it}, collected upon leaving in week t:

$$\text{Worker } i \text{ exits in week } t \iff VC_i(X_{it}) < \phi_{it}. \tag{6}$$

The exit value is drawn from an exponential distribution, the mean λ of which can depend on whether workers are returning from a holiday.[31] She arrives at VC_i by forming beliefs about her future workplace states (e.g., whether they are high-utility states or low-utility states) based on her available information, denoted X_{it}.

More formally, worker i receives the weekly utility flow U (expressed in monetary RMB):

$$U(Z_{it}, p_{it}; \theta_i) = p_{it} + Z_{it}^T \theta_i, \tag{7}$$

where p_{it} is disbursed pay (if payweek) and Z_{it} is her workplace state in week t. The vector of coefficients θ_i represents her preferences over states, and the non-income component of her utility flow is $Z_{it}^T \theta_i$. Thus, if worker i were to know her sequence of future states, $\{p_{it}, Z_{it}\}$, $\{p_{i,t+1}, Z_{i,t+1}\}$, $\{p_{i,t+2}, Z_{i,t+2}\}, \ldots$, then she could anticipate with certainty the utility flow she would receive in each period: $U(Z_{it}, p_{it}; \theta_i)$, $U(Z_{i,t+1}, p_{i,t+1}; \theta_i)$, and so on.

When worker i decides whether to exit in week t, she knows the amount of any pay p_{it} she is due to receive (pay cycle delay fixes that amount two weeks in advance), but has yet to learn her pending workplace state, Z_{it}. (She may even be transferred to a different assembly line.) To predict Z_{it} and beyond, she uses the information of her previous week's workplace state, her accruing

[31] In modeling call center attrition, Emadi and Staats (2018) additionally allow agents' exit values to depend on macroeconomic factors, including national rates of GDP growth, inflation, and unemployment. While macroeconomic influences on worker turnover are interesting, such factors showed minimal empirical variation during our shorter study period. For example, China's reported unemployment rate stayed between 4.0-4.1% over 2010 to 2016. Future work could compare workers' exit rates across products manufactured under differing macroeconomic conditions.

compensation, and the MPS update. Formally, her information set, $X_{it} \in \mathbb{X}$, consists of: (1) $Z_{i,t-1}$, her workplace state in week $t-1$ (plus that week's yield and share of new operators); (2) weeks to payweek; (3) her accrued pay to date and her base rate of hourly pay; and (4) week t's MPS trend. For notational convenience, because $Z_{i,t-1} \subset X_{it}$ (i.e., today's information includes last week's workplace state), we use $U(X_{it};\theta_i)$ to mean last week's utility received, $U(Z_{i,t-1}, p_{i,t-1};\theta_i)$.

Given worker i's available information $X \in \mathbb{X}$, her present-valued utility to stay, $VC_i(X)$, derives from the system of Bellman equations:

$$VC_i(X) = \mathbb{E}_{X',\phi'|X}\left[U(X';\theta_i) + \delta \cdot \max\{\phi', VC_i(X')\}\right] \text{ for all } X \in \mathbb{X}. \tag{8}$$

Worker i's available information X anticipates next week's new information, X', which includes her pending workplace state if she stays. Should she stay, she receives the utility flow $U(X';\theta_i)$ and, under the weekly discount rate δ, the higher of next week's value to stay, $VC_i(X')$, and next week's exit value, ϕ'. Without loss of generality, we typically denote the preference parameters as θ instead of θ_i, by pushing worker i's observable heterogeneity into her payoff-relevant workplace state, $Z_{i,t}$. Appendix E further details the optimal stopping model of worker turnover.

From worker i's perspective, her pending workplace state (e.g., her chances of being transferred to another assembly line and working overtime) depends on her co-workers' turnover levels. With the aims of analytical tractability and of limiting the sophistication required for workers' decision-making, we use the Experience-Based Equilibrium (EBE) proposed by Fershtman and Pakes (2012) to model workers' inter-related turnover decisions. The expectation operator of Bellman equations (8) uses the probabilities for workplace state transitions that result from co-workers' equilibrium rates of turnover.

5.3. Experience-Based Equilibrium

We define our model's restricted Experience-Based Equilibrium and assure existence.

Definition 1. A restricted Experience-Based Equilibrium formally consists of:

- A recurrent subclass $\mathcal{R}$ within the underlying Markov state space $\mathcal{S}$, where recurrence holds under the transition kernel incorporating equilibrium strategies m^*;

- Equilibrium turnover strategies $m^*: \mathbb{X} \times \mathbb{R}_+ \to \{0,1\}$, which map all feasible information sets $X \subset S \in \mathcal{S}$, coupled with private exit values $\phi \in \mathbb{R}_+$, into the agent's pure action space, 1 for turnover and zero otherwise;
- Defined over the same domain of information sets, the expected discounted values $VC_i: \mathbb{X} \to \mathbb{R}$ of the agent's utility flows conditional on staying;
- Optimality of agents' equilibrium turnover strategies, m^*, such that:

$$\text{Agent } i \text{ exits in week } t \iff VC_i(X_{it}) < \phi_{it}; \tag{9}$$

- Lastly, consistency of agents' expected continuation values:

$$VC_i(X) = \mathbb{E}_{X',\phi'|X}\left[U(X';\theta_i) + \delta \cdot \max\{\phi', VC_{\theta_i}(X')\}\right], \tag{10}$$

  for all information sets $X \in \mathbb{X}$ recurrent under equilibrium play, i.e., $X \subset S \in \mathcal{R}$. As before, X' and ϕ' denote the next period's information set and exit value.

Proposition 1. *A restricted Experience-Based Equilibrium (EBE) exists for our model.*

Proposition 1's proof and remarks on the EBE are available on the first author's website.

5.4. Estimation Procedure

We estimate the model's worker preferences by adapting Bajari et al. (2007)'s simulation-based, minimum-distance estimator to restricted EBE. We find similar results from Pakes et al. (2007)'s estimation approach modified by constraint-sampling techniques (De Farias and Van Roy (2003, 2004), Farias et al. (2012)) needed to computationally handle an immense state space (i.e., sparse transition matrix exceeding 1TB). As Appendix E.3 explains, both estimation approaches are computationally intensive, e.g., simulating 200 million sample paths (12 billion including bootstrap) for the results below, and both flexibly (nonparametrically) estimate the workplace state's equilibrium transition probabilities. In estimating worker preferences, we do not assume that the firm follows Section 5.6's production planning model.

The simulation-based, minimum-distance estimator exploits that workers' observed exit timing, formalized as an exit policy, should outperform alternative policies when workers' rewards are computed and compared by simulation. The critical idea is to recover the workers' exit policy from the data in the form of state-dependent exit thresholds.

Incentive-compatibility constraint (9) dictates that worker i exits if and only if her exit value ϕ_{it} in week t exceeds her expected value to stay. Without loss of generality, we normalize her exit value by the state-dependent mean of its distribution, resulting in an equivalent threshold rule. Under this rule, the worker exits if and only if her normalized exit value $\xi_{it} := \phi_{it}/\lambda_{\theta_i}(X_{it})$ exceeds the re-scaled threshold T:

$$\text{Worker } i \text{ exits in week } t \iff T(X_{it}) := \frac{VC_{\theta_i}(X_{it})}{\lambda_{\theta_i}(X_{it})} < \xi_{it}, \tag{11}$$

where the normalized exit values ξ_{it} are now standard exponential. Because worker characteristics and assembly line conditions are incorporated into state X_{it}, T can be thought of as custom-tailoring the exit thresholds to workers and conditions so as to standardize the exit values.

Worker i's probability of exit in week t is the probability of the event $T(X_{it}) < \xi_{it}$. Given first-stage estimates, $\hat{\rho}(X_{it})$, of the workers' state-dependent probabilities of exit, we invert ξ_{it}'s exponential CDF to recover the set of estimated exit thresholds:

$$\hat{T}(X) = -\log(\hat{\rho}(X)) \text{ for all } X \in \mathbb{X}. \tag{12}$$

In the subsequent simulation steps, $\hat{T}(X)$ is sufficient for evaluating the policy's payoffs. In each encountered state X, a simulated exit value ξ is compared against $\hat{T}(X)$ to decide whether the worker exits, and the resulting rewards are collected and recorded for each simulation run.

Appendix E.3 provides a detailed estimation procedure. Through 200 million simulations covering 100,000 perturbed policies, the estimated preferences are chosen to minimize violations where the observed policy represented by $\hat{T}$ is outperformed in expected reward by a perturbed policy. Bootstrap confidence intervals are obtained using an additional 12 billion simulations.

Table 16 Structural Estimates of Worker Preferences

	Utility (RMB) by worker type	
	Regular employee	Dispatch employee
Workplace state (weekly)		
Gender (F)	22.0^{****}	47.9^{**}
Experience in days	-0.23^{****}	-3.15^{****}
Marginal weekday hour	-11.2^{**}	-10.0^{**}
Marginal weekend or holiday hour	-58.8^{*}	-54.7^{****}
Assembly line workrate (UPH)	1.7	4.1^{**}
Attrition rate	-27.1^{*}	-21.6^{***}
Turnover present value (mean)	$8,797^{\dagger}$	$4,087^{\dagger}$
Convenience to leave at contract end	N/A	135
Exit by transfer (next model)		
Transfer present value	$8,246^{\dagger}$	$8,579^{\dagger}$

Significance levels → **** - 0.001 *** - 0.01 ** - 0.05 * - 0.1

415,859 employee-weeks for 47,579 compensation-matched employees on the product's FATP lines during weeks of normal activity within Sept. 2014 – Jun. 2015. All utility payoffs are expressed in terms of RMB monetary equivalents. No holiday effect (Chinese New Year not considered) on exit values is found for either type of hire. Superscript † indicates parameters that were constrained to be non-negative by the model.

5.5. Estimation Results

Table 16's two columns show the resulting estimates of workers' preferences, normalized to RMB, for each worker type (regular and dispatch). Recall for context that workers average 1080 RMB in total pay per week worked.

We highlight several findings. First, workers are sensitive to workloads and schedules. Both types of workers prefer weekday overtime, yet starkly dislike weekend and holiday overtime despite its 33% higher hourly compensation. Specifically, working a marginal weekday hour incurs 10-11 RMB worth of disutility, which is balanced by the base salary's hourly compensation of 10.6 RMB. At 15.9 RMB hourly, overtime compensation under the base salary exceeds workers' disutility of work for weekdays. In contrast, workers incur 55-59 RMB in disutility for weekend and holiday work, which well exceeds its hourly overtime compensation (21.2 RMB hourly for weekends and 31.7 RMB for holidays). In our data, 91% of line-weeks include some weekday overtime, whereas about 80% of weekends entail five or more overtime hours. The standard deviation in weekend work is 6.7 hours per shift, i.e., 364-389 RMB of worker disutility before compensation. These findings suggest that overtime can influence and even calibrate worker turnover. Because weekend work is highly compensated, evidence of aversion from workers surprised our data partner.

Table 17 Experienced Turnover by Worker Type

For dispatch employees		
	Duration	Extra monthly compensation to cap exit rate at new hires' exit rate
Average experience level	8.2 weeks (57 days)	773 RMB
Median experience level	6.9 weeks (48 days)	651 RMB
Under current retention bonuses: Never retained at same rate as new hires		
For regular employees		
	Duration	Extra monthly compensation to cap exit rate at new hires' exit rate
Average experience level	28 weeks (196 days)	191 RMB
Median experience level	23 weeks (162 days)	158 RMB
Under current retention bonuses: Retained at same rate as new hires at 309 days of experience		

Based on estimated worker preferences shown by Table 16.

Second, workers become difficult to retain as they gain experience. In Table 17, we show the extra monthly pay needed to keep experienced workers' exit rates no higher than new hires'. A regular employee at the median 23 weeks of experience requires 158 RMB, while a dispatch worker at the dispatch median of 7 weeks already requires 651 RMB. The firm's retention bonuses contain regular workers' exit rates below those for new hires through 309 days of experience, while exit rates always increase with experience for dispatch workers.

Third, little else distinguishes regular and dispatch workers' preferences. Dispatch employees' exit values are appreciably lower, suggesting poorer outside employment options in the mainland's interior provinces to which they typically return. While workers dislike weekend and holiday labor, we find their exit values unaffected by holidays. Dispatch workers suffer a mild inconvenience, 135 RMB, to convert to regular employment at dispatch contract's end.

Fourth, our evidence suggests that workers are affected by coworkers' attrition.[32] For workers who experience an average level of weekly attrition instead of none, the week's workplace experience devalues by 167-210 RMB (87-109 RMB at median attrition).

Finally, turnover is gender-imbalanced, and each hired cohort trends female over time. At hire, 19.0% of dispatch workers and 30.5% of regular workers are female. At 180 days of employment, they are 27.4% and 34.5% female.

[32] Our data cannot rule out alternative explanations for turnover on individual lines occurring "clumpily".

5.6. Workforce and Production Planning

In stage (3) of the model's weekly sequence of events (Section 5.1), the firm optimizes its planning decisions in two steps. In the first step, the firm decides how much to produce. In deciding the production quantity, it takes stock of existing inventories and anticipated build orders. It can pre-build or compensate for anticipated yield by chosing a level of production that exceeds the orders due. Reinforcement learning is used to solve the dynamic optimization problem of deciding weekly assembly line-level production quantities.

In the second step, the firm implements its decided production by staffing and managing the workgroups. The firm decides each workgroup's staffing and workrate, hence hours. To achieve tractability, we organize these workforce planning decisions into a large-scale, convex optimization problem. All production costs are incurred in this step's workforce planning problem, which feeds back into the first step's reinforcement learning problem as its weekly cost (reward) function.

`Decision variables` are denoted in `type font`, except when inside vectors, and we underline the first step's $\underline{\texttt{quantity decision variables}}$. The main decision variables of the first step's production planning problem are the week's volume of production, denoted $\underline{\texttt{Vol}_t}$, and the (log) share of production, $\underline{\texttt{logVolShare}_{lt}}$, for each assembly line $l \in L$. To determine the weekly reward, the convex optimization for the second step's workforce planning problem solves for hundreds of thousands of decision variables, including the staffing of the assembly lines by individual workers.

