Gig Workers and Performance Pay:
A Dynamic Equilibrium Analysis of an On-Demand Industry

Michal Hodor†

click here for latest version

October 23, 2020

Abstract

In many online product markets, firms manufacture and supply products almost immediately after receiving orders. Thus, firms need to ensure that their workers satisfy product demand, which can vary over time in a cost-effective way. This paper develops and estimates a dynamic equilibrium model of firm and worker behavior in an "on-demand" production context. The firm solves a dynamic discrete choice-cost minimization model in which it faces uncertainty about future product demand and workers' productive capacity. The firm chooses to employ two types of workers – gig workers and permanent workers – and sets parameters of a compensation scheme that is a mix of salary and performance-based incentives to elicit worker effort. Heterogeneous workers solve a daily-effort choice problem given the compensation scheme offered by the firm. I estimate the model and perform an out-of-sample validation of it using panel data from an online, global manufacturer that produces customized items. The data include detailed measures of workers' output and output quality under varying compensation schemes. I find that gig workers and permanent workers exhibit different production patterns, and that gig workers are much more responsive to incentive pay. I embed workers' optimal effort decisions into the firm's dynamic cost minimization problem and use simulation methods to derive optimal labor force composition and compensation schemes. I show that varying the compensation scheme over time and using a mix of gig and permanent workers provides the flexibility that a firm needs to effectively operate in an on-demand customized production environment.

Keywords: Labor Force Composition, Wage Level and Structure, Personnel Economics.
JEL Codes: J21, J22, J23, J31, M5.

I am immensely grateful to my advisors Petra Todd, Hanming Fang, and Iwan Barankay for their invaluable support throughout all stages of this project. I also thank Aviv Nevo, Juan Pablo Atal, Holger Sieg, Andrew Shephard, and seminar participants at the University of Pennsylvania. All errors are my own.
† University of Pennsylvania, Economics Department (mhodor@sas.upenn.edu)

1 Introduction

Major changes have occurred in the global economy as online commerce continues to grow and firms increasingly offer on-demand products.[1] Traditionally, manufacturing firms operated assembly lines that produced large quantities of products that were kept in storage until delivery. In on-demand manufacturing, an adjustable process aims to produce customized items based on real-time data with minimal possible lead time.[2] This paper explores the optimal operation of these "new" types of firms and implications for the growth of the gig economy.

The advantages that come with customized, on-demand manufacturing are two-fold: customers obtain exactly the product that they want, and the firm only has to produce as much as it sells. That means no excess inventory or over-production in times of high expected demand.[3] However, operating an on-demand production system poses enormous challenges in planning and committing to a workforce capacity.

This issue naturally raises the question of how a firm optimally operates a customized on-demand production system. There are two instruments that relate to contract design and workforce composition the firm can integrate into its labor management practices to respond quickly to changing demand. The first is a performance-based compensation scheme that can be used to induce higher labor input on the intensive margin by eliciting greater output production from the existing workforce (see Oyer & Schaefer, 2010, for review). The second is an adjustable labor force input that can be achieved by hiring temporary on-demand workers (hereafter, gig workers), while also maintaining a stable core of more permanent workers.[4] Gig workers provide workforce flexibility on the extensive margin. In practice, one can think about these two instruments as a trade-off between labor quantity and effort, and the decision over how to optimally combine them is an empirical one.

The major challenge in determining the firm's optimal labor force composition and compensation pay structure in response to changing demand conditions is that worker productivity responds endogenously to the firm's pay structure. Moreover, worker productivity may vary over time (with experience) and by worker type. To analyze this problem, I develop a structural equilibrium framework that includes both the firm's and the worker's decisions. I solve the model in two stages. In the first stage, workers solve for an optimal daily effort decision in relation to varying compensation schemes and daily workloads.[5] The worker's problem embodies features that explain productivity differences between permanent and gig workers. Specifically, workers

[1] In the U.S., online retailers brought in nearly half a trillion dollars over the past year, representing about 9% of total sales. Globally, the trend is even stronger, with around 1.66 billion online shoppers having spent $2.3 trillion in 2017. By 2021, sales could more than double from today's levels (U.S. Census Bureau).

[2] Lead time is the time between initiation of a production process until the product is delivered to a consumer.

[3] Excess inventory or over-production can lead to markdowns and a negative impact on the manufacturer's margin.

[4] The U.S. Department of Labor Statistics defines gig work is a contingent or alternative employment arrangements consists of income-earning activities outside of traditional, long-term employer-employee relationships. Gig workers are hired to complete a particular task or for a certain period of time, and thus they tend to be temporary or project-based workers.

[5] Daily workload is defined as the product demand divided by the number of workers in a given day.

are heterogeneous in their total factor productivity, intrinsic motivation, and effort costs.[6] In the second stage, the firm solves a dynamic discrete-choice cost minimization problem by choosing both an optimal compensation scheme (flat wage or performance incentive pay) and an optimal labor-force composition (composed of both gig and permanent workers). Worker effort decisions, job experience, and on-the-job learning processes, as well as uncertainty in their productive capacity, are integrated into the firm's problem through an incentive compatibility constraint. In addition, the firm's problem accounts for product demand uncertainty, inventory limitations, and short-demand lead times – all features that characterize firms producing on-demand customized products, and more generally, features common among global export-oriented firms.[7]

I estimate and perform out-of-sample validation of this model using a novel data set that I obtained from an online, global, mid-size manufacturer of customized apparel and accessories (hereafter, the firm).[8] The data contain rich information on individuals that worked for the manufacturer in 2015 and in 2018, including detailed objective measures of daily production quality, daily wages, and information on daily product demand and pay structure.[9] This firm is especially suitable for the analysis carried out in this paper, as the majority of its employees are gig workers hired on a seasonal basis with short-term contracts. The firm also maintains a small permanent workforce. Interestingly, the firm sometimes deviates from a flat-wage to a bonus pay scheme in response to peaks in product demand. This pay change is abrupt and short-lived, and it provides exogenous variation that is used to identify the models' parameters.

In addition to constructing the model, I provide descriptive regression evidence about the production patterns and the response to incentives of gig and permanent workers. Whereas permanent workers produce close to constant amounts over various levels of job experience, gig workers exhibit a parabolic production pattern over time spent in the firm. Furthermore, gig workers exhibit statistically and economically significant responses to incentives that amount to between 12% and 17% productivity gains on average, but permanent workers' average production response to incentives is not statistically different from zero for almost all job experience levels. These results support other recent evidence that workers in jobs with flexible schedules and locations tend to be more productive (Angelici & Profeta, 2020; Bloom, Liang, Roberts, & Ying, 2015). The structural model estimation explores the origins of these differences.

The equilibrium model is solved given the actual incentive pay schedule offered by the firm, under the assumption that the firm and workers optimally respond to this pay structure. I use

[6] Intrinsic motivation incentives are generally defined as actions taken through non-monetary rewards. In the psychology literature, Deci (1971) and De Charms (2013) argued that individuals undertake many activities without expecting an extrinsic reward. In the management literature, Amabile (1993) introduced worker's intrinsic motivation into models of labor supply, and Amabile, Conti, Coon, Lazenby, and Herron (1996) identified job flexibility as an important managerial practice to foster creativity at work.

[7] Inventory limitations could be due to customized production or high inventory costs, both of which impact many manufacturers in the Far East.

[8] For confidentiality reasons, the manufacturer's identity, location, and product description are not revealed.

[9] The documents jobs are of a low-skill, repetitive, and autonomous nature and include various processes in the assembly-line chain. The nature of these blue-collar tasks is common among other global labor-intensive export-oriented manufacturers.

the data from 2018 for estimation, and the data from 2015 to validate the model by forecasting worker and firm behaviors and then compare the forecast to actual data. The first stage of the model solves for a worker's daily-effort decisions exploiting temporal pay scheme changes. In particular, I use an indirect inference estimation procedure to evaluate the underlying parameters defining workers' productivity by matching coefficients of regressions over workers' productivity, wage gains, and measures of defective production. I find that the personal motivation and total factor productivity of gig workers are more than twice as large as those of permanent workers. The personal-motivation parameter captures the idea that individuals exert higher effort (or refrain from shirking) not only because they get paid, but also for non-pecuniary reasons, such as being task-oriented, experiencing less job burn-out, valuing job flexibility, or simply because they like performing certain activities.[10]

At the second stage, I solve the firm's decision problem for each week in a calendar year as a dynamic discrete-choice problem. This estimation exploits the observed time variation in product demand associated with extensive hiring of gig workers and observed pay changes. Specifically, I apply a simulated maximum-likelihood approach to estimate the hiring costs associated with gig and permanent workers, as well as the lay-off costs and natural separation rate of permanent workers.[11] I find that gig workers' hiring costs are significantly lower than those of permanent workers. This finding is consistent with the notion that permanent workers go through a rigorous screening process and are hired on a long-term contract whereas gig workers are hired to "fill the gap."

With regard to the labor-force composition decisions, the firm bears the costs of recruiting workers, which vary by worker type due to availability, screening process, and job proficiency. Hiring gig workers allows firms to avoid paying fringe benefits and layoff costs (Mas & Pallais, 2020), but comes at the expense of productivity, which is higher for the permanent labor force. With regard to the contract design, each worker type responds differently to monetary incentives.[12]

Knowing that the firm experimented with the performance- based pay parameters as well as with times of implementing the incentive pay during my observation period, I do not impose that those choices were made optimally when estimating the model. Rather, I investigate whether the firm can improve its current choices by performing simulations based on the estimated model in which I characterize equilibrium choices made by workers and the firm in different scenarios. Initially, I maintain the same incentive structure parameters as those offered by the firm, but I assume that the incentive pay was offered more frequently throughout the year, especially around times of expected demand surges. Interestingly, although these assumptions depart only slightly from the original settings, letting the firm solve for the optimal labor-force composition given this

[10] Deci, Olafsen, and Ryan (2017) offered a review of studies that have examined employees' intrinsic and extrinsic motivations in a workplace. Recent papers in economics have studied wages and incentive schemes when workers are intrinsically motivated (Benabou & Tirole, 2003; Besley & Ghatak, 2005; Francois, 2000; Glazer, 2004).

[11] This analysis is done under the assumption that gig workers are hired for predetermined periods, and thus no layoff costs or separation rate from the firms are associated with these workers.

[12] More generally, from a monopoly perspective, one can think about gig workers as enabling the firm to increase employment at the marginal costs without raising wages on the permanent workforce.

setup yields a reduction of more than 8% in labor costs.

Next, I optimize over the set of parameters that define the pay structure, given uncertainty about future product demand and workers' productive capacity. The solution shows that the optimal performance-based incentive structure and the corresponding optimal compensation schedule and labor-force composition choice throughout the year result in a 26% reduction in labor costs. However, this outcome comes at the expense of workers' utility, which is reduced by 8% on average. These results thus raise the question of whether there exists an incentive pay structure that is different than the one originally offered and such that the firm and the workers can benefit from it. I answer this question by adding workers' participation constraint into the optimization problem over the set of parameters that define the incentive pay structure. I find an incentive pay structure that reduces the firm's labor costs by 12% and keeps workers' participation constraint binding. Importantly, I further find that maintaining workers' base wage at the original level is the key feature of the payment structure that prevents harming workers' utility even in the face of generous performance incentives.