The remainder of this section formulates the weekly cost minimization problem, i.e., step two. Appendix F.2 solves for production quantities using recent reinforcement learning techniques (Osband et al. (2017)).

5.6.1. Stage Reward as Convex Objective. Each week, the manufacturer incurs variable costs (A) from producing, including from compensating workers and from incurring yield losses, and (B) from stock outs and lost sales when falling short of the Master Production Schedule's order quantities (i.e., underage costs).

For production costs, c_M is the cost of waste incurred when yield Y_{lt} incrementally reduces output by one unit. For underage, the state variable MPSNeed_t represents week t's order quantity plus backlog, and c_U is the per-unit underage cost when production is short of MPSNeed_t.[33] The planner minimizes the firm's weekly variable costs, plus its cost-to-go V_{t+1}^{Firm}:

$$\textbf{Minimize} \quad \overbrace{c_M \times \underline{\texttt{Vol}_t} \times \sum_{l \in L} \Big[(1 - \mathbb{E}Y_{lt}) \times \exp\{\underline{\texttt{logVolShare}_{lt}}\}\Big]}^{\text{Expected cost of material waste (lost yield)}} + \overbrace{\sum_{i \in W} \text{WagesPaid}_{it}}^{\text{Compensation if payweek}} + \cdots$$
$$\underbrace{c_U \times \mathbb{E}\Big[\text{MPSNeed}_t - \underline{\texttt{Vol}_t} \times \sum_{l \in L} \exp\{\underline{\texttt{logVolShare}_{lt}}\} \times Y_{lt}\Big]^+}_{\text{Expected underage cost (producing short of weekly MPS and backlog)}} + V_{t+1}^{Firm}, \tag{13}$$

In step two, the production quantities are fixed state variables, and the problem's decision variables are those affecting the distribution of yields Y_{lt} as random variables through the constraints (these effects are specified using Section 4's empirical estimates), and affecting the accrual of wages to be paid in the future. Appendix F.2 describes how simulations "learn" the cost-to-go, V_{t+1}^{Firm}.[34]

5.6.2. Workgroup activity decisions. Assembly lines' hours and speeds (units per hour) must respect their previously decided production quantities.

$$\text{For each line } l \in L: \quad \log(\underline{\texttt{Vol}_t}) + \underline{\texttt{logVolShare}_{lt}} = \texttt{logSpeed}_{lt} + \texttt{logHours}_{lt}. \tag{14}$$

Hours and speeds are additionally bounded within the ranges observed in our data.

5.6.3. Workforce staffing decisions. Indicator variable, $\texttt{assign}_{il}$, takes the value 1 if worker i is assigned to assembly line l, and zero otherwise. Omitting the week t subscript, let W be the set of last week's workers who did not turnover, W_l the subset previously staffing line l, and T and N the sets of newly available operators who are experienced (from the prior device generation) and inexperienced (new hires), respectively. The distribution of the assembly lines' yields, Y_{lt}, depends on the staffing decisions, $\{\texttt{assign}_{il},\ i \in W \cup T \cup N, l \in L\}$. The following constraints derive attrition rates, $\texttt{attritionrate}_{lt}$, and new hire staffing rates, $\texttt{newrecruitpercentage}_{lt}$:

[33] Except for the terminal week at 90%, we set per-unit underage to 40% of a lost sale. The firm's optimized production planning behavior is similar even when simulated underage costs are significantly less.

[34] We cannot guarantee this step's cost-to-go is convex, because the state transition includes workers' equilibrium turnover. To preserve convexity, we approximate the cost-to-go as an affine function under reinforcement learning.

$$\forall i \in W \cup T \cup N: \quad \sum_{l \in L} \texttt{assign}_{il} = 1 \qquad \text{(one assignment per head)}$$

$$\forall l \in L: \quad \texttt{lineexperience}_{lt} = \sum_{i \in W \cup T} \text{experience}_{it} \times \texttt{assign}_{il}$$

$$\forall l \in L: \quad \texttt{retainedexperience}_{lt} = \sum_{i \in W_l} \text{experience}_{it} \times \texttt{assign}_{il}$$

$$\forall l \in L: \quad \text{lineexperience}_{l,t-1} \times \left(1 - \texttt{attritionrate}_{lt}\right) = \texttt{retainedexper}_{lt}$$

$$\forall l \in L: \quad \text{headcount}_{lt} = \sum_{i \in W \cup T \cup N} \texttt{assign}_{il}$$

$$\forall l \in L: \quad \texttt{newrecruitcount}_{lt} = \sum_{i \in N} \texttt{assign}_{il}$$

$$\forall l \in L: \quad \text{headcount}_{lt} \times \texttt{newrecruitpercentage}_{lt} = \texttt{newrecruitcount}_{lt}$$

Line l's attrition rate, $\texttt{attritionrate}_{lt}$, is the fraction of last week's assigned experience stock that is not retained. Post-optimization, lines' experience stocks, $\texttt{lineexperience}_{lt}$, become state variables for week $t+1$'s optimization. Headcounts are constrained to balance across assembly lines, and variable range bounds are omitted for exposition. For tractability, the binary assignment variables, $\texttt{assign}_{il}$, are relaxed to real values between zero and one. We solve the relaxed problem to optimality and assign workers by treating their line-assignment values as probability weights.[35]

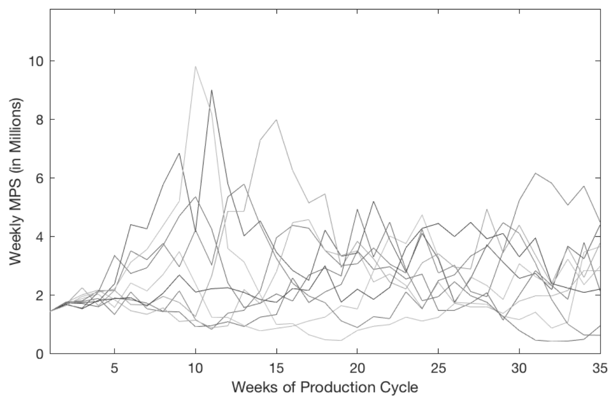
Step two's framework readily accommodates additional constraints, e.g., that specialized worker types meet required staffing levels or percentages, that may be important in other settings.

6. Simulated Counterfactuals

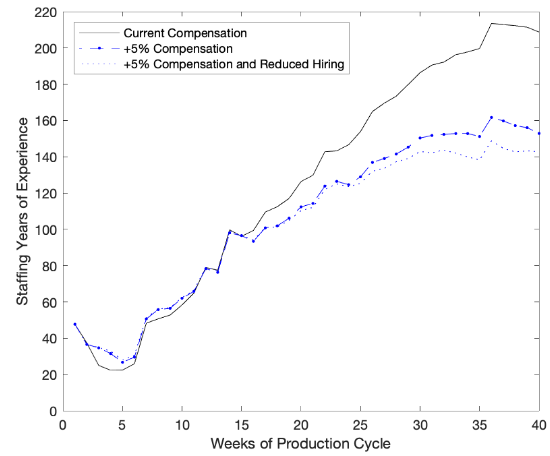
Our simulation analysis evaluates worker turnover's disruptive impact on the responsive manufacturer's production. Our analysis evaluates manufacturing performance under three counterfactual scenarios: (1) under the firm's current arrangements for wages and hiring, (2) under a counterfactual, across-the-board wage hike of 5%, and finally (3) under the counterfactual wage hike while curtailing hiring by 50% over the second half of the product's life cycle. Through these scenarios, we investigate the value of stabilizing the workforce and reducing the need to operationally hedge against disruptive turnover using inventories, extra workforce capacity, and overtime.

[35] Typically, only a subset of new workers receive nontrivial assignment weights split over a few lines. Existing workers typically receive de facto assignments from $\texttt{assign}_{il}$ values nearly indistinguishable from 1.

Figure 18 Simulated Counterfactuals over 40-week Production Cycle



(a) Weekly MPS Sample Paths



(b) Projected Experience Stocks

Ten simulated sample paths of weekly MPS quantities are displayed, over 35 full-activity weeks of production. For each compensation policy, displayed experience stocks average weekly outcomes from 100 sample paths.

To simulate order arrivals, we fit an autoregressive vector time series to the MPS order quantities. In each simulation run, order quantities are incrementally revised and revealed to the firm as a rolling horizon. Figure 18a plots ten realizations of simulated MPS. The firm's problem is to minimize the expected costs of satisfying stochastic sequences of orders like these.

All counterfactual simulations allow inventory pre-building (that is, producing more than the current week's order quantity and backlog), without inventory costs. Our resulting comparison is conservative, as the value of stabilizing the workforce is smaller when disruptions can already be contained in part through inventories at no holding cost.

Results. Table 19 summarizes each policy's results through outcomes averaged from 100 simulation runs. The simulation appears well-calibrated for the status quo at the current wage: Assembly line yields' median, mean, and standard deviation all virtually match those in our data, which are shown by Table 2. The projected and empirical rates of attrition are likewise similar. Comparatively, the simulated policy tends to avoid weekend labor, which workers dislike.

From the status quo simulation, we benchmark the costs and patterns of production. First, at nearly \$1.8B, underage costs substantially exceed both material waste at \$1B and wage compensation at merely \$81 million. Thus, it may be in the firm's interests to focus its compensation policy

Table 19 Counterfactual Production and Staffing

			5% Wage Increase	
		Current Wage	Same Hiring	Reduced Hiring
Production	Assembly Line Yield			
	Mean	93.5%	93.6%	93.7%
	Median	93.7%	93.7%	93.7%
	Std. dev.	1.7%	1.2%	1.1%
	Std. dev. within-week variation	1.4%	1.2%	1.0%
	Std. dev. between-week variation	0.7%	0.3%	0.3%
	Surplus Produced (% over weekly need)			
	Mean	96%	113%	111%
	Std. dev.	45%	17%	21%
	Std. dev. between-week variation	13%	6%	9%
Staffing	Attrition Rate (mean)	7.9%	9.9%	9.4%
	Experience Stock (in worker-years)			
	Mean (assembly line)	126	102	102
	Weeks 1 to 20 / Weeks 21 to 40	66 / 185	66 / 137	68 / 136
Utilization	Hours (mean)	47.3	45.7	45.8
	Weeks 1 to 20 / Weeks 21 to 40	45.3 / 49.2	43.9 / 47.6	45.2 / 46.4
	Weekday overtime utilization (mean)	35.9%	33.1%	32.3%
	Weeks 1 to 20 / Weeks 21 to 40	31.3% / 40.4%	28.1% / 38.0%	30.3% / 34.4%
	Weekend utilization (mean)	17.8%	14.5%	15.1%
	Weeks 1 to 20 / Weeks 21 to 40	13.2% / 22.4%	10.7% / 18.4%	13.4% / 16.8%
Costs	Wage Compensation	\$81M	\$70M	\$68M
	Underage Cost	\$1,772M	\$1,196M	\$1,129M
	Material Cost of Lost Yield	\$1,018M	\$999M	\$995M
	Total Cost	\$2,789M	\$2,195M	\$2,124M

For 100 sample paths simulated under each policy.

on controlling underage rather than controlling wages. Moreover, theory predicts that the manufacturer should strategically pre-build inventories in response to yield variation, e.g., Figure 11's right panel. The status quo's standard deviation in production surplus is 45%, implying that within within one standard deviation of the average, production ranges from 51% to 141% of weekly need (the weekly order quantity plus backlog). The wide swings in production surplus and deficit reflect how the firm's inventories strategically shift production activity between weeks, because producing surpluses in some weeks enables strategic deficits in others.

We compare this baseline against outcomes under the two counterfactual firm policies, with three high-level questions in mind. First, do workforce planning policies significantly affect production, especially by reducing underage costs? Second, do higher wages actually improve yields? Taken in isolation, lower attrition should improve yields, yet because managers already route work to the most productive assembly lines and weeks, the benefit is dampened. Additionally, higher wages may not improve retention much if marginal improvements reduce attractive overtime. Lastly, we examine whether effective workforce planning reduces the firm's operational hedging costs.

Offering workers 5% higher wages indeed alters production activity. After the workers and the firm adjust to their new trade-offs, yields improve modestly but become significantly more reliable. Yield variability drops both within weeks (i.e., yields diverge less between assembly lines producing in the same week) and between weeks. Reliable yields translate into steadier production. Because the firm need not sequester production into relatively more productive weeks, its production surplus varies significantly less, ranging from 96% to 130% within a standard deviation of the average. More balanced utilization improves the firm's ability to flexibly meet demand and cuts expected underage costs by $576 million. Illustrating the importance of the workforce in production planning, the policy's impact on underage outweighs the firm's *entire* expenditure on wages.

Compensation's influence on workforce dynamics is shown in Figure 18b, which plots the accruing stocks of worker experience under the status quo and the 5% wage increase scenarios. Higher wages improve retention during production's first ten weeks, when the product's assembly lines are being seeded with workers. However, the experience stock then settles at a *lower* level than before (see also Table 19), saving the firm nearly 14% in labor costs. That the workforce shrinks in response to higher wages is arguably counterintuitive, when wage hikes are intended to improve retention. However, consider the hedging perspective. As the higher wage stabilizes production, the manufacturer *should* reduce its labor-based capacity hedge resulting in a leaner workforce. Because the counterfactual scenario does not permit hiring reductions, the firm manipulates overtime to incentivize and control attrition on its assembly lines (for a leaner workforce, the attrition rate necessarily exceeds the status quo). As in Kesavan et al. (2018), work scheduling influences worker turnover, even as compensation's effect varies across industries (e.g., Emadi and Staats (2018) find little effect in call centers).

All told, the 5% wage hike saves the responsive manufacturer $594 million, or 21% of variable costs, by significantly curtailing underage costs. As the third column of Table 19 shows, reducing hiring in tandem, by half over weeks 21-40, stabilizes yields further and cuts underage by an additional \$67 million. The total savings grow by another 2.6%. Attrition decreases, but remains higher than the status quo baseline; the firm would benefit from scaling back hiring further.

7. Concluding Remarks

We study manufacturing worker turnover using comprehensive data regarding a responsive manufacturer's production plans, station productivity, and staffing and compensation covering 52,214 workers. Despite standardization and deskilling in assembly manufacturing, empirical tests consistently support that employees embody substantial tacit know-how. For the product we study, worker turnover's yield disruptions can cost an estimated \$146-178 million in waste.