The paper proceeds as follows. Section 2 describes how the paper contributes to the existing literature. Section 3 discusses the data. Section 4 outlines the static and dynamic models of the worker's and firm's problems, and Section 5 discusses identification. Section 6 presents reduced form regressions results. Section 7 describes the estimation methods, and Section 8 discusses the results. Section 9 discusses counterfactuals, and Section 10 concludes.

2 Related Literature

This paper relates to several strands of literature concerning the relationship between a worker and a firm. Broadly, it relies on agency theory, as developed Mirrlees (1976) and Hölmstrom (1979) who proposed the canonical principal-agent framework that has extended in multiple directions over the past decades (see Laffont & Martimort, 2009, for review).[13] This mainly theoretical literature has established the qualitative arguments for the relationship between the firm and the worker under moral hazard. Recent work by Georgiadis and Powell (2020) improved the practical applicability of classic agency theory of incentive contracts using experimental data under the assumption that the principal does not know the workers' objective (see also, Gottlieb and Moreira (2015) and Carroll (2015)). I develop an empirical framework and use it to construct a performance-based contract design relevant to an actual working environment (under the assumption that only the distribution of output production is known to the firm). I contribute to this literature by using data from a real firm and thus incorporating features of a realistic working environment, including uncertainty about both product demand and the total productive capacity

[13] For example, team production(Holmstrom, 1982), dynamic incentives(Holmstrom & Milgrom, 1987), multitask problems (Holmstrom & Milgrom, 1991), commitment problems (Laffont & Tirole, 1988), behavioral agents (Bénabou & Tirole, 2002; Benabou & Tirole, 2003), and private information (Carroll, 2015; Chade & Swinkels, 2019; Foarta & Sugaya, 2019).

of the firm's workforce. Moreover, I tie the contract design decision to the labor-force composition decision and embed these components in an equilibrium framework.

The linkage between worker productivity and pay contract structure has been explored by empirical studies that established a positive relationship between incentives and productivity in a variety of settings. For example, Lazear (2000) studied the effect of change from flat wage to a piece-rate scheme on windshield repairers; Shearer (2004) and Bellemare and Shearer (2011) conducted a field experiment and used structural estimation to study the response to incentives among tree planters; and Fehr and Goette (2007) showed in a field experiment that bicycle messengers elasticity of effort per hour is negative. More recent papers by Hong, Hossain, List, and Tanaka (2018), Guiteras and Jack (2018) and Balbuzanov, Gars, and Tjernström (2019) examined the response of workers to incentives in developing countries and focused on the tradeoff between output quality and quantity. The current paper investigates similar aspects including productivity, effort elasticity, and a quality-quantity tradeoff; however, it expands upon the scope of existing studies by comparing and contrasting gig and permanent workers.

Gig jobs and other flexible and non-traditional work arrangements have been examined in a literature that explores recent changes in labor markets structure (see Mas & Pallais, 2020, for review). Multiple studies have focused on the operational and pricing decisions for service jobs of a self-scheduling nature, such as ride-hailing drivers. For example, Chen, Rossi, Chevalier, and Oehlsen (2019) found that the labor surplus from Uber drivers' ability to adjust their work schedules is twice the surplus they would have in less-flexible arrangements, and Hall, Horton, and Knoepfle (2020) used differences in timing, city size, and exogenous price-index changes to identify driver responses to fare changes. Other papers focus on workers' productivity. The seminal paper by Bloom et al. (2015) conducted a randomized controlled trial within a Chinese call center employees and found that working from home increases productivity. Following this work, Beckmann (2016) showed that work-time autonomy promotes worker and firm performance in German firms, and Angelici and Profeta (2020) employed a randomized experiment on a sample of workers in a large Italian company and found that job time and space flexibility increased worker productivity. In addition, papers have also examined the extent to which workers value job flexibility. Specifically, studies used surveys (Eriksson & Kristensen, 2014; Maestas, Mullen, Powell, Von Wachter, & Wenger, 2018; Wiswall & Zafar, 2018) and a discrete choice experiment (Mas & Pallais, 2017) to study workers' willingness to pay for a broad set of job characteristics, and found that workers are generally willing to pay for non-wage attributes. Along similar lines, the current paper examines a broad set of aspects of workers in a non-traditional work arrangement, including productivity and effort responses to flexibility and incentives, as well as exploring workers' intrinsic motivation and their effort cost structure. This paper also provides a structural analysis to illuminate the origins of these behaviors, preferences, and production patterns

The current study goes beyond existing studies in three major ways. First, it considers an alternative work arrangement in manufacturing.[14] Second, it studies a different aspect of job flexibility by which workers do not have flexibility over the shift's length or the job location (shifts are assigned by a manager), but rather over the length of the employment duration (which ranges between several days and two months). Third, the paper considers how a firm optimally uses a blended workforce of permanent and gig workers.

The firm's decision about its labor force composition relates to studies that solve for firms' labor decision problem. Models that consider flexible work arrangements date back to the traditional framework proposed by Rosen (1986). In this framework, the firm's decision of whether to adopt flexible scheduling depends on whether the (monetary and non-monetary) benefits outweigh the costs (under a flat, market-clearing wage assumption).[15] This simple structure has been extended in various ways. Papers that are particularity relevant to the current paper are personnel-scheduling models in the operational management literature that have studied the labor-force composition decision with varying employee classifications and demand settings. Examples are Pinker and Larson (2003) who developed a theoretical model for an optimal workforce decision composed of permanent and temporary workers, and **?**, who considered a similar problem in service organizations incorporating differences in employee skill levels and demand uncertainty. Most relevant is the paper by Dong and Ibrahim (2017), who investigated the hiring decisions of gig and permanent workers in service jobs and showed that such decisions depend on operational costs, labor supply-side uncertainty, and flexibility to meet the time variation in customers' demands. Unlike their paper, which is highly theoretical, the current paper applies an empirical and data-driven approach in a manufacturing environment and additionally considers the pay contract design – an inseparable aspect of the labor composition problem.

3 Data and Settings

3.1 General

The recent micro-data used in this paper are from an online, global, mid-size manufacturer that produces customized apparel and accessories. The data contain rich information on individuals

[14] The practice of hiring gig workers (in addition to permanent workers) among manufacturers is no longer considered a short-term solution and has become a common human resource strategy to adjust production to demand variation and short-demand lead times (Foote & Folta, 2002). However, there has been little academic research on the management practices of gig workers in a manufacturing environment.

[15] Models that departed from the fixed-wage assumption adopted the efficiency wage hypothesis, which claims that a wage should be higher than the market-clearing wage in order to encourage workers to increase their productivity (and potentially reduce turnover) (Akerlof, 1982; Shapiro & Stiglitz, 1984). Other models followed the tournament theory, which suggests workers can be rewarded by their rank in an organization (Lazear & Rosen, 1981; Shapiro & Stiglitz, 1984). Studies that have taken these approaches have been mostly theoretical, focusing on worker productivity and abstaining from examining the labor-force size or composition. This paper adopts a similar approach and departs from a fixed-wage assumption by considering bonus performance pay incentives. It then uses this approach to expand the traditional labor-decision problem by constructing an empirical framework that couples the wage decision with a workforce size and composition decision.

that worked for the firm in 2015 and in 2018.[16,17] During these years, the firm's management adopted new compensation practices. Specifically, around the months of February 2015 and December 2018, the firm deviated from its usual pay structure – a flat hourly wage – and introduced a performance-pay scheme with the goal of augmenting productivity in peak demand times. The data from 2018 are used as the main data source throughout the paper, as it includes a greater number of workers and provides detailed information about the quality of production and production score (subsequently discussed in detail). The data from 2015 are used to perform an out-of-sample validation test of the model by evaluating its forecasting performance.

Table 1: DESCRIPTIVE STATISTICS

	Worker Type		
	Permanent	*Gig*	t-Stat
Female	0.86	0.83	0.57
Age	32.27	24.27	4.85
Shift Length	7.22	8.00	-3.78
Experience $_{\text{Days}}$	328.68	23.35	14.67
Avg. Production $_{\text{Day}}$	122.02	97.50	5.12
Avg. Defective Production $_{\text{Day}}$	2.45	3.41	-2.88
Avg. Production Score $_{\text{Day}}$	155.41	138.67	1.65
Number of Workers	44	216	

Notes: Production, defective production and production score measures are adjusted to 8 hours of work.

3.2 Employment

The classification of workers into types – permanent or gig – is based on employment data recorded by the firm's personnel department. For each worker observed, there are records of employment dates[18] used to infer whether a worker was hired as a gig worker with a short-term contract or as a permanent worker with a long-term contract. I combined the employment information with daily attendance records to generate an experience measure that counts the number of days a worker actually worked.[19] Table 1 shows that during the year of 2018, 216 (83%) of the 260 workers in the assembly department were gig workers. As expected, the average number of days of experience among gig workers (23) is significantly lower than that of permanent workers who were employed on average for one year.

[16] Months February through December.

[17] Although the data provide information on production process steps, the analysis restricts attention to the assembly department for several reasons. First, this department is the only one along the production chain that is solely based on human-capital labor and involves no machines. This fact enables the identification of a trade-off between labor effort and labor-force size and eliminates alternative trade-offs, such as a labor-capital substitution. Second, the assembly department is the largest in the plant, as 98% of the products require assembly. This fact makes the total number of orders received in a given day a good proxy for workers' workload. Lastly, the assembly department is the last station through which the items pass before reaching the quality-assurance department, which makes the quality measure most accurate and prevents records of false failures.

[18] In cases of repeated employment, all documented dates are considered.

[19] For employees whose employment-start date occurred prior to the beginning of the data, the number of days of experience is approximated based on the observed data.

The employment records include information about the workers' gender and for most employees, their age.[20] Table 1 shows that the majority of assembly workers are females among both worker types: the share of females comprise 86% and 83% for permanent workers and gig workers, respectively, and the difference is not statistically significant. The statistics pertaining to workers' age by type indicate an interesting pattern and imply that permanent workers and gig workers are inherently different. Gig workers are on average eight years younger than permanent workers, a difference that is statistically significant.

3.3 Production

The production data set is assembled from several sources. First, the total production records were generated by a sophisticated monitoring system that documents daily personal output for each worker at each workstation.[21,22] Second, production quality measures are inferred from the quality-assurance department records. For each item, this department's records indicate whether it passed or failed a quality check, and in the case of a failure, the record specifies the reason(s). As each item is linked to all of the workers involved in its production, I was able to deduce the amount of low-quality output generated by each worker. Lastly, each worker-day observation was matched with a shift-length record documented by the attendance system. I used this data to adjust workers' production by the number of hours worked and thus created a production measure that is comparable across workers.

In addition to the production volume and quality, the data include records of production score for each worker-day observation. The key difference between production score and "regular" production is that the former takes into account the complexity associated with assembling each item and gives higher score to items that require longer assembly time.[23] The production score, as well as total production quantities, are known to workers and presented in their station monitor. In practice, workers' bonuses are determined by this adjusted production measure. Because the assembly department operates in a way in which each worker chooses the items she assembles, such adjustment is essential. It helps to eliminate a scenario in which workers choose only items that can be finished quickly to increase their bonus.

Table 1 presents a comparison between the average daily production and score of each worker

[20] Originally, the personnel department only recorded a worker's date of birth if she was hired as a permanent worker. Once we began collaborating, however, they started to record gig workers' birthdates as well. Thus, the data set includes the ages of all permanent workers and the ages of gig workers that were hired towards the end of 2018.

[21] The system was initially installed for reasons other than the implementation of the performance-pay system, thus its cost can be ignored when analyzing the trade-off between productivity and profitability.

[22] Once an order is placed online, the production process begins with the generation of a barcode for the ordered item. When a worker starts to work on an item, she scans this barcode, an action that links the item to herself and her workstation. Aggregation of this information generates a detailed database of workers' daily production at each step in the production chain.

[23] The scoring menu was established during 2018 and finalized and presented to workers several weeks before implementation of the incentive scheme. Figure A.3 in the Appendix shows the production and imputed production based on the production score used to adjust the threshold of the performance-based pay regime.

type and implies a dramatic difference whereby permanent workers are more productive than gig workers. Although this conclusion may seem intuitive, as permanent workers are presumably more skilled than gig workers, it is in fact misleading. Gig workers are usually hired before or during times of demand surge, whereas permanent workers work (and hired) throughout the year. Workers' productivity is a function of the workload, as well as their experience. As such, gig and permanent workers are working under intrinsically different conditions and merely comparing the average production yields an incomplete picture. Comparing the productivity levels of permanent and gig workers requires a more comprehensive framework that controls for demand conditions and varying pay schemes, as is performed in this paper.[24]

3.4 Demand

The daily demand variable represents the sum of all orders placed online in a given day. Figure 1a displays a time series of the product demand during 2018. The dashed gray line represents the daily orders and demonstrates the high variability of product demand. The solid black line represents the average of these daily orders over weeks and emphasizes that the product demand is characterized by extreme seasonality around three annual holidays – Valentine's Day, Mother's Day, and Christmas – as represented by the gray solid lines, in this order. The firm's strict production policy is a key feature in explaining this demand pattern. Within the plant there are up to four days of production, starting on the day the order is placed and ending on the day the item is shipped. The firm guarantees its consumers the minimum possible shipment lead time.[25] Combining this policy with producing items entirely customized for consumer preferences creates the seasonal demand pattern observed in the figure. The demand peaks are as much as five times higher on average than the average volume in normal periods and they last from a week to two months. In fact, this pattern is not specific to the year of 2018 and is evident in the 2015 data as well.[26] My model assumes that the firm's forecast depends on that seasonal demand pattern with additional daily idiosyncratic shocks.

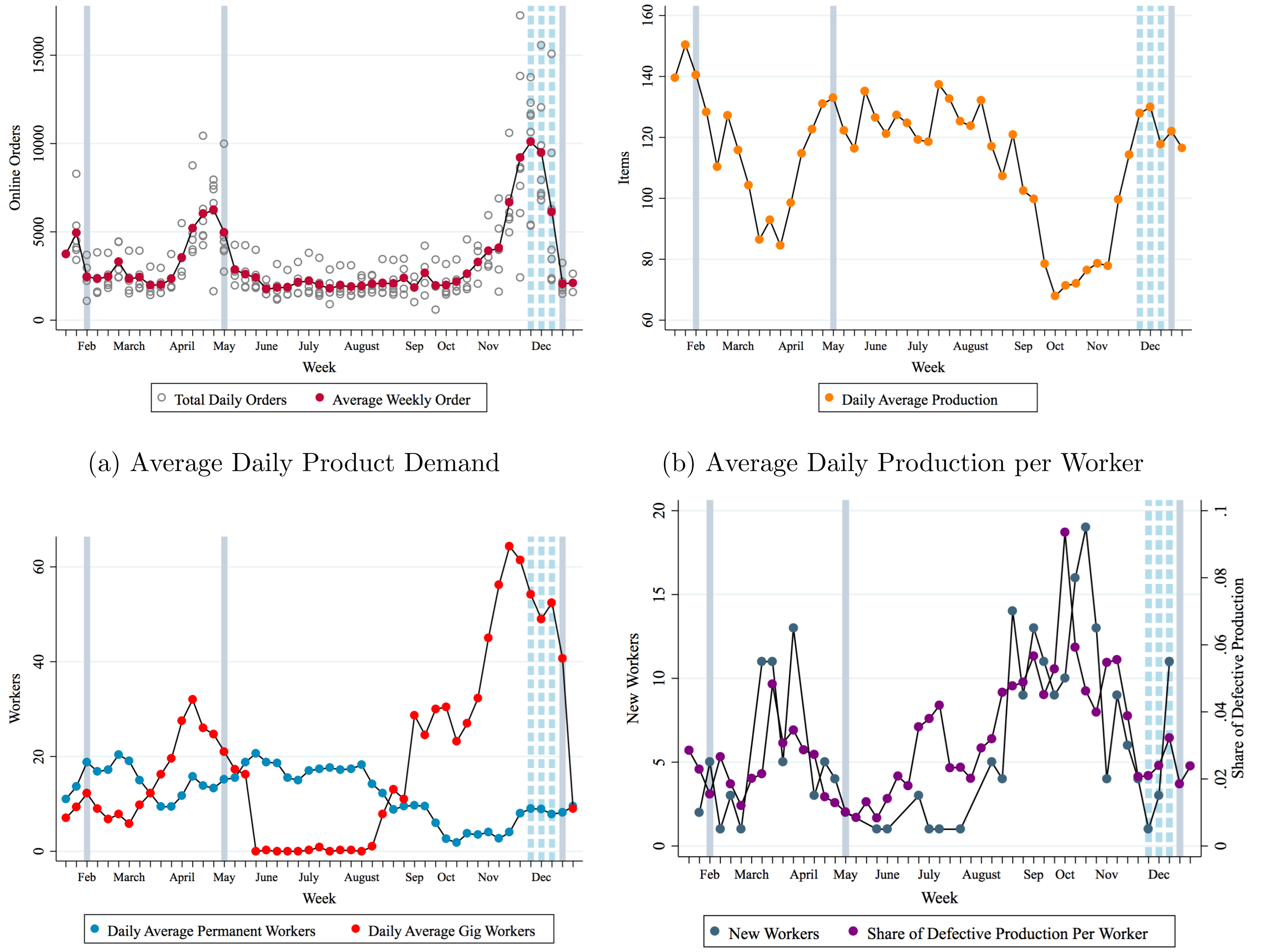
3.5 Labor Force Size

Figures 1b and 1c presents a time series of average production per worker and average number of workers by type over the weeks during 2018. They show that production is positively related to demand surges and also affected by the number of workers. However, the picture is not that simple. As illustrated in Figure 1c, while the number of permanent workers remains stable throughout the year, during high-demand periods many new and inexperienced gig workers join the labor-force pool. If job skills and experience are significant elements in determining production, workload might remain relatively high even in the face of mass hiring. At the same time, the performance-

[24]Also, Table 1 shows that permanent workers work less on average than gig workers, however, the manager, not the worker, determine shift length is determined.

[25]Shipment lead time depends on where the item is being shipped.

[26]Figure A.1 illustrates the average product daily demand over weeks during 2015.



(a) Average Daily Product Demand

(b) Average Daily Production per Worker

(c) Average Daily Number of Workers by Type

(d) Average Share of Daily Defective Items and Daily Number of Newly Recruited Workers

Figure 1: Data Description

Notes: Daily product demand is the sum of all the orders placed online in a given day. The gray dots in Figure 1a represent the daily orders and the red dots represent the average of these orders over weeks. Production per worker and defective production per worker are adjusted for eight hours. Gray solid lines represent (in this order) Valentine's Day, Mother's Day, and Christmas. Blue dashed lines represent times in which the performance-pay scheme is in effect.

pay schedule also changes workers' production incentives irrespective of the number of workers and demand.

Figure 1d sheds light on the relationship between experience and job proficiency by showing a time series of the share of defective items out of total production and the number of new workers over the weeks of 2018. New workers do not undergo an official training program; in practice, they have on-the-job training. Knowing this, if experience serves as an important input in workers' production function, one would predict an increase in the proportion of defective items associated with the massive hiring of new gig (or permanent) workers. The figure presents evidence to support this prediction. The time series presented in Figure 1d are highly correlated, which implies that workers learn their jobs by becoming familiar with the production process over time. However, the increase in the share of defective items might also be associated with increases in demand. That is, workers might rush production during times of high demand at the cost of reducing output quality. This idea is of particular importance when workers are compensated based on production quantities in times of incentivizing payment schemes. Disentangling all the abovementioned simultaneous forces and understanding the respective roles and levels of importance when searching for a firm's optimal labor-management behavior is a goal of this paper.

3.6 Pay Incentives

During the time under consideration, the firm switches between two pay regimes: a flat wage and a performance-based wage. During periods in which the flat-wage scheme is in effect, workers are paid based on a fixed hourly rate regardless of production. During periods in which performance-based pay is in effect, workers are subject to a wage structure of incentive pay with a minimum guarantee. That is, production above a certain threshold rewards worker with additional per-unit pay on top of their fixed wage, while production below the threshold reverts back to the flat-wage regime. Figure A.2 in the Appendix presents this exact wage structure, which was offered during the 2018 year.

There are several reasons to adhere to a performance-based incentive pay with a minimum guarantee. From the firm point of view, this wage structure facilitates compliance with labor laws, unlike a pure performance-pay that could create problems related to minimum wage, overtime compensation, and record-keeping obligations. Viewed by the worker, this incentive pay structure completely eliminates the risk associated with pure performance-pay structure and moreover, it guarantees that workers cannot be worse off under the performance-based scheme by giving them the opportunity to increase their pay.

Generally, the mechanism by which workers respond to such a performance-pay system is not obvious, and several issues exist related to the identification of effort response. First, an econometric challenge arises when the incentive contracts are endogenous to the firm performance; that is, when changes in incentives could reflect other changes in the firm's management practices.

Second, the literature has emphasized a concomitant change in workforce composition when introducing performance pay, because high-ability workers are attracted to this form of pay.[27] Last, an issue exists with the output quality generated under performance-pay schemes, as such schemes might motivate workers to speed up production by compromising quality.[28]

The data set used in this paper overcomes all of these challenges. First, the information comes from a firm that introduced exogenously timed variation in its incentive structures, which is orthogonal to its management methods. Therefore, concerns pertaining to endogenous management behavior are eliminated. Moreover, neither the firm nor the worker know beforehand when or for how long incentives will be offered. In practice, incentive pay has been offered for a short period of approximately one month, so workforce sorting and selection concerns are irrelevant. Finally, the quality issue is addressed directly by the firm because workers are only rewarded for points earned on high-quality items. When the quality assurance department marks an item as a failure, the score that the workers see on their personal monitor immediately reflects it, thus ensuring workers are attentive to quality during production. This system prevents workers from exploiting the performance-pay system and guarantees its efficacy in increasing high-quality production.