Prescriptively, this paper considers the role of uncertainty and hedging in workforce planning. Policies improving worker retention make production more reliable. Thus while better compensation incentivizes workers to stay, the workforce could counterintuitively *shrink* as the firm sheds labor capacity that previously hedged against unreliable production. Indeed, when the five-dollar day reduced Ford's turnover by 87% in 1914, its workforce became 11% leaner in the same year (Raff and Summers (1987)).[36] In our setting, increasing wages by 5% reduces underage costs by \$576 million, compensation by \$11 million, and overall variable costs by \$594 million (21%).

For the responsive manufacturer, a turnover-prone workforce generates productive disruptions, whereas a stable workforce supports its ability to produce flexibly. In producing flexibly, managing the firm's workforce may be as critical as any other source of flexibility in the supply chain.

References

Adlakha, Sachin, Ramesh Johari, Gabriel Y Weintraub. 2015. Equilibria of dynamic games with many players: Existence, approximation, and market structure. *J. Econom. Theory* **156** 269–316.

Akşin, Zeynep, Bariş Ata, Seyed M Emadi, Che-Lin Su. 2013. Structural estimation of callers' delay sensitivity in call centers. *Management Sci.* **59**(12) 2727–2746.

Argote, Linda. 2013. *Organizational learning: Creating, retaining and transferring knowledge*. 2nd ed. Springer Science & Business Media.

[36] This may also attribute to concurrent innovations in the assembly process and technology.

Argote, Linda, Sara L Beckman, Dennis Epple. 1990. The persistence and transfer of learning in industrial settings. *Management Sci.* **36**(2) 140–154.

Arthur, Jeffrey B. 1994. Effects of human resource systems on manufacturing performance and turnover. *Acad. Management J.* **37**(3) 670–687.

Bajari, Patrick, C Lanier Benkard, Jonathan Levin. 2007. Estimating dynamic models of imperfect competition. *Econometrica* **75**(5) 1331–1370.

Balseiro, Santiago R, Omar Besbes, Gabriel Y Weintraub. 2015. Repeated auctions with budgets in ad exchanges: Approximations and design. *Management Sci.* **61**(4) 864–884.

Bloom, Nicholas, Benn Eifert, Aprajit Mahajan, David McKenzie, John Roberts. 2013. Does management matter? Evidence from India. *Quart. J. Econom.* **128**(1) 1–51.

Borgatti, Stephen P, Rob Cross. 2003. A relational view of information seeking and learning in social networks. *Management Sci.* **49**(4) 432–445.

Brandes, Stuart D. 1976. *American Welfare Capitalism, 1880-1940*. University of Chicago Press.

Bray, Robert L, Haim Mendelson. 2012. Information transmission and the bullwhip effect: An empirical investigation. *Management Sci.* **58**(5) 860–875.

Cappelli, Peter, Keith Chauvin. 1991. An interplant test of the efficiency wage hypothesis. *Quart. J. Econom.* **106**(3) 769–787.

CEES. 2017. How are Chinese manufacturing firms coping with rising labor costs: A report of China employer-employee survey (2015-2016) .

Darr, Eric D, Linda Argote, Dennis Epple. 1995. The acquisition, transfer, and depreciation of knowledge in service organizations: Productivity in franchises. *Management Sci.* **41**(11) 1750–1762.

De Farias, Daniela Pucci, Benjamin Van Roy. 2003. The linear programming approach to approximate dynamic programming. *Oper. Res.* **51**(6) 850–865.

De Farias, Daniela Pucci, Benjamin Van Roy. 2004. On constraint sampling in the linear programming approach to approximate dynamic programming. *Math. Oper. Res.* **29**(3) 462–478.

Egelman, Carolyn D, Dennis Epple, Linda Argote, Erica RH Fuchs. 2017. Learning by doing in multiproduct manufacturing: Variety, customizations, and overlapping product generations. *Management Sci.* **63**(2) 405–423.

Emadi, Seyed M, Bradley R Staats. 2018. A structural estimation approach to study agent attrition. *SSRN Working Paper* (No. 3180225).

Epple, Dennis, Linda Argote, Kenneth Murphy. 1996. An empirical investigation of the microstructure of knowledge acquisition and transfer through learning by doing. *Oper. Res.* **44**(1) 77–86.

Farias, Vivek, Denis Saure, Gabriel Y Weintraub. 2012. An approximate dynamic programming approach to solving dynamic oligopoly models. *RAND J. Econom.* **43**(2) 253–282.

Fershtman, Chaim, Ariel Pakes. 2012. Dynamic games with asymmetric information: A framework for empirical work. *Quart. J. Econom.* **127**(4) 1611–1661.

Fine, Charles H, Robert M Freund. 1990. Optimal investment in product-flexible manufacturing capacity. *Management Sci.* **36**(4) 449–466.

Fisher, Marshall, Janice Hammond, Walter Obermeyer, Ananth Raman. 1997. Configuring a supply chain to reduce the cost of demand uncertainty. *Prod. Oper. Management* **6**(3) 211–225.

Fisher, Marshall, Ananth Raman. 1996. Reducing the cost of demand uncertainty through accurate response to early sales. *Oper. Res.* **44**(1) 87–99.

Fisher, Marshall L. 1997. What is the right supply chain for your product? *Harvard Bus. Rev.* **75** 105–117.

Fisher, Marshall L, Christopher D Ittner. 1999. The impact of product variety on automobile assembly operations: Empirical evidence and simulation analysis. *Management Sci.* **45**(6) 771–786.

Gaur, Vishal, Marshall L Fisher, Ananth Raman. 2005. An econometric analysis of inventory turnover performance in retail services. *Management Sci.* **51**(2) 181–194.

Gaur, Vishal, Sridhar Seshadri. 2005. Hedging inventory risk through market instruments. *Manufacturing Service Oper. Management* **7**(2) 103–120.

Glebbeek, Arie C, Erik H Bax. 2004. Is high employee turnover really harmful? an empirical test using company records. *Acad. Management J.* **47**(2) 277–286.

Goyal, Manu, Serguei Netessine. 2011. Volume flexibility, product flexibility, or both: The role of demand correlation and product substitution. *Manufacturing & Service Oper. Management* **13**(2) 180–193.

Grant, Robert M. 1996. Toward a knowledge-based theory of the firm. *Strategic Management J.* **17**(S2) 109–122.

Graves, Stephen C. 1981. A review of production scheduling. *Oper. Res.* **29**(4) 646–675.

Graves, Stephen C. 1988. Safety stocks in manufacturing systems. *J. Manufacturing and Oper. Management* **1**(1) 67–101.

Graves, Stephen C, Sean P Willems. 2000. Optimizing strategic safety stock placement in supply chains. *Manufacturing Service Oper. Management* **2**(1) 68–83.

Holtom, Brooks C, Terence R Mitchell, Thomas W Lee, Marion B Eberly. 2008. Turnover and retention research: A glance at the past, a closer review of the present, and a venture into the future. *Acad. Management Annals* **2**(1) 231–274.

Hossain, Tanjim, John A List. 2012. The behavioralist visits the factory: Increasing productivity using simple framing manipulations. *Management Sci.* **58**(12) 2151–2167.

Hounshell, David A. 1984. *From the American System to Mass Production, 1800-1932: The Development of Manufacturing Technology in the United States*. The Johns Hopkins University Press.

Huselid, Mark A. 1995. The impact of human resource management practices on turnover, productivity, and corporate financial performance. *Acad. Management J.* **38**(3) 635–672.

Ichniowski, Casey, Kathryn Shaw, Giovanna Prennushi. 1997. The effects of human resource management practices on productivity: A study of steel finishing lines. *Amer. Econom. Rev.* **87**(3) 291–313.

Jacoby, Sanford M. 1983. Industrial labor mobility in historical perspective. *Industrial Relations* **22**(2) 261–282.

Jordan, William C, Stephen C Graves. 1995. Principles on the benefits of manufacturing process flexibility. *Management Sci.* **41**(4) 577–594.

Kacmar, Michele K, Martha C Andrews, David L Van Rooy, R Chris Steilberg, Stephen Cerrone. 2006. Sure everyone can be replaced? but at what cost? turnover as a predictor of unit-level performance. *Acad. Management J.* **49**(1) 133–144.

Kesavan, Saravanan, Susan Lambert, Joan C Williams. 2018. The stable scheduling study .

Kim, Sang Won, Marcelo Olivares, Gabriel Y Weintraub. 2014. Measuring the performance of large-scale combinatorial auctions: A structural estimation approach. *Management Sci.* **60**(5) 1180–1201.

Krueger, Alan, Cecilia Rouse. 1998. The effect of workplace education on earnings, turnover, and job performance. *J. Labor Econom.* **16**(1) 61–94.

Lazear, Edward P. 2000. Performance pay and productivity. *Amer. Econom. Rev.* **90**(5) 1346–1361.

Li, Jun, Nelson Granados, Serguei Netessine. 2014. Are consumers strategic? Structural estimation from the air-travel industry. *Management Sci.* **60**(9) 2114–2137.

MacDuffie, John Paul, Kannan Sethuraman, Marshall L Fisher. 1996. Product variety and manufacturing performance: Evidence from the international automotive assembly plant study. *Management Sci.* **42**(3) 350–369.

Macher, Jeffrey T, David C Mowery. 2009. Measuring dynamic capabilities: Practices and performance in semiconductor manufacturing. *British J. Management* **20** S41–S62.

March, James G. 1991. Exploration and exploitation in organizational learning. *Organ. Sci.* **2**(1) 71–87.

Melnicoe, Mark. 2016. China: Companies must shed 'dispatch workers'. *Bloomberg BNA* (Apr. 4).

Mincer, Jacob. 1962. On-the-job training: Costs, returns, and some implications. *J. Political Econom.* **70**(5, Part 2) 50–79.

Moon, Ken, Kostas Bimpikis, Haim Mendelson. 2018. Randomized markdowns and online monitoring. *Management Sci.* **64**(3) 1271–1290.

Moreno, Antonio, Christian Terwiesch. 2015. Pricing and production flexibility: An empirical analysis of the US automotive industry. *Manufacturing Service Oper. Management* **17**(4) 428–444.

Musalem, Andres, Marcelo Olivares, Daniel Yung. 2018. Balancing service speed and agent utilization for an online insurance platform. Presentation at 40th ISMS Marketing Science Conference.

Narayanan, Sriram, Sridhar Balasubramanian, Jayashankar M Swaminathan. 2009. A matter of balance: Specialization, task variety, and individual learning in a software maintenance environment. *Management Sci.* **55**(11) 1861–1876.

Osband, Ian, Daniel Russo, Zheng Wen, Benjamin Van Roy. 2017. Deep exploration via randomized value functions. *arXiv preprint:1703.07608* .

Owen, Laura J. 1995. Worker turnover in the 1920s: What labor-supply arguments don't tell us. *J. Econom. History* **55**(4) 822–841.

Pakes, Ariel, Michael Ostrovsky, Steven Berry. 2007. Simple estimators for the parameters of discrete dynamic games (with entry/exit examples). *RAND J. Econom.* **38**(2) 373–399.

Plevan, Bettina, Aaron J Schindel, Howard Z Robbins, Yasmine Tarasewicz, Anthony J Oncidi, Jeremy M Mittman, Allan H Weitzman, Daniel Ornstein. 2011. Ten frequently asked questions about China's labor law. *Proskauer Rose LLP: International HR Best Practices* (January newsletter).

Raff, Daniel M, Lawrence H Summers. 1987. Did Henry Ford pay efficiency wages? *J. Labor Econom.* **5**(4, Part 2) S57–S86.

Schlichter, Sumner H. 1919. *The Turnover of Factory Labor*. Appleton and Company.

Shaw, Jason D. 2011. Turnover rates and organizational performance: Review, critique, and research agenda. *Organ. Psych. Rev.* **1**(3) 187–213.

Shaw, Jason D, Nina Gupta, John E Delery. 2005. Alternative conceptualizations of the relationship between voluntary turnover and organizational performance. *Acad. Management J.* **48**(1) 50–68.

Shih, Willy C. 2014. What it takes to reshore manufacturing successfully. *MIT Sloan Management Rev.* **56**(1) 55.

Siemsen, Enno, Sridhar Balasubramanian, Aleda V Roth. 2007. Incentives that induce task-related effort, helping, and knowledge sharing in workgroups. *Management Sci.* **53**(10) 1533–1550.

Taneja, Sonia, Mildred Golden Pryor, Leslie A Toombs. 2011. Frederick W. Taylor's scientific management principles: Relevance and validity. *J. Appl. Management and Entrepreneurship* **16**(3) 60.

Tang, Christopher S. 1990. The impact of uncertainty on a production line. *Management Sci.* **36**(12) 1518–1531.

Taylor, Frederick. 1911. *The Principles of Scientific Management*. Harper Brothers.

Tomlin, Brian. 2006. On the value of mitigation and contingency strategies for managing supply chain disruption risks. *Management Sci.* **52**(5) 639–657.

Ton, Zeynep, Robert S Huckman. 2008. Managing the impact of employee turnover on performance: The role of process conformance. *Organ. Sci.* **19**(1) 56–68.

Weintraub, Gabriel Y, C Lanier Benkard, Benjamin Van Roy. 2008. Markov perfect industry dynamics with many firms. *Econometrica* **76**(6) 1375–1411.

Yu, Jeannette. 2014. Labor dispatch in China: The end of an era. *Bloomberg BNA* (May 5).

Yu, Qiuping, Gad Allon, Achal Bassamboo. 2016. How do delay announcements shape customer behavior? An empirical study. *Management Sci.* **63**(1) 1–20.

Electronic Companion

Appendix A: Compensation

A.1. Pay Structure and Disbursal

Pay cycle and worker exit. Workers receive wage-based pay for work completed during monthly pay periods that close on the 25th. However, payday is the subsequent 9th (i.e., two weeks after closing).