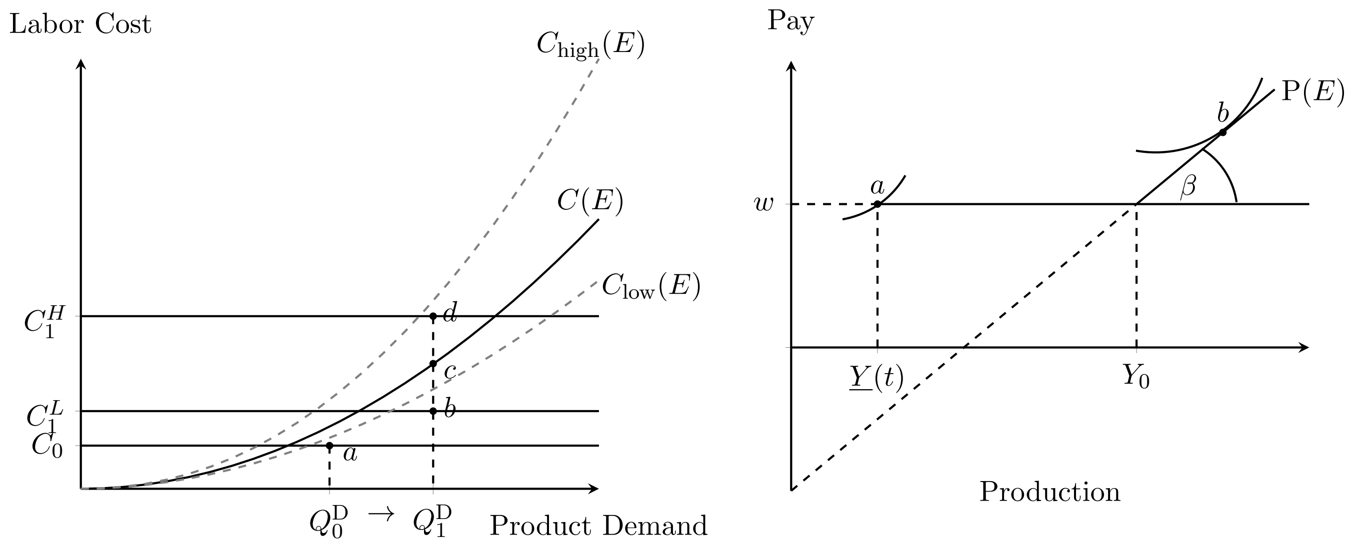
The literature has discussed the potential unintended consequences on workers and firm performance resulting from implementation of incentive pay scheme. Specifically, performance-based pay could lead to an inevitable change in the distribution of earnings across workers, which in turn may lead workers to reduce co-worker cooperation. (Baron & Pfeffer, 1994; Bewley, 1999; Lazear, 1989), workers' sabotaging co-workers' performance, or workers directly worse off in utility terms as a result of their structural adversity to pay inequality (Charness & Rabin, 2002; Fehr & Schmidt, 1999). These concerns are not pertinent under the setup studied in this paper. The task of assembly is completely autonomous - i.e., employees work in individual stations and no social interactions are required and thus concerns related to cooperation and sabotage become irrelevant. Additionally, the performance-pay scheme was instituted for a short term and thus any indirect adverse effects that arise as a result of a competitive pay system are less plausible.

4 Model

This section presents the equilibrium model of the firm's labor cost minimization problem. I first present a simple statistic model to gain intuition about the different options facing by the firm and workers and then detail the dynamic structural model that is estimated and used to simulate counterfactuals.

[27] Salop and Salop (1976) introduced the idea that firms can use compensation scheme as a self-selection device; Lazear (2000) provided empirical evidence that the implementation of a piece-rate incentive system was associated with an increase in the quality of newly hired workers; Lazear (2004) formalized the idea of self-selection and discussed how employees who believe themselves to be productive will expect to gain more under a performance-based pay.

[28] Recent papers by Hong et al. (2018), Guiteras and Jack (2018), and Balbuzanov et al. (2019) examined the response to incentives of workers in developing countries and focused on the tradeoff between output quality and quantity.



(a) The Firm's Problem

(b) The Worker's Problem

Figure 2: Static Model

4.1 Simple Static Model

4.1.1 The Firm's Problem

Figure 2a illustrates the firm's problem. The firm starts by facing demand level Q_0^{D} that is associated with C_0 labor costs under flat wages, as represented by equilibrium point a. Suppose the firm anticipates a demand increase that will last for a short period and lead to a new required production volume Q_1^{D}. With the goal of satisfying a new demand level, the firm faces three potential functions for its new labor costs that vary depending on the wage structure and the cost of recruiting and training new workers.

Consider a first scenario in which the new demand level can be satisfied with a small increase in the permanent labor force. If workers are hired under a flat-wage scheme, C_1^L represents the firm's new labor cost function. In another scenario, the firm needs to hire many permanent workers to meet the new demand level Q_1^{D}, and under a flat-wage scheme the firm faces a higher new labor-cost function represented by C_1^H. Under both scenarios, the firm could decide to establish a performance-pay schedule whereby workers are paid according to their measured productivity. In such a case, as productivity depends on the amount of effort a worker exerts, the firm faces a convex cost curve represented by $C(E)$ where E stands for effort. In addition to these alternatives, the firm could hire gig workers instead of permanent workers. Although gig workers are usually associated with lower hiring costs than permanent workers, this benefit comes at the cost of potential lower productivity. Moreover, as with permanent workers, a performance-pay compensation method could be offered to these workers.

Having these options, the firm needs to make the optimal decision and choose between the new potential equilibria b, c, or d. In the first scenario, it is optimal for the firm to be in b, in which the firm substitutes effort for employment by hiring new permanent or gig workers with flat-wage contracts. In the second scenario, the firm finds it optimal to choose c, because the increase in costs associated with a larger labor force outweighs the increase in workers' total pay when incentive-pay contracts are used. Therefore, the firm will substitute employment for effort by instituting a performance-pay wage structure.

In practice, the curvature of the effort-cost function is a key feature in determining the attractiveness of the equilibrium associated with the performance-pay option. As shown in Figure 2a, if the curvature of the effort-cost function is high, as represented by $C_{\text{high}}(E)$, the firm will no longer find it optimal to implement a performance-pay scheme because the flat wage alternative, represented by the equilibrium d, yields lower costs. If, however, the effort-cost function is low, as represented by $C_{\text{low}}(E)$, the performance-based scheme becomes even more attractive for the firm. These alternatives illustrate how fundamental the curvature factor is in determining the optimality of the new potential equilibria.

Additionally, although generous incentives may induce higher productivity, they result in higher labor costs. Meager incentives may not generate the desired increase in productivity. Thus, the optimal decision is the one that increases production while simultaneously keeping labor costs low.

4.1.2 The Worker's Problem

Workers are subject to a wage structure of a performance-pay with a minimum wage guarantee, as illustrated by Figure 2b, which presents a generalized form of this pay structure.[29] A worker must decide how much effort to exert under each compensation structure. Under a flat-wage scheme, a rational worker finds it optimal to exert the minimal effort level to produce the required output amount defined by $\underline{Y}(t)$, which means that a worker will maximize her utility at point a and earn a wage of w.[30]

Under performance-based pay, workers face the output production threshold Y_0, which is the production quantity after which the bonus system is applicable. If a worker produces above the threshold, she is compensated at performance rate β for every additional unit. If production does not exceed Y_0, the wage paid is equal to the flat-wage w. Any production level between $\underline{Y}(t)$ and Y_0 is not associated with financial benefits, which means that if a worker chooses to produce an amount in this range, the firm earns all the rent. Therefore, the worker's optimal effort decision under such incentive structure depends on her personal production capabilities. A worker that is capable of producing above Y_0 finds it optimal to be anywhere on the upward-sloping line – for example, point b – as long as the effort-cost associated with such production is smaller than

[29] Figure A.2 in the Appendix presents the exact wage structure, as it was offered during 2018.

[30] Note that the production-level threshold $\underline{Y}(t)$, reflects the demand variation the firm experiences through its dependence on demand time t, as in periods of higher expected demand, this threshold is set at a higher level.

the gained benefit.[31] A worker that is incapable of producing above Y_0 and chooses to produce anywhere between $\underline{Y}(t)$ and Y_0 has chosen an allegedly suboptimal effort level, because such effort levels are uncompensated. In practice, the data tells us that workers indeed choose to produce at such levels. This observation can be rationalized if workers are maximizing an objective that depends on non-pecuniary factors.

4.2 Model

4.2.1 The "New" Firm's Problem

With the goal of constructing a model for the labor decision of a firm that operates an on-demand customized production system, I depart from the 'traditional' firm's problem in several ways. In the neoclassical model of the labor market, the firm assumes that labor is hired as a factor of production, and it is put to work like capital, at a market-clearing wage and a rental rate. However, there is one major difference between labor and capital that is ignored by this assumption: The firm is free to use capital as it wishes; however, having hired a worker, it faces a considerable restriction on the worker's effort. Not only are there legal restrictions, but the firm must also usually obtain workers' cooperation to make the best use of them. This idea is even more pronounced when hiring a gig worker, as the interaction between her and the firm is usually short-term. For this reason, I consider workers' effort to be an input of production rather than treating all laborers as a homogeneous production factor.

Second, in the new firm's problem I redefine the extensive and intensive labor margins in a way that embodies worker effort. The "traditional" firm problem focuses on determining the size of their workforce. A more elaborate approach examines the labor-demand adjustment in the presence of a trade-off between the number of workers hired and the number of hours each employee works. This paper builds on this idea and analyzes the latent variable of effort as a new labor-intensive margin instead of hours worked.

When thinking about a performance-pay contract from the firm perspective, there is a tension between productivity and profitability. The main advantage of such contracts is that they not only improve labor productivity but also increase labor welfare. However, a major caveat is that firms maximize discounted profits, not productivity, and performance-contingent contracts may increase productivity but not profits. One of the main factors that could cause a negative relationship between productivity and profits in the face of performance-based contracts is quality reduction. Performance-based pay could motivate workers to speed up production by compromising quality. This issue is of particular concern when output is measurable but quality is not, and thus workers

[31] For example, assume that a worker produces Y_b, such that $Y_b > Y_0$. Also assume that the cost associated with this production level is $c(e_b)$, where $c(\cdot)$ represents the worker's effort-cost function, and e_b represents the effort exerted to produce Y_b, such that the cost function is increasing and convex: $c', c'' > 0$. The worker then finds it optimal to produce Y_b only if $w + \beta(Y_b(e_b) - Y_0) > c(e_b)$, which means that the production benefits outweigh the costs.

may increase measured production at the expense of unmeasured quality.[32] For this reason, output quality is integrated as a factor of production into the "new" firm's problem.