This arrangement engenders pay cycle turnover and disincentivizes switching between employers. Consider the worker who leaves the firm on January 10th after collecting a paycheck on the 9th. Typically, she then returns to her home province or begins new employment. On January 9th, she receives her wages for the period ending December 25th. Unless she returns to the facility on or after February 9th, she effectively forgoes her wages for work between December 26th and January 10th. In practice, when confronted by the hurdles of missing work and/or a trip across provinces, virtually all departing workers elect to forgo their final paychecks for these partially completed pay periods. By departing shortly following payday, these workers minimize their forgone earnings and generate the monthly pattern of cyclical turnover that peaks post-payweek. Because leaving an employer entails foregone earnings, workers are disincentivized from frequently switching employers, e.g., from moving between positions for different contract manufacturers. In addition to the fact that the manufacturers draw from a largely homogenous pool of unskilled labor, the friction to switch may blunt wage competition between manufacturers.

Retention bonuses. Retention bonuses are apportioned to employees in the following amounts: (A) 100 RMB awarded to an employee upon completing her first full paymonth (26th to the following calendar month's 25th); (B) 200 RMB awarded upon completing each of her second and third full paymonths; and (C) 300 RMB awarded for each full paymonth worked thereafter. Retention bonuses are altered from time to time, but infrequently and not in the midst of the product life cycle. See Appendix A.2.

Shifts and stipends. Employees earn additional amounts (up to 400 RMB monthly) when assigned to shifts at night (9 RMB per shift) or at uncomfortable positions, such as the clean room. They can earn additional amounts of 50, 100, and 150 RMB by passing operator skill tests at 3, 6, and 9 months into their employment, respectively. Employees receive pro-rated, monthly food stipends of 280 RMB.

Dispatch worker pay. Regarding dispatch hires, Yu (2014) and Melnicoe (2016) explain the significant changes in Chinese laws and regulations that were implemented prior to our study period. Under these, dispatch hires must be compensated with pay and benefits equal to their standard-employee coworkers. Moreover, the rules caused many contract manufacturers, including ours, to convert willing dispatch-sourced workers to regular employment at the end of their dispatch contracts. Otherwise, rules enacted on March 1, 2014 practically precluded further sourcing of labor from dispatch agencies. Because we do not have data on workers' dispatch labor contracts, our analysis of worker turnover in Section 5.2 assumes that dispatch contracts conclude at six months (the reported standard) and convert the worker into a regular employee if she stays.[37]

[37] Reports cite six months as the duration for standard dispatch labor contracts, including for agencies servicing our manufacturer. Six months is the maximum dispatch contract duration permitted under Chinese law (Yu (2014)). However, some agencies may operate under shorter contract durations. At least one public report suggests an agency offering a contract duration of four months with conversion permitted at three months.

A.2. Natural Experiment Suggesting Compensation Effects

Thanks to a reviewer's suggestion, we study how worker tenures responded to a compensation increase for the facility's FATP workers in a natural experiment. More specifically, we compare worker tenures in June and July 2015 against June and July 2013, when workers received lower retention bonuses and base salaries. We obtain and compare the distributions of employment durations for workers exiting the facility within these two time periods. We find statistically significant evidence that workers stayed 46-104% longer at the firm on average when better compensated in 2015. The analysis has limitations, including that we cannot link this employment duration data to the manufacturer's concurrent production data. Consequently, we cannot strongly rule out that other factors (e.g., availability of overtime) may have influenced how long workers chose to stay. We do confirm that gender is balanced across the two periods.

Let us first explain why we chose to compare the 2013 and 2015 time periods. Table EC.1 shows FATP workers' salaries and retention bonuses over 2013-2015 at the studied facility. Let us designate the post-April 2014 period as our benchmark, with compensation for workers (row 3) matching that applicable during our study. During the immediately preceding seven months from August 2013 to March 2014, workers received *higher* retention bonuses but *lower* base salaries (hence lower overtime rates) – compare rows 2 and 3 in the table. Thus, this time period (row 2) fails to provide a clean, unambiguous contrast in compensation against the benchmark. However, during the earlier period before August 2013, workers' base salaries and retention bonuses were *both lower* than the post-April 2014 benchmark – compare rows 1 and 3 in the table.

Table EC.1 Manufacturer's Base Salaries and Retention Bonuses for FATP Workers

		Retention Bonus in Month					
	Monthly Base Salary	1^{st}	2^{nd}	3^{rd}	4^{th}	5^{th}	$6^{th}+$
Before August 2013	1650	100	100	140	140	140	180
August 2013 to March 2014	1650	100	300	300	400	400	500
From April 2014	1820	100	200	200	300	300	300

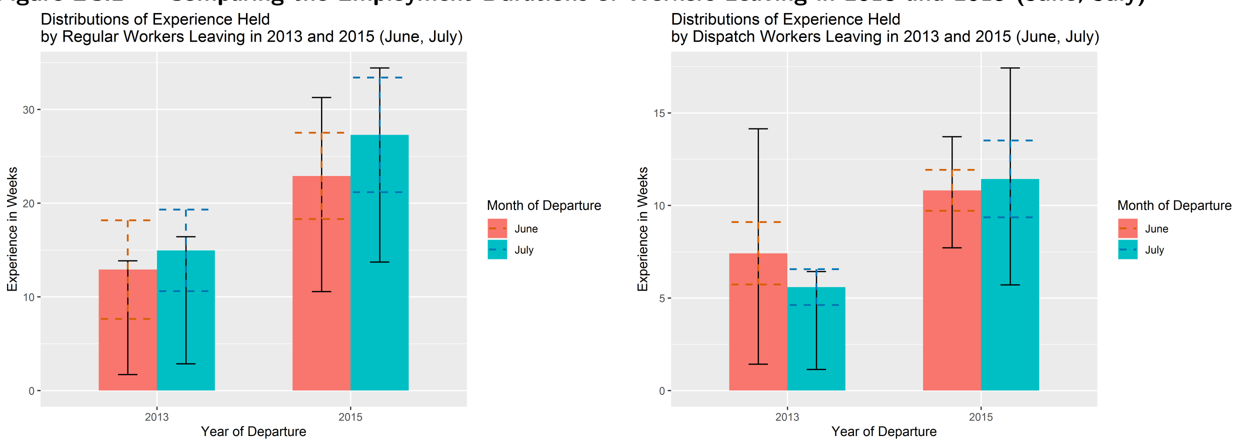
Retention bonuses are awarded for full pay cycles of work completed. Compensation is in RMB.

Our analysis is limited in several respects.

- From prior to August 2013, we obtain data covering June 10th to July 31st, 2013. We were unable to obtain any earlier data. For each employee terminating his or her employment during this time, we observe the duration of his or her employment at the facility.
- The 2013 data cannot identify the roles and assignments in the facility of the employees covered. For example, the data cannot identify whether an employee works in FATP or in subassembly, nor which product she assembles. Thus, the distribution of employment durations covers all workers leaving the facility in June and July, 2013.
- Because the 2013 data cannot link employees to production, we cannot control for factors, such as the availability of overtime, that may influence workers' employment durations.

We obtain employment duration data for all workers leaving the manufacturer's facility in June and July, 2015. This period is appealing as a comparison to June and July of 2013 for two reasons. First, selecting June and July accommodates seasonality in when products are generally introduced and produced in the year. Second, June 2015 is fourteen months after the compensation structure was revised in April 2014. Very few workers stay at the facility longer than fourteen months, so that we can reasonably treat June 2015's employment durations as resulting from the new compensation rules.

Figure EC.2 Comparing the Employment Durations of Workers Leaving in 2013 and 2015 (June, July)



For 2013 and 2015's months of June and July, we plot the distributions of employment durations (measured in weeks) for employees leaving the facility. The solid bars depict the average employment duration, and the dashed, colored error bars show 95% confidence intervals for these sample means. The solid black error bars span the 25th and 75th percentile employment durations among all workers leaving that month. The analysis covers 42,886 workers leaving the facility in 2013 and 2015 (June, July). 17,108 regular employees are used for the left panel, and 25,778 dispatch employees for the right panel. Data for June 2013 covers employees leaving on or after June 10th.

Figure EC.2 compares the distributions of employment durations found in the 2013 and 2015 data. We carry out separate comparisons for regular and dispatch employees. In 2013, regular employees averaged employment durations of 12.9 weeks and 15.0 weeks when exiting the facility in June and July. In contrast, in 2015, they averaged 22.9 weeks when exiting in June and 27.3 weeks in July. (Dashed confidence intervals for these sample averages are shown in Figure EC.2.) Similarly, dispatch employees' employment durations averaged 7.4 and 5.6 weeks when exiting in June and July of 2013, compared to 10.8 and 11.4 weeks in 2015. As Table EC.3 shows, nonparametric, rank-sum tests provide statistically significant evidence (all p-values are below 0.001%) that the 2015 employment durations stochastically dominate the durations from 2013.

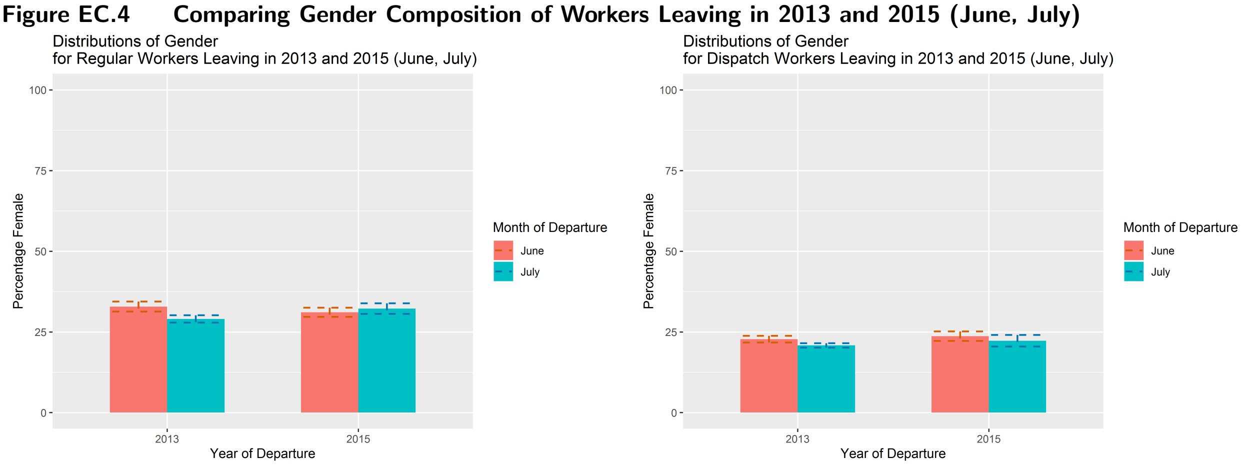
We close out a few minor points regarding the analysis. As shown in Figure EC.4, the facility's stable gender composition indicates that gender cannot plausibly explain the significant expansion in employment durations from 2013 to 2015. The facility also operates at a similar scale in both years. Applying Little's Law to the 2013 and 2015 data's exit counts and employment durations suggests that workforce size is stable between the two years (after proportionally scaling the June 2013 exit count to account for its unobserved days), with a smaller fraction made up of dispatch workers in 2015. Lastly, we cannot rule out that macroeconomic factors may contribute to the longer employment durations in 2015.

Table EC.3 Rank-sum Tests for 2015 Data's Employment Durations Stochastically Dominating 2013

		For 2013 as subsample of 2013 and 2015 pooled		
		Rank sum	Subsample size	P-value
Regular Employees	**June**	9,983,624	3,487 out of 7,728	Below 0.001%
	July	24,345,753	6,247 out of 9,380	Below 0.001%
Dispatch Employees	**June**	25,910,976	6,241 out of 9,452	Below 0.001%
	July	110,039,187	14,257 out of 16,326	Below 0.001%

For each month and employee type, we conduct Mann-Whitney-Wilcoxon rank-sum tests of the null hypothesis that the 2013 and 2015 data's employment durations share a common distribution. Against the one-sided alternative hypothesis that employment durations in the 2015 data stochastically dominate those in the 2013 data, p-values are calculated using 100,000 randomly drawn rank permutations, and the finite sample test is exactly level. Ranks are assigned from lowest to highest, and ties are randomly broken.

Figure EC.4 Comparing Gender Composition of Workers Leaving in 2013 and 2015 (June, July)



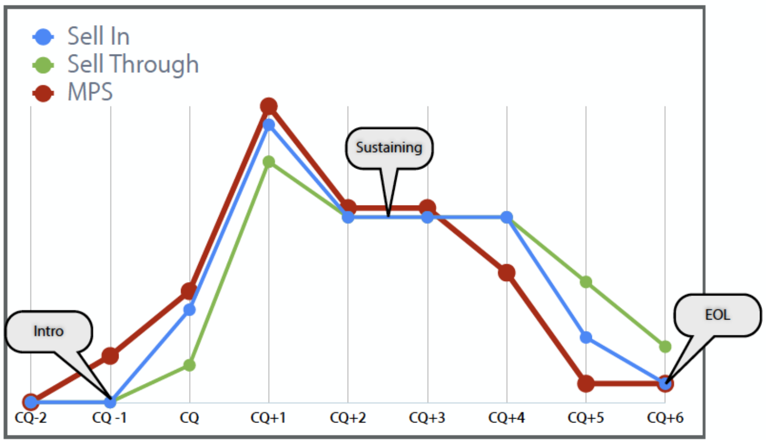
For 2013 and 2015's months of June and July, we plot the distributions of gender for employees leaving the facility. The solid bars depict the percentage female, and the dashed, colored error bars show 95% confidence intervals for these percentages as sample means. The analysis covers 42,886 workers leaving the facility in 2013 and 2015 (June, July). 17,108 regular employees are used for the left panel, and 25,778 dispatch employees for the right panel. Data for June 2013 covers employees leaving on or after June 10th.

Appendix B: Master Production Schedule

Twice weekly in order to coordinate a supply chain involving over 700 suppliers, the firm disseminates the MPS (Master Production Schedule) to its procurement teams and contract manufacturers. The MPS reports weekly order quantities for final assembly, organized into a rolling horizon covering from the immediately pending week to a quarter or more out into the future. Our dataset compiles the primary MPS reports disseminated weekly before each Monday during our study period. See Graves (1981) for an overview of master scheduling.

MPS data. For simplicity, Figure EC.5 depicts a quarterly (instead of weekly) example MPS.

Figure EC.5 Illustration of Master Production Schedule (MPS)



MPS production targets are derived by adjusting the concurrently available sell-through data and forecasts to build up, sustain, and wind down safety-stock inventory, in light of robust product margins. Real-time capacity and supply constraints are separately accounted for by the report's clear-to-build quantities.