Additionally, the distribution of earning gains between the firm and its workers is a key aspect linking productivity and profitability. The firm is passing along some of the productivity gains to its workforce in the form of the additional compensation. The parameter that governs this allocation is the incentive rate (denoted by β in Figure 2b). In fact, a major challenge of constructing a performance-pay system lies in determining this rate. On the one hand, setting it too high could lead to a scenario in which workers receive most of the earnings gains or even obtain a share that is larger than the earnings increment. On the other hand, a rate set too low could lead to no production response at all. The incentive rate is incorporated directly into the firm's problem and an analysis that finds the optimal contract is presented in the counterfactual analysis.

Last, the new type of firm is characterized by unorthodox labor arrangements – specifically, the hiring of gig workers. This is a prevalent practice today among manufacturers that operate on-demand production lines to hire gig workers to perform a specific task, for a project, or for a season. Thus, I incorporate the labor force composition decision as a key aspect in the firm's problem.

4.2.2 Model Details

The firm faces a labor cost minimization problem in a calendar year. At the outset of each week t, subject to an output demand shock, the firm solves for the optimal number of permanent workers P_t, and gig workers G_t, in addition to the decision of which compensation method to apply: performance pay (PP) or flat wage (FW), denoted by $z \in \{FW, PP\}$. The pay-scheme type decision is binary –PP or FW – so that the firm's problem does not solve for the optimal incentive structure but rather takes the one chosen by the firm as given. The goal is thus to solve the problem and identify the parameters that govern the firm's labor and pay scheme decisions. This allows me to construct a reliable model that describes worker response to the existing incentives and isolate it from other factors such as the working environment and personal attributes. After I validate and estimate the model, I turn to computing the optimal incentive structure, an analysis that is presented in the counterfactual section.

Permanent Workers: Costs and Structure

Permanent workers are hired based on a long-term contract with an undetermined length. Thus, conditional on the number of permanent workers that were employed in the previous period P_{t-1}, the firm needs to make decisions regarding the number of new permanent workers it chooses to

[32] Freeman and Kleiner (2005) illustrated this concept in an empirical study and showed that the abolition of performance pay reduced productivity but increased profits, as quality rose in the absence of incentives. Holmstrom and Milgrom (1991) had a similar theoretical finding in the context of a multi-tasking model in which incentive contracts could cause agents to under- or over-invest sub-optimally in different tasks.

hire, denoted by P_t^N, as well as the number of permanent workers it chooses to lay off, denoted by P_t^L.

The number of workers per se is not enough to make a knowledgeable decision, as workers are heterogeneous in their experience, and worker productivity changes over time. Therefore, the number of workers is accompanied by information on each worker's tenure. However, as documented in Section 4, job experience has little effect on productivity for permanent workers because their productivity levels remained almost unchanged over time for in both total and high-quality production. Moreover, even if one accounts for job experience as a continuous measure, doing so is difficult due to the "curse of dimensionality." Hence, to make the problem tractable and aligns with the empirical evidence, permanent worker tenure is classified to be in one of the following three categories,

$$TC_t = \begin{cases} C_1 & \text{if} \quad X_t \leq 3 \\ C_2 & \text{if} \quad 3 < X_t \leq 30 \\ C_3 & \text{if} \quad 30 < X_t, \end{cases}$$

where X_t denotes a worker's experience in week t. In fact, TC_{t-1} and P_{t-1} define the problem's state variables and describe the information the firm needs at the beginning of time t to make an informed decision.

The firm's labor force composition decision is made under the assumption that layoffs are independent of tenure. This idea is conceptualized by randomly choosing P_t^L out of the pool of existing permanent workers, P_{t-1}, so that $\boldsymbol{P_t^L}$ reflects the sum of workers laid off from all categories. Combining all these components, the law of motion that defines the number of permanent workers at time t looks as follows:

$$P_t = (1-\mu)P_{t-1} - \boldsymbol{P_t^L} + P_t^N,$$

where μ represents a natural separation rate of permanent workers, according to which a permanent worker might terminate her job for exogenous reasons.

The costs associated with permanent workers include both recruiting and layoff costs. Let R^P be the recruiting costs of permanent workers and L^P the costs associated with their job termination. Hence, the total costs associated with this group of workers are equal to

$$R^P \cdot P_t^N + L^P \cdot \boldsymbol{P_t^L}.$$

Gig Workers: Costs and Structure

Gig workers are hired for a predetermined short-term period and therefore layoff decisions and costs associated with job termination are irrelevant. I assume that gig workers are selected from an existing pool of temporary workers at the beginning of every week, denoted by t. In practice, I assume that the number of gig workers the firm chooses to hire is determined as a residual component, so that the firm would successfully fulfill its demand within a given week.

Accordingly, the job experience of gig workers evolves weekly and depends on the workers' actual days of experience in the firm, denoted by X_{it}. I further assume that gig workers' contract length is one week, so that the number of new gig workers at week t, G_t^N defines the total number of gig workers at this week, G_t, that is, $G_t = NG_t$. Define R^G to be the recruiting costs of gig workers, so that the labor costs associated with recruiting this group of workers is defined by

$$R^G \cdot G_t^N.$$

Wage Costs

Under a flat-wage scheme, wages are fixed and independent of worker effort while under a performance-pay scheme, wage is a function of the effort level that maximizes workers' private payoff. Worker production is endogenously incorporated into the firm's problem through the incentive compatibility constraint

$$\begin{aligned} Y(E_{it}^{*z}) &= \arg\max_E (U_{it}|z) \\ &= \arg\max_E (W_{it} - C_{it}|z), \end{aligned}$$

which stipulates that worker i in week t chooses an optimal effort E_{it}^{*z} to maximize utility U_{it} conditional on the pay scheme type: $z \in \{PP, FW\}$. A worker's utility is equal to her total wage gains W_{it} net the effort costs C_{it} associated with this wage. In practice, by constructing the incentive compatibility constraint in such a way, I capture the idea that the firm has a comprehensive understanding of how worker production responds to monetary incentives.

Worker i's wage at time t is given by $W_{it}(E_{it}^*, z)$, which is a function of the optimal effort a worker exerts and the pay method $z \in \{FW, PP\}$. The total wage payments made by the firm at time t are a sum of wages made to permanent workers P_t and gig workers G_t, defined as:

$$\sum_{i=1}^{P_t} \mathbb{E}\left[W_{it}(E^*, z|TC_{it})\right] + \sum_{i=1}^{G_t} \mathbb{E}\left[W_{it}(E^*, z|X_{it})\right],$$

where expectations are taken with respect to workers' idiosyncratic shock (a detailed explanation of the workers' problem is subsequently presented). The main difference between gig workers and permanent workers is their tenure structure in the model. The wage function reflects this feature, as the permanent workers' wage depends on their tenure category, TC_{it}, while the gig workers' wage evolves weekly and depends on their actual day of experience, X_{it}.

Total Labor Costs

By combining these labor cost components, one can assemble the objective function the firm wishes to minimize:

$$\text{Cost}_t(P_t, TC_t) = \sum_{i=1}^{P_t} \mathbb{E}\left[W_{it}(E^*, z|TC_{it})\right] + \sum_{i=1}^{G_t} \mathbb{E}\left[W_{it}(E^*, z|X_{it})\right] + \left(R^P \cdot P_t^N + L^P \cdot \boldsymbol{P_t^L}\right) + \left(R^G \cdot G_t^N\right).$$

Product Demand

On-demand customized production systems operate under a rigid production schedule. I mimic such strict production policy by assuming that the firm must satisfy the demand it experiences within a week. Furthermore, to be as close as possible to the real setting in which firms use past product demand and rely on seasonal patterns to generate demand forecasts, I use the demand observed in 2015 as a benchmark level and add a fixed-percentage increase to reflect the average product demand predicted for 2018. Specifically, I define $D_t(\zeta_t)$ to be the weekly product demand where ζ_t represents weekly product demand uncertainty shock, such that $\zeta_t \sim \mathcal{N}(0, \sigma_D)$. These shocks are assumed to be serially independent where σ_D is derived from the data observed in the year 2015.

Aggregating the information presented thus far, the following constraint guarantees that the varying weekly product demand is equal to workers' total production

$$\sum_{i=1}^{P_t} \mathbb{E}\left[Y_{it}(E^*, z|TR_{it})\right] + \sum_{i=1}^{G_t} \mathbb{E}\left[Y_{it}(E^*, z|X_{it})\right] = D_t(\zeta_t),$$

where $Y_{it}(E^*, z)$ denotes the production level of worker i at time t and pay scheme z when exerting the optimal effort level.

The Firm's Dynamic Problem

The firm faces a dynamic stochastic discrete choice problem in which it seeks to minimize labor costs. At each week t, subject to an output demand shock, the firm minimizes the present discounted value of remaining lifetime labor costs with respect to the optimal number of new permanent workers, P_t^N, the total number of permanent workers that are being laid off, P_t^L, in addition to the decision of which compensation method to apply, $z \in \{FW, PP\}$. The firm also decides on the total number of new gig workers that are being hired, G_t^N; however this decision will not be deliberate and instead, the number of gig workers will be chosen as a residual term to satisfy product demand.

In any period t, the firm faces $K(t)$ mutually exclusive alternatives, defined by the vector (P_t^N, P_t^L, z), where K varies over time with the number of permanent workers at the previous period. Define the indicator function $d_k(t) = 1$ if alternative k is chosen by the firm at time t, and $d_k(t) = 0$ otherwise, such that $\sum_k d_k(t) = 1$. Further define Ω_t to be the state space at t that consists of all the relevant factors affecting current or future labor cost – that is, the exhaustive set of all possible number of permanent workers P_{t-1} and their possible combinations of tenure categories TC_{t-1}.

The minimized present discounted value of lifetime labor costs at t, which defines the value function of the problem, is given by

$$V_t(\Omega_t) = \max d_k(t)\, \mathbb{E}\left\{ \sum_{\tau=t}^{T} \psi^{\tau-t}\text{-Cost}_t \middle| \Omega_t \right\},$$

where the problem starts from initial week $t = 1$ and ends at week T, which is the terminal decision period for that year. The expectation is taken over product demand shock, and the employees productivity shock. The solution to the firm's optimization problem is a set of decision rules that relates the optimal choice at any t from among the feasible set of alternatives to the elements of the state space at t. Recasting the problem in a dynamic programming framework, the value function can be written as the maximum over alternative-specific value functions, $V_t^k(\Omega_t)$ – i.e., the expected discounted value of alternative $k \in K(t)$, which satisfies the Bellman equation, namely

$$V_t(\Omega_t) = \max_{k \in K(t)} \left[V_t^k(\Omega_t) \right], \tag{1}$$

$$V_t^k(\Omega_t) = \begin{cases} -\text{Cost}_t^k(\Omega_t) + \psi\, \mathbb{E}\left[V_{t+1}(\Omega_{t+1}) \,|d_k(t) = 1, \Omega_t \right] & \text{for } t < T, \\ -\text{Cost}_T^k(\Omega_T) & \text{for } t = T, \end{cases}$$

such that

$$\begin{aligned}
&P_t = (1-\mu)P_{t-1} - \boldsymbol{P_t^L} + P_t^N \\
&P_t^L \leq (1-\mu)P_{t-1} \\
&G_t = G_t^N \\
&Y(E_{it}^{*z}) = \arg\max_E (U_{it} - C_{it}|z, \varepsilon_{id}) \\
&\sum_{i=1}^{P_t} \mathbb{E}\left[Y_{it}(E^*, z|TC_{it}) \right] + \sum_{i=1}^{G_t} \mathbb{E}\left[Y_{it}(E^*, z|X_{it}) \right] = D_t(\zeta_t) \\
&\zeta_t \sim \mathcal{N}(0, \sigma_D).
\end{aligned}$$

4.2.3 The Worker's Effort Decision

Workers solve for the optimal effort level when subject to a wage structure of a performance pay with a minimum wage guarantee.[33] The problem accounts for incentive gains as well as features other than monetary gains that may affect a worker's optimal effort level, including intrinsic motivation, total factor productivity, and a measure that represents workload tightness.[34]

4.2.4 Model Details

On each given day, d, a risk-neutral worker, i of type ν, such that $\nu \in \{g, p\}$ representing gig and permanent types, wishes to maximize her utility and decide on the effort she exerts, E_{id}. The decision over the optimal effort depends on the days of experience a worker has accumulated in the job thus far, X_{id}, which evolves daily: $X_{id} = X_{id-1} + 1$. Also, the worker solves the problem given the pay scheme offered by the firm and a daily idiosyncratic productivity shock. The optimal effort is a latent variable that is not directly observed but rather inferred through the model's solution. As such, in order for the problem to generate valuable information, I restrict effort to lie between 1 and 10.