At a high level, the MPS is constructed as follows. First, approximately one year in advance of producing the device, the firm sets its planned capacity for each week of production. Once production commences, the firm begins to collect and receive data regarding the product's rates of both sell-in (orders made into distribution channels) and sell-through (sales to end customers in retail). Each week, the latest data is used to generate demand forecasts for the individual weeks in the planning horizon. Applying the planned production capacities to the latest forecasts, the MPS report's weekly production quantities are generated, as shown quarterly in Figure EC.5.

Because of the product's sizeable margins, it is appropriate for the MPS to pace production to build up and wind down retail safety stocks over the product's life cycle. (See, e.g., Gaur et al. (2005).) During the initial months of the life cycle, the rate of producing the product should exceed sell-through, resulting in the accumulation of inventories. In the intermediate phase, inventories are sustained by matching the pace of production with that of sell-through. Finally, inventories are wound down for the product life cycle's end. These dynamics are depicted in Figure EC.5 and built into the MPS quantities.

By design, the MPS omits supply-side factors and constraints. For example, while the MPS *reports* include information on the incoming supplies of components, the MPS itself does not reflect, for example, when the firm expects a component shortage to bindingly constrain weekly production. Instead, a parallel sequence of weekly CTB (clear-to-build) quantities is reported alongside the MPS. For each week, the CTB projects the production quantities that are actually feasible after considering supply-side (i.e., procurement and workforce capacity) constraints and shocks. Thus, the MPS represents ideal production quantities generated from the latest demand forecasts while assuming the pre-planned production capacity, whereas the CTB further accounts for the latest news about available supply and capacity.

MPS variables. We define the MPS-based variables used in the study. The schedule generated in week t, denoted MPS_t, is the vector of order quantities:

$$\mathrm{MPS}_t := \{\mathrm{MPS}_{t,0}, \ldots, \mathrm{MPS}_{t,H}\}, \tag{EC.1}$$

where $\mathrm{MPS}_{t,0}$ is the quantity demanded for week t, $\mathrm{MPS}_{t,1}$ is the same for week $t+1$; and so forth through the planning horizon of H weeks in the future. In the original data, these quantities are cumulative for the ideal total quantity shipped by the target week's close. We thus apply the lag-differencing operator, $1 - L_h$, to convert these into weekly planned production quantities. (For example, if the current week's cumulative MPS quantity is 3 million and was 2 million last week, then lag-differencing yields 1 million as the quantity to produce this week. The subscript on lag L_h makes clear that the differencing spans consecutive weeks within a single MPS report, not differences between MPS reports consecutively released.)

Similarly, week t's report communicates the sequence of clear-to-build (CTB) quantities:

$$\mathrm{CTB}_t := \{\mathrm{CTB}_{t,0}, \ldots, \mathrm{CTB}_{t,H}\}. \tag{EC.2}$$

From the contract manufacturer's perspective, the CTB are predicted actual order quantities.

From the report in each week t, the first triplet of variables captures the trend of "ramping" (up or down) in MPS quantities. By log-differencing between weeks in the horizon, we measure the percentage growth in the planned production between the weeks. For instance, the first ramping variable below captures the growth in week t's scheduled production over the previous week's. The second and third variables measure growth in scheduled production one and two weeks out, respectively, against the current week's.

$$\begin{bmatrix} \mathrm{MPSRamp}_{t,0} \\ \mathrm{MPSRamp}_{t,1} \\ \mathrm{MPSRamp}_{t,2} \end{bmatrix} := \begin{bmatrix} \log\big((1-L_h)\cdot \mathrm{MPS}_{t,0}\big) \\ \log\big((1-L_h)\cdot \mathrm{MPS}_{t,1}\big) \\ \log\big((1-L_h)\cdot \mathrm{MPS}_{t,2}\big) \end{bmatrix} - \begin{bmatrix} \log\big((1-L_h)\cdot \mathrm{MPS}_{t,-1}\big) \\ \log\big((1-L_h)\cdot \mathrm{MPS}_{t,0}\big) \\ \log\big((1-L_h)\cdot \mathrm{MPS}_{t,0}\big) \end{bmatrix}. \tag{EC.3}$$

An additional "surprise" variable focuses on how the current week's scheduled production is *revised by* the latest report's new information:

$$\mathrm{MPSInfo}_t := \log\big((1-L_H)\cdot \mathrm{MPS}_{t,0}\big) - \log\big((1-L_H)\cdot \mathrm{MPS}_{t-1,1}\big). \tag{EC.4}$$

Note that the production scheduled for the current week, $\mathrm{MPS}_{t,0}$, is compared against last week's forecast for the same week, $\mathrm{MPS}_{t-1,1}$ (i.e., the 1-week-ahead forecast from $t-1$'s report). After lag-differencing to obtain the scheduled weekly production quantities involved, log-differencing derives the percentage change.

Appendix C: First-stage Regressions & Tests for Instrumental Variables

Tables EC.6 and EC.7 provide the first-stage regressions for the yield analyses of Tables 9 and 12. For convenience, we have highlighted in boldface the significant coefficients that support the relevance of the instrumental variables. For example, in Table EC.6, other buildings' turnover rates significantly predict assembly lines' attrition rates, whereas the MPS ramping variables significantly predict slow-downs in assembly lines' speeds. The economic background behind these relationships is explained by the main text.

Each table shows the associated first-stage F-statistics. We summarize the test results of both the first stage and the main text. First, all F-statistic tests for weak instruments result in p-values below 0.001, which provides evidence supporting that the instruments are not weak. As shown in Tables 9a and 12, all Sargan tests fail to detect statistical evidence of violations of the exclusion restrictions. All Wu-Hausman tests for endogeneity are significant at p-values less than 0.001, signaling significant endogeneity addressed by the instrumental variables.

Finally, Table EC.8 shows the first-stage regressions supporting Table 14's first-pass yield analyses. In addition to the instrumental variables used to study yields, the first-pass yield analyses rely on lagged line turnover predicting workloads and on "surprising" revisions to the MPS predicting the prevalence of staffing new hires. As in the prior tables, all F-statistic tests support that the instruments are not weak, and the Sargan tests in Table 14a do not find evidence of violating the exclusion restrictions.

As the main text discusses, the instruments appear to resolve the expected endogeneity biases in the correct directions. Reflecting how managers endogenously allocate work to the best performing lines, the instrumental variables dissipate or reverse the positive associations of line productivity with production speeds and workloads. (Such positive associations are difficult to explain without such endogeneity.)

Table EC.6 IV First-stage Regressions for Table 9a

	Dependent Variable	
	Attrition Rate	$\log(\text{UPH}_{it})$
Other buildings' mean turnover rate	**0.807*** **	0.047
	(0.084)	(0.173)
Current-week ramping of MPS by log-differencing	−0.004	**−0.065** **
	(0.010)	**(0.020)**
1-week-ahead ramping of MPS by log-differencing	0.021	**−0.060* **
	(0.014)	**(0.029)**
2-weeks-ahead ramping of MPS by log-differencing	−0.014	**−0.087*** **
	(0.011)	**(0.023)**
R^2	0.07	0.03
F-statistic	26.5	10.3

Significance levels → *** - 0.001 ** - 0.01 * - 0.05

Unbalanced panel for 1,516 production-active line-weeks in Sept. 2014 - Jun. 2015. Regressions include fixed effects for forty-four assembly lines observed.

Table EC.7 IV First-stage Regressions for Table 12

	Dependent Variable		
	Attrition Rate	log(UPH$_{it}$)	Net Attrition Rate
Other buildings' mean turnover rate	**0.785*****	0.022	0.907
	(0.084)	(0.173)	(1.182)
Lagged assembly line turnover rate	0.156**	0.170	**1.954***
	(0.060)	(0.124)	**(0.845)**
Current-week ramping of MPS by log-differencing	−0.003	**−0.064****	−0.406**
	(0.010)	**(0.020)**	(0.137)
1-week-ahead ramping of MPS by log-differencing	0.016	**−0.066***	−0.389
	(0.014)	**(0.030)**	(0.201)
2-weeks-ahead ramping of MPS by log-differencing	−0.007	**−0.080*****	0.078
	(0.011)	**(0.023)**	(0.159)
R^2	0.07	0.03	0.01
F-statistic	22.6	8.6	3.4

Significance levels → *** - 0.001 ** - 0.01 * - 0.05

Unbalanced panel for 1,516 production-active line-weeks in Sept. 2014 - Jun. 2015. Regressions include fixed effects for forty-four assembly lines observed.

Table EC.8 IV First-stage Regressions for Table 14a

	Dependent Variable			
	Attrition Rate	New Hire Headcount %	log(L_{it})	log(UPH$_{it}$)
Other buildings' mean turnover rate	**0.785*****	0.487***	−1.143*	0.021
	(0.084)	(0.048)	(0.582)	(0.174)
Lagged assembly line turnover rate	0.160**	0.087*	**3.002*****	0.151
	(0.061)	(0.034)	**(0.419)**	(0.125)
Current-week ramping of MPS by log-differencing	−0.002	0.090***	0.215**	**−0.065****
	(0.010)	(0.006)	(0.067)	**(0.020)**
1-week-ahead ramping of MPS by log-differencing	0.017	−0.006	−0.013	**−0.069***
	(0.014)	(0.008)	(0.100)	**(0.030)**
2-weeks-ahead ramping of MPS by log-differencing	−0.008	0.065***	0.230**	**−0.079*****
	(0.011)	(0.006)	(0.079)	**(0.023)**
MPS revision by log-differencing	0.050	**−0.322*****	0.506	−0.214
	(0.084)	**(0.048)**	(0.580)	(0.173)
R^2	0.07	0.24	0.04	0.03
F-statistic	18.9	76.5	10.5	7.4

Significance levels → *** - 0.001 ** - 0.01 * - 0.05

Unbalanced panel for 1,516 production-active line-weeks in Sept. 2014 - Jun. 2015. Regressions include fixed effects for forty-four assembly lines observed.

Appendix D: Robustness Checks

D.1. Regression Analyses Excluding Building C

We re-run the analyses of Tables 9a, 12, and 14a after excluding Building C's assembly lines. We apply the same instruments for each analysis but generally lose some statistical power by excluding 14 of 44 lines.

Table EC.9 Workgroup Yield and First-pass Yield Analyses excluding Building C

(a) Log Yield Regression Results with IV (Tables 9a and 12)

Workgroup Regressors	(1)	(2)	(OLS)	(3)	(4)	(OLS)
Attrition rate	-0.068^{***}	-0.062^{***}	0.001	-0.066	-0.068^{*}	-0.001
	(0.016)	(0.016)	(0.003)	(0.108)	(0.026)	(0.003)
Net attrition rate				-0.019	-0.001	0.000
				(0.029)	(0.005)	(0.000)
$\log(\text{UPH}_{lt})$	-0.098	0.020	0.099^{***}	-4.637	-0.381	0.099^{***}
	(0.787)	(0.729)	(0.023)	(19.359)	(0.548)	(0.023)
$\log(\text{UPH}_{lt})^2$	0.015	0.003	-0.006^{**}	0.445	0.042	-0.006^{**}
	(0.075)	(0.069)	(0.002)	(1.834)	(0.051)	(0.002)
Dummies for weeks to payweek		Y	Y		Y	Y
R^2	0.22	0.28	0.38	0.15	0.13	0.38
Wu-Hausman p-value	<0.001	<0.001		<0.001	<0.001	
Sargan p-value	0.529	0.410		0.995	0.177	
Significance levels → *** - 0.001	** - 0.01	* - 0.05				

Unbalanced panel of 1,161 production-active line-weeks in Sept. 2014 – Jun. 2015. Fixed effects for 30 assembly lines. IV regressions (1) and (2) use the same instrumental variables as Table 9, and (3) and (4) use the same instrumental variables as Table 12. Weak-instrument F-test p-values all fall below 0.001 to reject null.

(b) Log First-pass Yield Regression Results with IV (Table 14a)

Workgroup Regressors	(1)	(2)	(OLS)
New hire staffing %	-2.171^{***}	-2.772^{**}	-0.921^{***}
	(0.595)	(0.862)	(0.061)
$\log(\text{UPH}_{lt})$	10.942	26.192	1.380^{***}
	(19.715)	(30.087)	(0.299)
$\log(\text{UPH}_{lt})^2$	-1.097	-2.543	-0.103^{***}
	(1.888)	(2.888)	(0.030)
$\log(L_{lt})$	-0.212^{*}	-0.232	0.023^{***}
	(0.091)	(0.157)	(0.006)
Dummies for weeks to payweek		Y	Y
R^2	0.08	0.15	0.40
Wu-Hausman p-value	<0.001	<0.001	
Sargan p-value	0.529	0.353	
Significance levels → *** - 0.001	** - 0.01	* - 0.05	

Unbalanced panel of 1,161 production-active line-weeks in Sept. 2014 – Jun. 2015. Fixed effects for 30 assembly lines. IV regressions use the same instrumental variables as Table 14. Weak-instrument F-test p-values all fall below 0.001 to reject null.

D.2. First-pass Yields Excluding Failed Units

We replicate the findings in Table 14 for first-pass yields, using an alternative version that excludes failed units from station inputs:

$$\text{First-pass Yield Excluding Fails}_{lt} := \prod_{s\in S} \frac{\#\ \text{Pass}_{lts} - \#\ \text{Retested}_{lts}}{\#\ \text{Pass}_{lts}}. \tag{EC.5}$$

The log version of this metric equals the log of first-pass yield minus the log of yield, where summary statistics (e.g., Table 2) indicate that the former is approximately 3 times more variable than the latter. The metric descriptively resembles first-pass yields, with sample mean, median, and standard deviation being 33.8%, 34.3%, and 5.6%, respectively, in comparison to first-pass yield's 31.7%, 32.1%, and 5.5% (Table 2).

The estimation results shown in Table EC.10 virtually replicate Table 14a.

Table EC.10 First-pass Yield Analysis (Table 14a) excluding Failed Units

Workgroup Regressors	(1)	(2)	(OLS)	(3)	(4)	(OLS)
New hire staffing %	-2.968^{***}	-2.625^{***}	-0.885^{***}	-3.016^{***}	-2.975^{***}	-0.872^{***}
	(0.299)	(0.439)	(0.054)	(0.301)	(0.428)	(0.053)
Attrition rate		-0.303	-0.025		0.039	-0.023
		(0.300)	(0.032)		(0.293)	(0.032)
$\log(\text{UPH}_{lt})$	7.320^{*}	5.410	0.885^{***}	8.664^{*}	8.491^{*}	0.895^{***}
	(3.605)	(3.854)	(0.232)	(3.502)	(3.701)	(0.230)
$\log(\text{UPH}_{lt})^2$	-0.743^{*}	-0.552	-0.059^{*}	-0.865^{*}	-0.847^{*}	-0.060^{*}
	(0.350)	(0.377)	(0.024)	(0.342)	(0.365)	(0.024)
$\log(L_{lt})$	-0.101^{*}	-0.096^{*}	0.010^{*}	-0.106^{*}	-0.105^{*}	0.011^{*}
	(0.043)	(0.040)	(0.005)	(0.043)	(0.043)	(0.005)
Dummies for weeks to payweek				Y	Y	Y
R^2	0.16	0.15	0.38	0.17	0.17	0.39
Wu-Hausman p-value	<0.001	<0.001		<0.001	<0.001	
Sargan p-value	0.634	1		0.444	0.199	
Significance levels → *** - 0.001	** - 0.01	* - 0.05				

Unbalanced panel of 1,516 production-active line-weeks in Sept. 2014 – Jun. 2015. Fixed effects for 44 assembly lines. IV regressions use instrumental variables: lagged own-line turnover, average line turnover in other two buildings, and MPS-based variables. Weak-instrument F-test p-values all below 0.001 to reject null. The mean and median first-pass yields excluding failed units are 33.8% and 34.3%, with a standard deviation of 5.6%.

Appendix E: Equilibrium & Estimation

E.1. Model Overview

By estimating our structural model of worker turnover, we effectively infer workers' preferences from when they choose to turnover in equilibrium.

Parameters. The model identifies preference parameters for each worker type. Organized into the vector θ_i, the parameters act as coefficients describing worker i's preferences. Encountering workplace state Z, she receives a weekly utility flow given by the product $Z^T\theta_i$. Outside of compensation paid and received, her workplace state Z consists of her weekday and weekend hours worked and her assembly line's workrate and attrition rate.

Worker i draws her weekly, private exit value ϕ_{it} from the distribution, $\Phi(\cdot; h_t, \theta_i)$. The exit value distribution depends on the worker's type (dispatch or direct) and on whether workers are returning from a holiday in week t. Especially for dispatch workers, exiting post-holiday avoids substantial transportation costs back to the factory and home again. We specify Φ as exponential with mean, $\lambda(h_t, \theta_i) = \sigma + \lambda \cdot h_t$: indicator $h_t = 1$ when returning from a holiday, while σ represents the baseline expected value of exit.[38] Φ's parameters, σ and λ, are distinct for dispatch and direct hires and included in θ_i. Lastly, we allow dispatch workers to incur a one-time inconvenience cost to continue as regular hire at the end of their dispatch contracts, which we assume uniformly at 6 months.

States. Conventional dynamic equilibrium models, notably the Markov-perfect equilibrium, require workers to track precisely the payoff-relevant workplace states of other workers, including the experience levels and accrued earnings of several hundred coworkers on her assembly line. In our setting, this is both implausible and intractable.

Instead, we specify the information, X_{it}, that worker i is able to use.[39] She recalls or tracks:

- *Previous workplace state.* Her own experience level and her line's workplace state, $Z_{i,t-1}$, yield, first-pass yield, and share of new hires in the previous week;
- *Pay cycle.* Weeks to payweek and any fractional attribution of week t's days across pay periods;
- *Compensation.* Her earned incomes due on the next and subsequent paydays and her base rate of hourly pay, reflecting her assigned shift – see Appendix A.1 on pay accrual;
- *Production trends.* (A) the Master Production Schedule (MPS)'s week-on-week ramping or sustaining trends, (B) unexpected corrections to MPS (MPSInfo) and (C) component supply shortages (CTB, expressed as fraction of MPS). See Appendix B for details on the MPS-sourced variables, $\text{MPSRamp}_{t,0}$, $\text{MPSRamp}_{t,1}$, MPSInfo_t, and CTB.

[38] We impose that $\min\{\sigma + \lambda \cdot h : h = 0, 1\} > 0$, necessarily implying $\sigma > 0$.

[39] That players track informationally relevant states contrasts with recent equilibrium concepts involving many players. In these, a stable overall distribution of players' types permits them to condition their strategies solely on their own types – and the population's asymptotic type distribution – with negligible loss of individual optimality. See, e.g, Weintraub et al. (2008), Adlakha et al. (2015), and Balseiro et al. (2015).

E.2. Estimation Overview

The estimation is carried out in two steps. First, we estimate equilibrium transition probabilities for the workplace state, or more precisely the state of the worker's available information which includes her workplace state. As captured by the model's Bellman equations (8), the worker's value to stay with the firm is a function jointly of her expectation over future states and of her preferences over states. With the expected states pinned down by estimated transition probabilities, the second stage estimates workers' preferences based on when they turnover. The two-step estimator's confidence intervals derive by bootstrap.

E.3. First-stage Estimation of State Transitions

The first stage nonparametrically estimates two equilibrium functions. First, we estimate workers' state-contingent turnover probabilities, $\overrightarrow{\rho}: \mathbb{X} \to [0,1]$, using a non-parametric, sieve-logit specification. That is, our logit regression predicts workers' binary turnover decisions using polynomial bases, expanding appropriately with the sample size, of the information set, $X_{it} \in \mathbb{X}$. Second, we estimate the state transition process or matrix, $M_C: \mathbb{X} \times \mathbb{X} \to [0,1]$, capturing how the worker who stays transitions through workplace states in equilibrium.

To estimate M_C, first take the data observations of worker-weeks in which the worker chose not to exit the firm. State transitions are consistently estimated on this subsample. See Pakes et al. (2007).

The worker will now either remain on the same assembly line, be transferred to another assembly line, or be transferred to the next generation product's assembly, and the likelihood of each outcome depends on the current state, X_{it}. On the subsample, estimate by sieve logit the agent's state-dependent probability (i.e., conditioned on states X_{it}) of being transferred to join another assembly line.

Then take the data's further subsample of worker-weeks in which the worker neither exited nor was transferred to another assembly line. On this subsample, estimate by sieve logit on X_{it} the agent's state-dependent probability of being transferred to the next generation product.

Finally, for each staffing outcome, we estimate the new workplace state. First collecting all observations where the agent remains on her assembly line, the following transition process is estimated for week t:

$$\begin{bmatrix} \text{Weekly Salary Earned}_{it} \\ \text{Line Hours}_{l(i,t),t} \\ \text{Weekend Utilization}_{l(i,t),t} \\ \text{UPH}_{l(i,t),t} \\ \text{Yield}_{l(i,t),t} \\ \text{First-pass Yield}_{l(i,t),t} \\ \text{Attrition}_{l(i,t),t} \\ \text{Newly Hired Share}_{l(i,t),t} \end{bmatrix} = M_{\text{SameLine}} \begin{bmatrix} \text{Hourly Pay Rate}_{it} \\ \text{Line Hours}_{l(i,t-1),t-1} \\ \text{Weekend Utilization}_{l(i,t-1),t-1} \\ \text{UPH}_{l(i,t-1),t-1} \\ \text{Yield}_{l(i,t-1),t-1} \\ \text{First-pass Yield}_{l(i,t-1),t-1} \\ \text{Attrition}_{l(i,t-1),t-1} \\ \text{Newly Hired Share}_{l(i,t-1),t-1} \\ \text{CTBShortage}_{t} \\ \text{MPSRamp}_{t,0} \\ \text{MPSRamp}_{t,1} \\ \text{MPSInfo}_{t} \end{bmatrix} + \overrightarrow{\epsilon}_t^{\text{SameLine}}, \qquad \text{(EC.6)}$$

where $l(i,t)$ represents agent i's assigned line in week t. The mean-zero vector of transition shocks, $\overrightarrow{\epsilon}_t^{\text{SameLine}}$, is assumed independent across weeks t, with an unrestricted covariance structure, Ω_{SameLine}, estimated for its elements within-week. For the case of transfer onto a new line, a transition process, M_{Transfer} and Ω_{Transfer}, is similarly estimated. Exogenous state transitions for the hourly wage and MPS are provided below. Note that workers' attributes (as part of X_{it}) can affect their likelihoods of being transferred, but are not modeled as affecting the assembly line's state conditional on being transferred. Week t's earned salary bases on a worker's hourly compensation rate, her assembly line's workload hours, and multiples-based adjustments for overtime hours. The above specification parsimoniously admits these relationships alongside important correlations with the workplace state and MPS.

Workers' hourly base compensation rates transition by the following process:

$$\text{Hourly Pay Rate}_{i,t+1} = \alpha_{\text{PayRate}} \cdot \text{Hourly Pay Rate}_{it} + \vec{\epsilon}^{\,\text{PayRate}}_{t+1}, \tag{EC.7}$$

which allows flexibility in shift type. We estimate the coefficient α_{PayRate} and the standard deviation σ_{PayRate}. We separately account for monthly bonuses when constructing i's accrued pay.

Lastly, we estimate the MPS transition as the exogeneous, vector-autoregressive stochastic process:

$$\begin{bmatrix} \text{CTBShortage}_{t+1} \\ \text{MPSRamp}_{t+1,0} \\ \text{MPSRamp}_{t+1,1} \\ \text{MPSInfo}_{t+1} \end{bmatrix} = M_{\text{MPS}} \begin{bmatrix} \text{CTBShortage}_{t} \\ \text{MPSRamp}_{t,0} \\ \text{MPSRamp}_{t,1} \\ \text{MPSInfo}_{t} \end{bmatrix} + \vec{\epsilon}^{\,\text{MPS}}_{t+1}, \tag{EC.8}$$

where CTBShortage_t represents the fractional shortage in weekly feasible production (due to procurement and other supply constraints) relative to week t's demand-based MPS target, as reported at week t's start:

$$\text{CTBShortage}_t := \frac{\max(\text{MPS}_{t,0} - \text{CTB}_{t,0}, 0)}{(1 - L_H) \cdot \text{MPS}_{t,0}}.$$

The covariance matrix, Ω_{MPS}, of the mean-zero, transition-shock vector, $\vec{\epsilon}^{\,\text{MPS}}_{t+1}$, assumed independent across weeks, is estimated along with matrix M_{MPS}.

E.4. Procedure to Estimate Workers' Preferences

To estimate worker preferences, we adapt Bajari et al. (2007)'s simulation-based, two-step estimator to the restricted EBE. This minimum-distance estimator exploits the idea that we observe workers' exit policies, $\vec{\rho}$, which should be optimal hence outperform alternative policies when their expected rewards are computed and compared by simulation.

The estimation has two critical steps. The first is estimating workers' turnover probabilities, $\hat{\rho}$, in each state. This is done by estimating $\hat{\rho}$ as a nonparametric function over the state space $\mathbb{X}$. Next, the procedure reverse engineers from each state X's exit probability, $\hat{\rho}(X)$, the workers' state-dependent exit threshold, i.e., the smallest exit value ϕ for which workers would leave. These thresholds fully describe the policy workers follow for deciding on staying or exiting. Thus, when workers' utility-based rewards are specified, we can simulate the policy's rewards by drawing exit values and state transitions and following the exit rule (i.e., exit if the drawn exit value exceeds the current state's threshold).

The remainder of the procedure carries out parallel simulation runs, under the recovered policy and under alternative policies generated by perturbing the obtained thresholds. Through 200 million simulations covering 100,000 perturbed polices, we construct an estimator based on the idea that the observed policy should outperform the perturbed policies in expectation under workers' true preferences. Intuitively, the estimated preferences are chosen to minimize violations where the observed policy is outperformed by one of the 100,000 perturbed policies. Our implementation is as follows.

1. *Recover and normalize exit thresholds from the data.* First, invert the estimated turnover probabilities $\hat{\rho}$ to obtain the exit thresholds presently followed by workers.

   Incentive-compatibility constraint (9) dictates that worker i exits if and only if her exit value ϕ_{it} in week t exceeds her expected value to stay. Without loss of generality, we normalize her exit value by the state-dependent mean of its distribution, resulting in an equivalent threshold rule. Under this rule, the worker exits if and only if her normalized exit value $\xi_{it} := \phi_{it}/\lambda_{\theta_i}(X_{it})$ exceeds the re-scaled threshold T:

   $$\text{Worker } i \text{ exits in week } t \iff T(X_{it}) := \frac{VC_{\theta_i}(X_{it})}{\lambda_{\theta_i}(X_{it})} < \xi_{it}, \tag{EC.9}$$

   where the normalized exit values ξ_{it} are now standard exponential. Because worker characteristics and assembly line conditions are incorporated into state X_{it}, T can be thought of as custom-tailoring the exit thresholds to workers and conditions so as to standardize the exit values.

   Worker i's probability of exit in week t is the probability of the event $T(X_{it}) < \xi_{it}$. We already possess first-stage estimates, $\hat{\rho}(X_{it})$, of the probabilities of exit. By inverting the exponential CDF for ξ_{it}, we recover the set of estimated exit thresholds:

   $$\hat{T}(X) = -\log(\hat{\rho}(X)) \text{ for all } X \in \mathbb{X}. \tag{EC.10}$$

   In the subsequent simulation steps, $\hat{T}(X)$ is sufficient for evaluating the policy's payoffs. In each encountered state X, a simulated exit value ξ is compared against $\hat{T}(X)$ to decide whether the worker exits, and the resulting rewards are collected and recorded for each simulation run.

2. *Generate alternative exit policies by perturbation.* Next, fix a set of B alternative policies, each represented by their own exit thresholds $T_b^{'} : \mathbb{X} \to \mathbb{R}_+$ for $b = 1, \ldots, B$. We generate our set of alternative policies by: for each b, randomly perturb the estimated coefficients of $\hat{\rho}$, and then apply Step 1's inversion procedure to obtain the altered threshold rules $T_b^{'} \neq \hat{T}$. We generate and fix a distinct counterfactual exit policy $T_b^{'}$ for each of the 100,000 initial state draws described below, hence $B = 100,000$.