Utility

The worker's utility is defined as follows

$$U(E_{id}, X_{id}) = W_\nu(E_{id}, X_{id}) - C_\nu(E_{id}, X_{id}),$$

where $W_\nu(E_{id}, X_{id})$ denotes the wage a worker receives, and $C_\nu(E_{id}, X_{id})$ denotes the effort cost associated with this wage. The workers wage structure takes the following form:

$$W = \max\Big\{w, w + \beta(Y_\nu^{\mathrm{HQ}} - Y_0)\Big\},$$

such that

$$\begin{aligned} &\beta = 0 && \text{if wages are flat} \\ &\beta \in \mathcal{R}_+ && \text{if wages are based on performance,} \end{aligned}$$

where Y_0 is the incentive regime's production threshold, β is the incentive coefficient, and Y_ν^{HQ} is the total of high-quality items produced by worker i on day d.[35]

[33] Figure A.2 in the Appendix presents the exact wage structure, as it was offered by the firm during the year 2018, and Figure 2b provides a generalized form of this pay structure.

[34] Throughout this paper daily, workload tightness defined as the product demand divided by the number of workers in a given day, as presented in Figure A.4 in the Appendix

[35] All of these notations match the description in Figure 2b.

Production Function

Production on a given day is a function of effort and experience $Y_\nu^{\text{Total}} = f(E_{id}, X_{id})$, such that $f_1, f_2 > 0$, signifying that more experienced workers need effort to achieve a given level of output. As a parametric assumption, the worker's production technology takes the following Cobb-Douglas form:

$$Y_\nu^{\text{Total}} = \alpha_\nu E_{id} X_{id}^{\delta} \mathrm{e}^{\varepsilon_{id}}, \tag{2}$$

with $\delta > 0$ denoting the elasticity of experience, $\alpha_\nu > 0$ denoting the worker-type specific total factor productivity, and σ_ε represents an *iid* worker-day specific production shock, which is assumed to be normally distributed: $\varepsilon_{id} \sim N(0, \sigma_\varepsilon)$. The experience elasticity is constant across individuals and days and measures the output responsiveness to a change in the level of experience in production, ceteris paribus. The total factor productivity parameter varies by worker type (gig or permanent) and is introduced to measure production efficiency differences and account for the part of the variation in production across workers that is not explained by experience, effort, or personal production shock. In particular, this parameter is constructed to control for differences between worker types.

Production Quality

Define $\bar{\rho}(E_{id}, X_{id})$ to be the probability of worker i, of type ν, on day d to generate low-quality items out of her total production.[36] I parameterized this probability to be a function of the worker's effort and experience as follows:[37]

$$\bar{\rho}_{id}(E, X) = \frac{\exp\left(\phi_E E_{id} + \phi_X X_{id}\right)}{1 + \exp\left(\phi_E E_{id} + \phi_X X_{id}\right)}.$$

The sign and magnitude of ϕ_E and ϕ_X quantify the relationship between experience and effort on output quality. On the one hand, one would assume that the more experienced a worker, the less low-quality output she produces, so that $\rho_X < 0$. On the other hand, it may be that a byproduct of higher effort is a rapid and less-accurate production so that, $\rho_E > 0$. Having defined this probability, the total high-quality production of worker i of type ν on day d is defined as

$$Y_\nu^{\text{HQ}} = \Big[1 - \bar{\rho}(E_{id}, X_{id})\Big] Y_\nu^{\text{Total}}.$$

[36] Items of low quality are those that do not pass the quality assurance check and thus must undergo a fixing process to be sold. Under performance-based pay, workers are rewarded only for high-quality output and thus quality plays a crucial role in the worker problem.

[37] The *logit* transformation keeps the probability in the range between zero and one. I further multiply it by Pr(low quality), which is equal to the average of low-quality production observed in the data across all workers and periods in order for the probability to take values in the proper and relevant range.

Effort Cost

The effort cost function takes the following parametric form:

$$C_\nu(E_{id}, X_{id}) = \frac{\kappa_d}{X_{id}} E_{id}^{\gamma_\nu} - \eta_\nu E_{id}. \tag{3}$$

The first term of the function represents the direct effort cost a worker incurs, which decreases with experience and increases with effort. Further, it increases with the workload tightness level, κ_d, which is defined by the distribution of workload among workers as follows:[38]

$$\kappa_d = \frac{\text{Product Demand}_d}{\text{\# of Workers}_d}.$$

The effort-cost curvature parameter γ_ν determines the degree of convexity of the effort cost function for each type of worker with the restriction of $\gamma_\nu > 1$.

The second term of the effort cost function represents self-compensation gained from effort. The parameter η_ν can be interpreted as workers' intrinsic motivation, with the important restriction of $\eta_\nu > 0$ (further discussed). This parameter captures the idea that individuals exert higher effort (or refrain from shirking) not only because they get paid, but for non-pecuniary reasons, such as they are task-oriented, have less job burn-out, value job flexibility, or simply because they like to undertake certain activities.[39]

The Worker's Daily Effort Problem

To conclude, the worker's problem can be written as

$$\max_{E\in[1,10]} \left[\max\left\{w, w + \beta(Y_\nu^{\text{HQ}}(E_{id}, X_{id}) - Y_0)\right\}\right] - \left(\frac{\kappa_d}{X_{id}} E_{id}^{\gamma_\nu} - \eta_\nu E_{id}\right) \tag{4}$$

s.t.

$$Y_\nu^{\text{Total}}(E_{id}, X_{id}) = \alpha_\nu E_{id} X_{id}^{\delta} \mathrm{e}^{\varepsilon_{id}}$$
$$Y_\nu^{\text{HQ}}(E_{id}, X_{id}) = \left[1 - \rho(E_{id}, X_{id})\right] Y_\nu^{\text{Total}}$$
$$\mathbb{E}\left[Y_\nu^{\text{HQ}}\right] \geq \underline{Y}(d),$$

where the last requirement states that expected production is as least as high as the minimum required periodic production level for d being a day in week t.

[38] This parameter aims to capture the production intensity at the firm on a given day. Figure A.4 in the Appendix provides a visual representation of κ_d for the year 2018.

[39] Psychologists introduced intrinsic motivation and argued that individuals undertake many activities without expecting an extrinsic reward (De Charms, 2013; Deci, 1971). Amabile (1993) introduced workers' intrinsic motivation into models of labor supply. Deci et al. (2017) offered a review of studies that examined employees' intrinsic and extrinsic motivations in a workplace, and several recent papers have studied wages and incentive schemes when workers are intrinsically motivated (Benabou & Tirole, 2003; Besley & Ghatak, 2005; Francois, 2000; Glazer, 2004).

4.3 Model Predictions and Implications

4.3.1 Effort Cost Function

First, consider the worker's effort in the flat-wage scenario. Given that workers are guaranteed to receive their pay regardless of productivity level in this scenario and that production entails a cost, we might reasonably expect that a utility-maximizing agent finds it optimal to choose $E_{id} = 0$. The effort-cost production function is constructed to preclude this possibility by setting $\eta_\nu > 0$.

To derive this condition directly, consider the following marginal cost of effort

$$C_{\nu,E} = \frac{\kappa_d}{X_{id}} \gamma_\nu E_{id}^{\gamma\nu - 1} - \eta_\nu.$$

If a worker decides to not exert effort, the marginal cost of effort then becomes negative, as shown by

$$C_{\nu,E}\Big|_{E=0} = -\eta_\nu.$$

Thus, by restricting the personal motivation parameter to be greater than zero, I guarantee that workers supply positive effort levels under the fixed-wage schedule.

Several other features of the effort-cost function are that it increases with effort $C_{\nu,E} > 0$ and is convex with respect to effort $C_{\nu,EE} > 0$. These properties capture the idea that while little effort is not very costly, higher effort is associated with increased effort cost. Additionally, the parametric assumption of the effort cost function implies that (all else held equal) more experienced workers face a lower marginal effort-cost, $C_{\nu,EX} < 0$.

4.3.2 Optimal Effort

The worker's objective function is strictly concave with respect to E_{id}.[40] This property guarantees that the worker's problem yields a unique maximum. Under each pay scheme, the worker's effort solution takes a different form, as I next describe.

Flat Wage: $\beta = 0$

If the firm sets $\beta = 0$, earnings are then fixed and independent of performance. Solving for the optimal-effort level in this scenario, denoted by $E_{id}^{*\text{FW}}$, yields the following solution:

$$\frac{\kappa_d}{X_{id}} \gamma \left(E_{id}^{*\text{FW}}\right)^{(\gamma_\nu - 1)} = \eta_\nu. \tag{5}$$

This equation illustrates the idea that it is not a monetary incentive that leads workers to exert effort to produce above the minimum required level under a flat wage scheme, but rather intrinsic

[40]A proof of this property presented in Section B in the Appendix.

motivation. Specifically, (5) shows that on day d, worker i of type ν will maximize her utility by choosing the effort level that will equate her personal motivation to the marginal effort cost given her experience, effort-cost curvature, and the workload on that given day. Specifically, the optimal effort decreases with daily workload and increases with experience and intrinsic motivation.[41] Also, the optimal effort is decreasing with the effort-cost function curvature, a feature that reflects the idea that a higher curvature is associated with higher effort.

Performance Pay: $\beta > 0$

Under a performance-pay regime, workers choose their optimal effort $E_{id}^{*\mathrm{PP}}$ to equate the marginal benefit and the marginal costs of production, considering both monetary and non-monetary incentives. This idea is clearly illustrated when examining the first-order condition of the problem in the performance-pay scenario,

$$\beta\Big[1-\rho\Big]\alpha X_{id}^{\delta}\mathrm{e}^{\varepsilon_{id}}\mathbb{I}_{\mathrm{PP}} + \eta_{\nu} = \frac{\kappa_d}{X_{id}}\gamma\Big(E_{id}^{*\mathrm{PP}}\Big)^{(\gamma_\nu - 1)} + \beta\rho_E\alpha_\nu\Big(E_{id}^{*\mathrm{PP}}\Big)X_{id}^{\delta}\mathrm{e}^{\varepsilon_{id}}\mathbb{I}_{\mathrm{PP}}, \tag{6}$$

where ρ and ρ_E represent the probability and marginal probability of low-quality production, respectively. The indicator function $\mathbb{I}_{\mathrm{PP}}$ equals one if a worker reaches the incentive threshold such that $Y_{\nu}^{HQ} > Y_0$.

Consider first the scenario by which $\mathbb{I}_{\mathrm{PP}} = 1$. In such case, the left side of (6) represents the sum of the marginal monetary benefits of producing above $\underline{Y}(t)$, as represented by the product of the performance-rate β and the marginal productivity of high-quality output, as well as the non-monetary benefits embodied by the intrinsic motivation, η_ν. The right side of (6) represents the sum of the marginal cost of exerting effort and additionally the marginal cost of producing low-quality items above $\underline{Y}(t)$. If $\mathbb{I}_{\mathrm{PP}} = 0$, the optimal effort solution under the performance-pay and flat-wage schemes then coincide and the discussion relevant to (5) applies under this scenario.

4.3.3 Participation Constraint

Gig workers are a temporary labor force leveraged by the firm during times of high demand. Gig jobs are characterized by frequent transition – the most salient feature. As such, it is reasonable to assume that gig worker participation constraints are binding in order to reflect their indifference regard to the current job and alternative outside options. Knowing that firm plant locations in non-central areas have low job variety and that the nature of the examined job is a low-skill type with no education requirement, I extend this assumption to take effect on permanent workers as well.

Denote workers' alternative utility by $\underline{u}$ and define the participation constraint based on the

[41] These features directly support the empirical evidence presented in Section 6 and thus strengthen the model parametric assumption.

expected utility under flat wage as follows:

$$\mathbb{E}\left[U(E_{id}^{*\mathrm{FW}}, X_{id})\right] \geq \underline{u}.$$

By writing this condition explicitly when the constraints is binding, one can identify the guaranteed wage level w,

$$w = \underline{u} + \mathbb{E}\left[C_{\nu}^{*\mathrm{FW}}\right], \tag{7}$$

such that C_{ν}^{*FW} denotes the effort cost under optimal effort in the fixed-wage payment schedule in which the expectation is taken with respect to all workers and days.

4.3.4 Indirect Utility

The indirect utility function under the flat-wage scheme takes the form of $V^{*\mathrm{FW}} = w - \mathbb{E}\left[C_{\nu}^{*\mathrm{FW}}\right]$ and using the results obtained in (7), I derive the following result

$$V^{*\mathrm{FW}} = \underline{u}, \tag{8}$$

which shows that workers' maximum attainable utility under a sflat wage is equal to their outside option value. The indirect utility function under the performance-pay scheme is defined as follows:

$$V^{*\mathrm{PP}} = w + \beta\,\mathbb{E}\left[Y_{\nu}^{*\mathrm{HQ,PP}} - Y_0\right]\mathbb{I}_{\mathrm{PP}} - \mathbb{E}\left[C_{\nu}^{*\mathrm{PP}}\right].$$

If $\mathbb{I}_{\mathrm{PP}} = 1$, this equation can be simplified using the results in (7). Specifically, the indirect utility function under the performance-pay scheme can be written as:

$$V^{*\mathrm{PP}} = \underline{u} + \beta\,\mathbb{E}\left[Y_{\nu}^{*\mathrm{HQ,PP}} - Y_0\right] - \mathbb{E}\left[C_{\nu}^{*\mathrm{PP}} - C_{\nu}^{*\mathrm{FW}}\right], \tag{9}$$

where $Y_{\nu}^{*\mathrm{HQ,PP}}$ denotes the high-quality output generated under the optimal effort decision, and $C_{\nu}^{*\mathrm{PP}}$ denotes the effort cost under optimal effort in the bonus performance-pay schedule.

The illuminating result in (9) has several implications. First, it shows that under performance-based pay, workers' maximal attainable utility is equal to the monetary benefit of producing above Y_0, minus the cost for effort above what they would otherwise choose under the flat-wage scheme (on top of the outside option value, $\underline{u}$). Second, based on (9), one can infer that the maximal attainable utility decreases with the daily workload tightness κ_d, and increases with the personal motivation η_{ν}, with incentive rate β, and with worker's job experience X_{id}. Lastly, a worker chooses to put an effort level greater than that associated with her optimal decision under the flat-wage scheme only if the financial benefit of doing so exceeds the effort costs associated with this benefit, that is,

$$\beta\,\mathbb{E}\left[Y_\nu^{*\mathrm{HQ,PP}} - Y_0\right] > \mathbb{E}\left[C_\nu^{*\mathrm{PP}} - C_\nu^{*\mathrm{FW}}\right], \tag{10}$$

and in such a case, workers are strictly better off under the performance-pay scheme: $V^{*\mathrm{PP}} > V^{*\mathrm{FW}}$. If production falls below the incentive threshold so that $\mathbb{I}_{\mathrm{PP}} = 0$, then the indirect utility under performance-pay is then given by

$$V^{*\mathrm{PP}} = \underline{u} - \mathbb{E}\left[C_\nu^{*\mathrm{PP}} - C_\nu^{*\mathrm{FW}}\right]. \tag{11}$$

Under such a scenario the second term of (11) must be equal to zero for the worker to accept this job, a result that becomes apparent when comparing workers' maximal attainable utility under the flat wage scheme, as presented in (8).

4.3.5 Expected Earnings

The expected earnings in the performance-pay scenario are higher than the expected earnings in the flat-wage scenario if a worker is capable of reaching the performance regime threshold Y_0. From the worker's perspective, the earnings when incentives are available should compensate for the additional cost burden incurred from producing above $\underline{Y}(t)$. In practice, the firm earns all the rent over production levels between $\underline{Y}(t)$ and Y_0, and workers are directly compensated only for production above Y_0. This idea can be viewed in the following equation, which presents workers' expected earning performance-based pay,

$$\mathbb{E}\left(W_\nu | z = \mathrm{PP}\right) = w + \beta\,\mathbb{E}\left[Y_\nu^{*\mathrm{HQ,PP}} - Y_0\right], \tag{12}$$

where the expectation is taken with respect to the idiosyncratic production shocks. Thus, workers find it beneficial to produce above $\underline{Y}(t)$ when the expected earnings in (12) exceed expected production cost $\mathbb{E}\left(C_\nu^{*\mathrm{PP}}\right)$, an outcome that depends on a worker's intrinsic motivation.

5 Identification

The following are the sets of parameters in both the firm's and worker's problems

$$\Theta_\omega = (\alpha_p, \alpha_g, \delta, \gamma_p, \gamma_g, \eta_p, \eta_g, \phi_E, \phi_X, \sigma_\varepsilon)$$
$$\Theta_F = \left(\mu, R^P, L^P, R^G\right),$$

where ω stands for workers and F stands for firm. The sufficient condition for identification of these parameters is one-to-one mapping between them and a subset of moment restrictions generated from the data (of the same dimension). As neither the firm nor the worker models yield an analytical close-formed solution, I discuss the sources of variations that ensure the model's parameters are data identified in the following section.

5.1 The Worker's Problem Parameters

Under the flat wage or when $\mathbb{I}_{PP} = 0$, the first-order condition yields the following solution for the optimal effort

$$E^*_{id} = \left(\frac{\eta_\nu X_{id}}{\gamma_\nu \kappa_d}\right)^{\frac{1}{\gamma_\nu - 1}}.$$

This solution can be substituted into the production function and by applying logarithmic transformation, one obtains the log production of workers

$$\log\left(Y^{*FW}_\nu\right) = \left[\log(\alpha_\nu) + \left(\frac{1}{\gamma_\nu - 1}\right)\log\left(\frac{\eta_\nu}{\gamma_\nu}\right)\right] - \left(\frac{1}{\gamma_\nu - 1}\right)\log\left(\kappa_d\right) + \left(\delta + \frac{1}{\gamma_\nu - 1}\right)\log(X_{id}) + \varepsilon_{id}. \tag{13}$$

From (13) one can see that variation in X_{id} identifies the experience elasticity δ and that the variation in the workload parameter κ_d identifies the curvature of the effort-cost function.

The probability for low-quality output can be written it in the following odds ration form

$$\frac{\mathbb{P}(\text{Low Quality})_{id}}{\mathbb{P}(\text{High Quality})_{id}} = \exp\left(\phi_E E^*_{id} + \phi_X X_{id}\right). \tag{14}$$

By applying a logarithmic transformation and plugging in the optimal effort under a flat wage scheme (or when $\mathbb{I}_{PP} = 0$) I derive the following result

$$\log\left(\frac{\mathbb{P}(\text{Low Quality})}{\mathbb{P}(\text{High Quality})}\right)_{id} = \phi_E\left(\frac{\eta_\nu}{\gamma_\nu}\right)^{\frac{1}{\gamma_\nu - 1}}\left(\frac{X_{id}}{\kappa_d}\right)^{\frac{1}{\gamma_\nu - 1}} + \phi_X X_{id}. \tag{15}$$

The expression in (15) exemplifies additional variation used for identification.

Now consider the first-order condition when $\mathbb{I}_{PP} = 1$ and $\beta > 0$:

$$\begin{aligned}\beta&\left[\frac{1}{1 + \exp\left(\phi_E E^*_{id} + \phi_X X_{id}\right)}\right]\alpha_\nu X^\delta_{id} e^{\varepsilon_{id}} + \eta_\nu \\ &= \frac{\kappa_d}{X_{id}}\gamma_\nu\left(E^*_{id}\right)^{(\gamma_\nu - 1)} + \beta\left[\frac{\phi_E \exp\left(\phi_E E^*_{id} + \phi_X X_{id}\right)}{\left(1 + \exp\left(\phi_E E^*_{id} + \phi_X X_{id}\right)\right)^2}\right]\alpha_\nu\left(E^*_{id}\right)X^\delta_{id} e^{\varepsilon_{id}},\end{aligned} \tag{16}$$

a result that can be further simplified using (14). Because the optimal effort E^* is a unique solution of the problem, having solved for it one can substitute it into the production function to obtain the estimated production of workers i of type ν at day d, defined by:

$$Y^{*\text{PP}}_\nu = \alpha_\nu E^{*\text{PP}}_{id} X^\delta_{id} e^{\varepsilon_{id}}. \tag{17}$$

In practice, the estimation procedure relies on the model's parameters to match the estimated production levels with those observed in the real data, as defined in (13), (16) and (17). The estimation procedure also finds parameters that match the low-quality production quantity, and the performance-incentive wage gain as defined in (12) and (14).

5.2 Parameters of the Firm's Problem

The costs of hiring permanent and gig workers and laying off permanent workers depends on several actions taken by the firm, such as increasing the human resources department, including the searching costs of suitable workers, and conducting recruiting sessions and interviews. These costs are not directly observed and their quantification is not straightforward for the firm's management, as well as for a researcher. I use the structural model framework to evaluate these costs by using observed firm behavior.