3. *Simulation.* Draw 100,000 states, $\{X_{(k)} \in \mathbb{X},\ k = 1, \ldots, 100,000\}$, with replacement from the dataset's empirical distribution of states, $\{X_{it} \in \mathbb{X}\}$. For each of these draws, we simulate 1,000 sample paths under each of the observed and counterfactual exit policies, for 200 million sample paths simulated in total. For each draw k, the intent is to use these simulations to evaluate:

   $$V_{\hat{T}}(X_{(k)}; \hat{M}_C, \theta) - V_{T_k^{'}}(X_{(k)}; \hat{M}_C, \theta), \tag{EC.11}$$

   where V_T represents the worker's expected, present-valued payoffs when she follows the threshold exit policy T (after starting at state $X_{(k)}$ and under the state-transition beliefs $\hat{M}_C$). More precisely:

   $$V_T(X; \hat{M}_C, \theta) := \mathbb{E}_{T, \hat{M}_C}\left[\sum_{t=0}^{\tau-1} \delta^t \cdot U(X_{i,t+1}; \theta_i) + \delta^\tau \cdot \lambda_{\theta_i}(X_{i\tau}) \cdot \xi_{i\tau} \mid X_{i,0} = X\right], \tag{EC.12}$$

   where X_{it} transitions by $\hat{M}_C$, and τ is the random exit time under threshold exit policy T:

   $$\tau := \inf\{t \in \mathbb{Z}_+ :\ \xi_{it} > T(X_{it})\}. \tag{EC.13}$$

Let us define $D_k(\theta)$ as the difference expressed in (EC.11). Note that $D_k(\theta)$ is a function of θ: assuming the optimal policy is unique, optimality and first-stage consistency together imply that $D_k(\theta) > 0$ holds in probability as the sample size increases.

4. *Simulation output.* For each of the k sampled instances, construct the following simulated estimate of $D_k(\theta)$, which we denote by $\hat{D}_k(\theta)$. For each of $r = 1, \dots, R$ simulation runs, we collect the outcomes:

$$S_{\hat{T},k,r} := \left\{ X_{it}^{(k,r)}, t = 1, \dots, \tau_{(k,r)}, \xi_{i\tau_{(k,r)}}^{(k,r)}, \tau_{(k,r)} \right\} \text{ and } S_{T_k',k,r} := \left\{ X_{it}^{'(k,r)}, t = 1, \dots, \tau_{(k,r)}^{'}, \xi_{i\tau_{(k,r)}^{'}}^{'(k,r)}, \tau_{(k,r)}^{'} \right\}, \tag{EC.14}$$

where all simulations are initialized with $X_{i,0}^{(k,r)} = X_{i,0}^{'(k,r)} = X_{(k)}$, $\hat{M}_C$ governs the state transitions of $X_{it}^{(k,r)}$ and $X_{it}^{'(k,r)}$, $\xi_{it}^{(k,r)}$ and $\xi_{it}^{'(k,r)}$ are i.i.d. standard exponential, and $\tau_{(k,r)}$ and $\tau_{(k,r)}^{'}$ are realized exit times under the threshold policies $\hat{T}$ and $T_k^{'}$, respectively. We set $R = 1000$, and each sample path simulation is permitted to proceed up to $\bar{H} = 260$ weeks (five years).

Exploiting that weekly payoffs and exit values are linear in the parameters, let:

$$\hat{D}_k(\theta) := \theta^T \left[\sum_{r=1}^{R} \sum_{t=0}^{\bar{H}} \delta^t \left(Y_t^{(k,r)} - Y_t^{'(k,r)} \right) \right], \tag{EC.15}$$

where the sums are element-wise over the vectors Y given by:

$$Y_t^{(k,r)} := \begin{bmatrix} X_{it}^{(k,r)} \cdot \mathbf{1}_{\left\{t < \tau_{(k,r)}\right\}} \\ \xi_{it}^{(k,r)} \cdot \mathbf{1}_{\left\{t = \tau_{(k,r)}\right\}} \\ \xi_{it}^{(k,r)} \cdot \mathbf{1}_{\left\{t = \tau_{(k,r)}\right\}} \end{bmatrix} \text{ and } Y_t^{'(k,r)} := \begin{bmatrix} X_{it}^{'(k,r)} \cdot \mathbf{1}_{\left\{t < \tau_{(k,r)}^{'}\right\}} \\ \xi_{it}^{'(k,r)} \cdot \mathbf{1}_{\left\{t = \tau_{(k,r)}^{'}\right\}} \\ \xi_{it}^{'(k,r)} \cdot \mathbf{1}_{\left\{t = \tau_{(k,r)}^{'}\right\}} \end{bmatrix}. \tag{EC.16}$$

By $\theta^T X$ for $X \in \mathbb{X}$, we mean the dot product of parameters against X's payoff relevant subvector, and the replicated pair of normalized exit values ξ is multiplied against the parameters σ and λ, respectively.

5. *Estimator.* Finally, obtain the minimum-distance estimator:

$$\hat{\theta} := \operatorname{argmin}_{\theta \in \Theta} \sum_k \min\left(\hat{D}_k(\theta), 0 \right)^2. \tag{EC.17}$$

In principle, $\hat{D}_k(\theta)$ needs to be re-evaluated repeatedly over different $\theta \in \Theta$ to find the minimizer, $\hat{\theta}$. However, the simulated component of Steps 3 and 4 need only be carried out a single time for the entire estimation procedure. Upon completing the simulations, calculate and fix the matrix containing as its kth column the square-bracketed term in (EC.15) as a pre-optimization step. During (EC.17)'s optimization over $\theta \in \Theta$, left-multiplying the pre-fixed matrix by a given θ's transpose immediately gives the row vector, $\left\{ \hat{D}_k(\theta) : k = 1, \dots, B \right\}$. Additionally, Step 3's simulations can be parallelized.

E.5. Alternative Structural Estimation of Worker Preferences

We provide an alternative method for estimating worker preferences. The previous section's simulation-based estimator empirically recovered the workers' exit policy and simulated its rewards, without ever solving directly for an optimal policy. In contrast, the alternative method presented here solves the worker's optimal stopping problem. Whereas the previous estimator's computation burden arose from its extensive simulations, the methodological and computational challenge for the alternative method stems from obtaining policy solutions in an extremely large state space where the state transition matrix, $\hat{M}_C$, exceeds 1TB even when sparsely represented via sampling. Worker preferences are estimated by maximizing the log likelihood from solving the workers' optimal stopping problem, where the problem and its solution vary with workers' preference-based reward functions.

Our estimator solves the following constrained optimization problem, with the log likelihood as objective and indicator variable d_{it} signaling worker i's deciding to depart in week t:

$$\hat{\theta}_{MLE} := \text{argmax}_{\theta\in\Theta} \sum_{i,t} \left[d_{it}\cdot\log(\rho(X_{it})) + (1-d_{it})\cdot\log(1-\rho(X_{it}))\right]$$
$$\text{subject to Bellman constraints (8).} \tag{EC.18}$$

As before, ρ are the workers' state-contingent turnover probabilities, that is, in the model's terms:

$$\rho(X) := \Pr\{\phi > VC_{\theta_i}(X)\} \text{ for all } X\in\mathbb{X}. \tag{EC.19}$$

The same inversion used in Section E.4's step 1 to obtain a turnover policy's exit thresholds permits us to replace ρ to equivalently write:

$$\hat{\theta}_{MLE} = \text{argmax}_{\theta\in\Theta} \sum_{i,t} \left[-d_{it}\cdot\frac{VC_\theta(X_{it})}{\lambda_\theta(X_{it})} + (1-d_{it})\cdot\log\left(1-\exp\left\{-\frac{VC_\theta(X_{it})}{\lambda_\theta(X_{it})}\right\}\right)\right]$$
$$\text{subject to Bellman constraints (8).} \tag{EC.20}$$

In principle, solving (EC.20) requires re-solving the Bellman constraints (8) at every $\theta\in\Theta$ considered by the optimization routine. Alternatively, MPEC-type approaches would utilize constrained optimization routines that temporarily relax these constraints. However, neither type of approach tractably handles the full problem as presented, where the state space (hence number of constraints represented by (8)) enumerates approximately ten million.

We take two steps to tractably estimate $\hat{\theta}_{MLE}$. First, we use the first-stage estimates, $\hat{\rho}$, to convert the Bellman constraints (8)) into equivalent constraints that are linear in the state-dependent payoffs. As done in Pakes et al. (2007), we replace the Bellman equation's nonlinear term taking the maximum of next week's value function and exit value, by next week's value function plus the option value of exit. Second,

The Bellman equations (8) are equivalent to the following (EC.21):

$$VC_i(X) = \mathbb{E}_{X',\phi'|X}\left[U(X';\theta_i) + \delta\cdot VC_i(X') + \delta\cdot\lambda(h_t,d_i,\theta_i)\cdot\rho(X')\right]. \tag{EC.21}$$

(EC.21) replaces (8)'s nonlinear term, $\max\{\phi', VC_i(X')\}$, by the sum of $VC_i(X')$ and the option value of exit, $\lambda(h_t,d_i,\theta_i)\cdot\rho(X')$. The latter is given by the probability of using the option, which is $\rho(X')$, multiplied by the conditionally expected amount by which the exit value exceeds $VC_i(X')$, which is $\lambda(h_t,d_i,\theta_i)$ by the memoryless property of the exit value's exponential distribution.

Translating (EC.21) into matrix form and re-arranging yields:

$$\overrightarrow{VC_\theta} = M_C\left[\overrightarrow{U_\theta} + \delta\cdot\overrightarrow{\lambda_\theta}\circ\overrightarrow{\rho}\right] + \delta\cdot M_C\overrightarrow{VC_\theta} \tag{EC.22}$$

$$\iff \left[I-\delta\cdot M_C\right]\overrightarrow{VC_\theta} = M_C\left[\overrightarrow{U_\theta} + \delta\cdot\overrightarrow{\lambda_\theta}\circ\overrightarrow{\rho}\right], \tag{EC.23}$$

where all vectors are defined over the state space, $\mathbb{X}$, and M_C denotes the $|\mathbb{X}|\times|\mathbb{X}|$ equilibrium state transition matrix. (We suppress explicit notation that M_C and ρ depend on θ, especially as both are estimated in the first stage.)

Using (EC.23), we can readily evaluate $\overrightarrow{VC_\theta}$. For any parameters θ, the utility-based vectors $\overrightarrow{U_\theta}$ and $\overrightarrow{\lambda_\theta}$ are easily computed. First-stage $\hat{M}_C$ and $\hat{\rho}$ complete the linear constraints pinning down $\overrightarrow{VC_\theta}$. However, the number of constraints remains approximately ten million, with a commensurate number of decision variables.

Constraint sampling is enabled by linear programming techniques from the approximate dynamic programming literature (De Farias and Van Roy (2003, 2004), Farias et al. (2012)). In the optimization problem (EC.20) which defines the m-estimator $\hat{\theta}_{MLE}$, we replace the Bellman equations by the equivalent LP formulation taking constraints from (EC.23). Our estimator randomly samples a strict subset, size $L < |\mathbb{X}|$, of the constraints. With theoretical guarantees from the literature on constraint sampling, the estimator $\hat{\theta}'_{MLE}$ solves the following approximation:

$$\hat{\theta}'_{MLE} := \text{argmax}_{\theta\in\Theta} \sum_{i,t} \left[-d_{it} \cdot \frac{(\Phi\overrightarrow{s_\theta})(X_{it})}{\lambda_\theta(X_{it})} + (1-d_{it}) \cdot \log\left(1 - \exp\left\{ -\frac{(\Phi\overrightarrow{s_\theta})(X_{it})}{\lambda_\theta(X_{it})} \right\}\right)\right] \tag{EC.24}$$

where:

$$\overrightarrow{s_\theta} := \text{argmin}_{\overrightarrow{s}\in\mathbb{R}^K} \overrightarrow{c}^T \Phi \overrightarrow{s} \tag{EC.25}$$

subject to:

$$\left\{ \text{rows } r_1,\ldots,r_L \text{ of } \left[I - \delta\cdot\hat{M}_C\right]\Phi\overrightarrow{s} \geq \hat{M}_C\left[\overrightarrow{U_\theta} + \delta\cdot\overrightarrow{\lambda_\theta}\circ\overrightarrow{\hat{\rho}}\right]\right\},$$

where the constraint indices $r_1,\ldots,r_L$ are randomly sampled from the integers $1,\ldots,|\mathbb{X}|$ without replacement, Φ is an $|\mathbb{X}|\times K$ matrix with each of its K columns a polynomial basis function over the state space $\mathbb{X}$, and $\overrightarrow{c}$ a weighting vector of constants. VC_θ is replaced by an approximating linear combination $\overrightarrow{s_\theta}$ of Φ's set of columns, acting as basis functions.

Appendix F: Production Planning

We develop a tractable framework for production planning that embeds our restricted Experience-Based Equilibrium model of endogenous worker turnover (Section 5.2) into the firm's dynamic optimization of weekly production, staffing, and speed-hour workloads.

The weekly stage optimization solves for hundreds of thousands of decision variables, designating lines' speeds, hours, and staffing of individual workers, while also accounting for a dynamic cost-to-go function. Learned through recent reinforcement learning techniques (Osband et al. (2017)), the cost-to-go, V^{Firm}, crucially considers how decisions influence the coming weeks' worker turnover.

F.1. Firm's Weekly Cost Minimization

Our production planner faces a series of stochastic production demands over a planning horizon. The planner's objective function (13) is equivalent to the following, where we merely express the underage cost in terms of the assembly lines' lost yields instead of yields:

$$\begin{aligned}\textbf{Min } & c_M \times \texttt{Vol}_t \times \sum_{l\in L}\Big[(1-\mathbb{E}Y_{lt})\times \exp\big\{\texttt{logVolShare}_{lt}\big\}\Big] + \sum_{i\in W}\text{WagesPaid}_{it} + \cdots \\ & c_U \times \mathbb{E}\Bigg[\text{MPSNeed}_t - \texttt{Vol}_t + \texttt{Vol}_t \times \sum_{l\in L}\exp\big\{\texttt{logVolShare}_{lt}\big\}\times\Big(1-Y_{lt}\Big)\Bigg]^+ + V^{Firm}_{t+1}. \end{aligned} \tag{EC.26}$$

To preserve convexity, we re-run our productivity regressions using the log of lost yield, i.e., $\log(1-Y_{lt})$, as our dependent variable, instead of $\log(Y_{lt})$, with no qualitative difference in our regression results (i.e., the estimated marginal impacts are virtually identical). As before, we collect assembly lines' fixed effects and estimate the variance of the regression residual as σ_{LY}. We denote the vector of estimated coefficients for the yield regression as β_{LY}, and the productivity regressors for assembly line l in week t as $\overrightarrow{X}_{lt}$.

We outline how we evaluate the objective function, where the approach splits into two cases. Case 1 applies when the state variables $\text{MPSNeed}_t \geq \texttt{Vol}_t$, where $\texttt{Vol}_t$ has been decided in stage one.

Case 1. Consider the case where $\text{MPSNeed}_t \geq \texttt{Vol}_t$, that is, when the week's production need exceeds the firm's decided production volume. In this scenario, underage is guaranteed, even if yields were flawless. Thus applying the "positive part" function to the objective's underage cost term is superfluous, and a closed form expression for expected underage cost is readily obtained. Objective (EC.26) then becomes:

$$\begin{aligned}\textbf{Min } & c_M \times \texttt{Vol}_t \times \sum_{l\in L}\left[\exp\left\{\overrightarrow{X}^T_{lt}\beta_{LY} + \frac{\sigma_{LY}}{2} + \texttt{logVolShare}_{lt}\right\}\right] + \sum_{i\in W}\text{WagesPaid}_{it} + \cdots \\ & c_U \times \left[\text{MPSNeed}_t - \texttt{Vol}_t + \texttt{Vol}_t \times \sum_{l\in L}\exp\left\{\overrightarrow{X}^T_{lt}\beta_{LY} + \frac{\sigma_{LY}}{2} + \texttt{logVolShare}_{lt}\right\}\right] + V^{Firm}_{t+1}. \end{aligned} \tag{EC.27}$$

Case 2. Otherwise, we evaluate expected underage by simulating B draws of lost yield outcomes, $\Big(1 - Y^{(b)}_{lt}\Big)$, $b = 1,\ldots,B$, and using the soft maximum function to approximate the positive part computation. (While the hard maximum preserves convexity, convex optimization software packages encounter issues with its discontinuous derivative.) We draw the stochastic component of lost yield, $\epsilon^{(b)}_{lt}$, $b = 1,\ldots,B$, from the standard Normal distribution to replace $\left(1-Y_{lt}\right)$ by $\left(1-Y_{lt}^{(b)}\right)=\exp\left\{\overrightarrow{X}_{lt}^{T}\beta_{LY}+\sigma_{LY}\times\epsilon_{lt}^{(b)}\right\}$ over B simulations:

$$\begin{aligned}
&\textbf{Min } c_M\times\text{Vol}_t\times\sum_{l\in L}\left[\exp\left\{\overrightarrow{X}_{lt}^{T}\beta_{LY}+\frac{\sigma_{LY}}{2}+\log\text{VolShare}_{lt}\right\}\right]+\sum_{i\in W}\text{WagesPaid}_{it}+\cdots\\
&\quad c_U\times\frac{1}{B}\times\sum_{b=1}^{B}\left[\text{MPSNeed}_t-\text{Vol}_t+\text{Vol}_t\times\sum_{l\in L}\exp\left\{\overrightarrow{X}_{lt}^{T}\beta_{LY}+\sigma_{LY}\times\epsilon_{lt}^{(b)}+\log\text{VolShare}_{lt}\right\}\right]^{+}+V_{t+1}^{Firm}
\end{aligned}\tag{EC.28}$$

$$\begin{aligned}
\approx\ &\textbf{Min } c_M\times\text{Vol}_t\times\sum_{l\in L}\left[\exp\left\{\overrightarrow{X}_{lt}^{T}\beta_{LY}+\frac{\sigma_{LY}}{2}+\log\text{VolShare}_{lt}\right\}\right]+\sum_{i\in W}\text{WagesPaid}_{it}+\cdots\\
&\quad c_U\times\frac{1}{B}\times\sum_{b=1}^{B}\cdots\\
&\quad\log\left(1+\exp\left\{\text{MPSNeed}_t-\text{Vol}_t+\text{Vol}_t\times\sum_{l\in L}\exp\left\{\overrightarrow{X}_{lt}^{T}\beta_{LY}+\sigma_{LY}\times\epsilon_{lt}^{(b)}+\log\text{VolShare}_{lt}\right\}\right\}\right)+V_{t+1}^{Firm}.
\end{aligned}\tag{EC.29}$$

The soft maximum is introduced in moving from (EC.28) to (EC.29).

F.2. Reinforcement Learning

Reinforcement learning techniques update the workers' value functions, VC_i, and the planner's cost-to-go, V^{Firm}, over many simulations, until the learning updates converge to an equilibrium. We focus on our implementation, with Osband et al. (2017) serving as a general methods reference for the interested reader. We start with the workers' updates, as their learning problem is simpler than the firm's.

F.2.1. Learning Workers' Continuation Values. In deciding whether to turnover, worker i self-interestedly compares her weekly exit value, ϕ_{it}, against her expected value from continuing her employment, $VC(X_{it})$. Treating the latter as a belief, $\hat{VC}$, how can the worker's beliefs be updated so as to ultimately converge to $VC(X_{it})$ that satisfies equilibrium consistency condition (10)?

For intuition, the following outlines how we implement reinforcement learning:

1. Suppose that the worker holds *beliefs*, $\hat{VC}$, about her continuation values. We can derive $\hat{m}^*$, the decision rules satisfying incentive-compatibility under her beliefs $\hat{VC}$, such that:
$$\hat{m}^*(X_{it},\phi_{it})=1\iff\hat{VC}(X_{it})<\phi_{it}.\tag{EC.30}$$
2. In turn, for any set of decision rules, $\hat{m}^*$, we can generate new continuation values, $\hat{VC}'$, consistent with the stochastic process on $\mathcal{S}$ under the transition kernel incorporating strategies $\hat{m}^*$.
3. Given 1 and 2, it is natural to consider iterating between the steps described in 1 and 2 until $\hat{VC}'$ converges (foregoing a theoretical guarantee of convergence for now) to VC. However, step 2 involves computing $\hat{VC}'$ to high confidence using many, computationally taxing simulation runs. Instead, reinforcement learning updates beliefs more rapidly. In place of step 2, we carry out only a single simulation run after updating the transition kernel with step 1's latest strategy $\hat{m}^*$. The simulation outcome is added to the agents' accumulating body of "experiences" from past simulations, i.e., replay memory, which are then used to immediately form agents' new beliefs $\hat{VC}'$. Thus, reinforcement learning boosts

computational efficiency in learning by more frequently updating workers' strategies and passing them into the state transition kernel.

4. Computational efficiency is further enhanced by tailoring reinforcement learning to incentivize agents to explore in ways that accelerate overall learning. Three such enhancements are useful in our application. First, agents are seeded with optimistic beliefs about continuation values. A pessimistic agent tends to "over"-exit, in which case no updating experience is generated, whereas an optimistic agent is incentivized to over-explore, hence accumulating experiences and learning over time which continuation values are worthwhile. Optimistic beliefs incentivize the agent to explore the entire state space. Second, we bootstrap instead of utilizing the entire body of simulated experiences to form the new beliefs, $\hat{VC}$, seeding each next simulation run. Constructing the new beliefs, $\hat{VC}$, from a bootstrapped sample of past experiences helps to generate variance driving policy exploration. Third, the new beliefs are constructed with noise added to each observation in the (bootstrapped) set of experiences used. Such noise will tend to "cancel out" in well-explored regions, where many observations are effectively "averaged" to arrive at the believed continuation value, $\hat{VC}$. In contrast, underexplored regions that are sparsely observed will exhibit higher variance in payoffs under noise-infused beliefs. Because such variance results in instances of optimism, noise serves to target exploration where it is sparse. See Osband et al. (2017) for discussion on these techniques, including how to inject correlated noise to increase exploration.

For the kth simulation run, we model and update agents' beliefs as the continuation values, $\hat{VC}^{(k)}$, projected by a non-parametric regression over the relevant body of accumulated experiences through simulation run $k-1$ (appropriately bootstrapped with noise). Then when simulation run k occurs, the fitted regression is applied to each state X_{it} encountered by the agent in the simulation, with the output being her believed continuation value, $\hat{VC}^{(k)}(X_{it})$. Her simulated exit decision, $\hat{m}^*_{(k)}(X_{it}, \phi_{it})$, derives immediately from comparing this belief against her exit value – see (EC.30).

F.2.2. Learning Manufacturer's Cost-to-go Reinforcement learning similarly learns the firm's cost-to-go but involves greater complexity. The planner learns a nested pair of cost-to-go functions: (A) V_t^{Firm}, used in the planner's "inner" convex optimization covered by Section 5.6; and (B) an "outer" cost-to-go term, V_t^{Vol}, that facilitates the planner's weekly decisions on overall production volume, $\texttt{Vol}_t$, and how it is distributed among the assembly lines, $\texttt{logVolShare}_{lt}$.

Weekly optimization's cost-to-go. As the firm decides on assembly lines' workloads and staffing, the "inner" cost-to-go, V_t^{Firm}, accounts for the effects on variable costs incurred in future weeks. (For instance, overtime hours incurred now do not pay out in wages until the next payweek.)

For tractability, it will be important to preserve the convexity of the production planning objective (13). For this reason, we allow V_t^{Firm} to be highly flexible in the state variables, such as the MPS, the chosen production quantities, $\texttt{Vol}_t$ and $\texttt{logVolShare}_{lt}$, worker experience retained, new hires, and unpaid compensation; however, we constrain V_t^{Firm} to be affine in the decision variables of objective V_t^{Firm}. Note, nonetheless, that V_t^{Firm}'s intercept and slopes for these decision variables may flexibly depend on the state variables, and V_t^{Firm} is time-inhomogenous, allowing for such slopes to depend on the week's position within the pay cycle and the product life cycle. As a concrete example, the cost-to-go's sensitivity to assembly line attrition rate is permitted to depend on whether the MPS is ramping and on whether such attrition occurs early or late in the product's life.

We permit V_t^{Firm}'s intercept and decision-variable coefficients (i.e., the slopes representing cost sensitivities to the decision variables) to depend non-parametrically on the following state variables, which capture production needs, attrition rates, compensation due, and the available worker mix:

- ◇ The MPS order quantity, backlog, and next week's anticipated MPS order quantity;
- ◇ The manufacturer's decided production quantity, $\texttt{Vol}_t$, as a percentage (possibly exceeding 100% to permit pre-building) of the MPS order quantity and backlog;
- ◇ Average worker experience;
- ◇ Share of dispatch workers;
- ◇ Share of new hires; and
- ◇ Workers' total unpaid compensation accrued to date.

The cost-to-go, V_t^{Firm}, depends on the following decision variables reflecting staffing and workloads. Of these, the final two decision variables summarize whether the manufacturer chooses to staff and operate lines unevenly by worker type (e.g., concentrating regular or dispatch hires together on assembly lines).

- ◇ Average assembly line attrition;
- ◇ Average log hours of operation;
- ◇ Average log speed;
- ◇ The Herfindahl-Hirschman Index (HHI) measuring the concentration of experience stocks among lines;
- ◇ The HHI measuring concentration of log hours among lines;
- ◇ An HHI measure of imbalances in lines' staffing levels of dispatch and direct hires (by experience stock); and lastly
- ◇ Interactions of lines' lagged hire-type imbalance HHIs with their current attrition, log hours, and new hire staffing percentages.

We operationalize V_t^{Firm} by nonparametric, sieve-polynomial regression, on expansions of the decision and state variables while restricting the degree of the decision variables to one in each polynomial term. Once the regression coefficients are estimated, we organize them into a matrix that projects any vector of state variables down to a vector representation of V_t^{Firm} as a linear function of the decision variables. While we believe that our specification pushes the feasible limit of computational resources at our disposal, more complex and general specifications can be used at computational cost.

Optimizing production quantities. We similarly learn V_t^{Vol}, the manufacturer's cost-to-go on its production quantity decisions. Like V_t^{Firm}, the cost-to-go V_t^{Vol} is a time-inhomogenous function of state and decision variables jointly. Unlike for V_t^{Firm}, we do not restrict V_t^{Vol} to be linear in its decision variables, and we impose only the weaker condition that V_t^{Vol} be additively separable into sets of terms involving each decision variable.

The cost-to-go V_t^{Vol}'s optimization intends to solve for the production-quantity decision variables, $\texttt{Vol}_t$ and $\texttt{logVolShare}_{lt}$. However, we re-normalize these variables and reduce the dimensionality of the decision space by treating the following as the decision variables in V_t^{Vol}:

- ⋄ Production volume as a percentage (possibly exceeding 100%) of the MPS order quantity and backlog, that is, $\texttt{Vol}_t/\text{MPSNeed}_t$; and
- ⋄ Instead of directly handling the production shares of the forty-four assembly lines, $\texttt{logVolShare}_{lt}$, as the decision variables (i.e., forty-three degrees of freedom), we "score" the assembly lines on the basis of select metrics, such that the scores determine their relative workloads. The decision variables are the coefficients used to score the assembly lines based on the following metrics: turnover rate, average experience level, and net difference in experience stocks of the two hire types (i.e., a measure of balance in mix of the two worker types). For any set of decided coefficients, the resulting scores, $s_l \in \mathbb{R},\ l \in L$, assign to assembly line l the production quantity share of $\frac{\exp\{s_l\}}{\sum_{k\in L}\exp\{s_k\}}$.

V_t^{Vol}'s intercept and decision-variable coefficients are permitted to vary flexibly as non-parametric functions of the state variables:

- ⋄ The MPS order quantity, backlog, and next two weeks' anticipated MPS order quantities;
- ⋄ Turnover rate;
- ⋄ Average worker experience;
- ⋄ Share of dispatch workers;
- ⋄ Share of new hires;
- ⋄ Workers' total unpaid compensation accrued to date;
- ⋄ HHI measuring the concentration of experience stocks among lines; and
- ⋄ HHI measuring imbalances in lines' staffing levels of dispatch and direct hires (by experience stock).