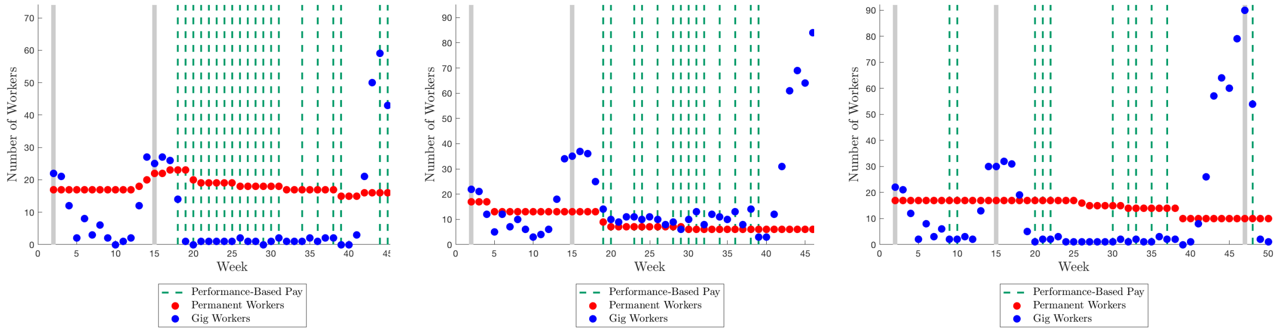
To understand what information in the data would enable identification of these costs, it is useful to consider an illustrative model. Consider a single period (myopic) decision regarding the hiring composition and whether to implement a performance-based pay scheme. For simplicity, assume that in the prior period the firm had one permanent worker paid by a flat wage scheme and that demand is fixed and known. The firm then faces two alternatives related to the labor force composition. In the first, the firm lays off the permanent worker, an action that entails L^P costs, and hires gig worker(s) instead, a process that involves R^G costs. In the second alternative, the firm stays with the one permanent worker from the previous period and thus has only wage costs to consider. The firm also faces a decision of whether to implement a performance-based pay ,$z = PP$, or remain with a flat wage pay ,$z = FW$. These decision are dependent on the per production unit wage costs associated with each type of worker and pay scheme, which I define as $W_{\nu,z}/Y_{\nu,z}$, for $\nu \in \{G, P\}$. Under this scenario, for a given pay scheme z, the firm chooses to change the labor force composition if and only if $L^P + R^G + W_{G,z}/Y_{G,z} \leq W_{P,z}/Y_{P,z}$.[42]

Now, suppose the firm expects an increase in demand. The firm's alternatives are then between hiring gig workers or permanent workers (under a given pay scheme). In such a scenario, the firm chooses to change the labor force composition if and only if $R^G + W_{G,z}/Y_{G,z} \leq R^P + W_{P,z}/Y_{P,z}$.

As these simple examples indicate, knowledge of R^G,R^P and L^P are necessary to forecast the optimal labor input decisions of the firm. Variation in the opportunity per production unit-wage costs, defined by workers' wages and productivity for a given pay scheme, serve as a substitute for direct labor costs. Coupling this information with the product-demand fluctuations during the year and changes in the pay schemes types provide the sources of variation for identification. Hence, to obtain separate estimates of the firm's model parameters, it is both necessary and sufficient that the firm's hiring choices vary across time and pay scheme types.

Figure 3 illustrates the solution of the firm's problem under varying observed assumptions on the labor costs R^G,R^P, and L^P for a natural separation rate of 1% for all considered scenarios. In Figure 3a the cost of hiring gig and permanent workers is equal and lower than the cost of laying off permanent workers ($R^G = R^P < L^P$). Under this settings, the firm avoids layoffs of permanent workers (which are subject to natural separation rate) and hires very few gig workers in times of low product demand. The firm also engages in massive hiring of gig workers around times of anticipated product-demand increases and finds it optimal to offer a performance-based

[42]The firm may be required to hire more than one gig worker to fulfill the demand forecast.



(a) $R^G = R^P < L^P$ (b) $L^P < R^G < R^P$ (c) $R^G < L^P < R^P$

Figure 3: Simulations of the Firm's Problem Solution

pay schemes at times when product demand is low and stable and around times of massive gig worker hiring.

Figure 3b demonstrates a different scenario by which $L^P < R^G < R^P$. In this case, the firm gradually lays off most of its permanent labor force and retains only a few permanent workers to provide a production safety net. Moreover, the practice of hiring gig workers is employed throughout the year, with low hiring around times of low-anticipated product demand and massive hiring around times of product demand surges. Interestingly, the firm finds it optimal to implement incentive pay only at times of low-anticipated product demand as if to guarantee that product demand is fulfilled while keeping the costs of compositional changes of workforce low.

Figure 3c illustrates the firm's optimal behavior when faced with high layoff costs of permanent workers such that that $R^G < L^P < R^P$. In this scenario, the firm avoids laying off permanent workers and also avoids hiring gig workers.

6 Descriptive Regression Results

In this section I examine the evolution of production patterns and response to incentives over job experience in the firm separately for gig and permanent workers. Based on the findings I establish five facts that lay the groundwork for assumptions made in constructing the structural model.

The estimation model in Equation (18) seeks to explore the relationship between productivity of worker i at day d: Y_{id}, with the working environment at day d, represented by the daily demand, daily number of workers and the pay scheme type, as well as worker i's attributes, including worker type (permanent or gig), job experience, and gender:

$$\begin{aligned}\log(Y_{id}) = \gamma_0 &+ \beta_{\text{Gig}}\text{Gig}_i + \beta_{\text{PP}}\text{PP}_d + \beta_{\text{F}}\text{Female}_i + \beta_{\text{Exp.}}\text{Exp.}_{id} + \beta_{\text{Exp.}^2}\text{Exp.}^2_{id} \\ &+ \beta_{\text{Exp.}\times\text{Gig}}\left(\text{Exp.} \times \text{Gig}\right)_{id} + \beta_{\text{Exp.}^2\times\text{Gig}}\left(\text{Exp.}^2 \times \text{Gig}\right)_{id} \\ &+ \beta_{\text{Exp.}\times\text{PP}}\left(\text{Exp.} \times \text{PP}\right)_{id} + \beta_{\text{Exp.}^2\times\text{PP}}\left(\text{Exp.}^2 \times \text{PP}\right)_{id} \\ &+ \beta_{\text{PP}\times\text{Gig}}\left(\text{PP} \times \text{Gig}\right)_{id} + \beta_{\text{PP}\times\text{Exp.}\times\text{Gig}}\left(\text{PP} \times \text{Exp.} \times \text{Gig}\right)_{id} + \beta_{\text{PP}\times\text{Exp.}^2\times\text{Gig}}\left(\text{PP} \times \text{Exp.}^2 \times \text{Gig}\right)_{id} \\ &+ \alpha_1\text{Demand}_d + \alpha_2\text{Demand}^2_d + \beta_1\left(\text{ \# of Workers}\right)_d + \beta_2\left(\text{ \# of Workers}\right)^2_d \\ &+ \psi_i + C_{id} + \epsilon_{id}.\end{aligned} \tag{18}$$

The variable Gig_i denotes an indicator for whether worker i is a gig worker, PP_i denotes an indicator for days in which the incentive performance-pay system was instituted, and Exp_{id} represents the number of days of job experience worker i has accumulated up to day d. The model incorporates worker experience both linearly and as a square term in a three-way interaction pattern to capture how the relationship between worker type and pay scheme differs along varying levels of worker job experience. Female_i takes the value of 1 if worker i is a female and 0 otherwise. Demand_d represents the daily product demand measured by the number of orders placed online in day d and $(\text{ \# of Workers})_d$ denotes the number of workers present in day d. ψ_i are individual fixed effects that summarize the impact of permanent differences among individuals in observed and unobserved characteristics. The vector C_{id} consists of observed characteristics of individuals and days, both time-varying and time-constant, which include a dummy variable that denotes whether worker i has worked in several departments within a plant and a day-of-the-week fixed effect. Finally, the error term ϵ_{id} is assumed to have constant variance and is uncorrelated across individuals and days.

The results associated with this model are presented in Table 2. Whereas the model in (18) takes the most comprehensive form, the results shown in this table present various specifications of this model.

Fact 1: Worker type and working environment are crucial factors in determining production.

The coefficient of the dummy variable for gig workers as shown in Column (1) in Table 2, indicates that on average gig workers are 16% less productive than permanent workers, ceteris paribus. This estimate implies that worker type plays a crucial rule in determining productivity; however, it does not address the effect of experience on productivity differences between worker types. As gig workers and permanent workers are hired for fundamentally different time periods, overlooking the differential experience effect could yield misleading estimates.

The work environment variable estimates – i.e., product demand and number of workers – presents interesting findings. Worker productivity responds positively to an increase in product demand; however, the increasing productivity response to higher product-demand levels is dimin-

Table 2: Gig Workers, Incentives, and Production

	Dependent Variable: Log of Daily Worker's Production (Adj)					
	(1)	(2)	(3)	(4)	(5)	(6)
Gig Worker=1	-0.162**	-0.292**		-0.264**	-0.284**	
	(-6.72)	(-11.87)		(-10.65)	(-11.21)	
Performance Pay=1				0.131**	-0.257*	0.0815**
				(7.76)	(-1.95)	(2.39)
Performance Pay=1 × Gig Worker=1					0.617**	0.0679*
					(4.27)	(1.75)
Gig Worker=1 × Experience		0.326**	0.496**	0.279**	0.313**	0.375**
		(14.97)	(16.30)	(12.40)	(12.84)	(13.45)
Gig Worker=1 × Experience × Experience		-0.0496**	-0.0969**	-0.0413**	-0.0484**	-0.0861**
		(-7.00)	(-7.85)	(-5.79)	(-6.54)	(-10.76)
Female=1	0.0979**	0.0577**		0.0642**	0.0679**	
	(6.81)	(4.01)		(4.48)	(4.72)	
Experience		0.00194	0.00638	0.00182	0.00185	-0.0239**
		(0.78)	(0.73)	(0.74)	(0.73)	(-2.95)
Experience × Experience		0.0000733	-0.000983**	0.0000712	0.0000530	-0.0000590
		(1.28)	(-2.99)	(1.25)	(0.88)	(-0.17)
Product Demand	0.0596**	0.0664**	0.0376**	0.0858**	0.0825**	0.0442**
	(7.38)	(8.50)	(5.35)	(10.50)	(10.05)	(5.53)
Product Demand × Product Demand	-0.00152**	-0.00233**	-0.000743	-0.00340**	-0.00326**	-0.000830
	(-2.70)	(-4.25)	(-1.53)	(-6.05)	(-5.79)	(-1.55)
Number of Workers	-0.00554**	-0.00654**	-0.00593**	-0.00913**	-0.00876**	-0.00860**
	(-4.47)	(-5.46)	(-5.71)	(-7.37)	(-7.01)	(-7.37)
Number of Workers × Number of Workers	0.0000257**	0.0000238**	0.0000355**	0.0000382**	0.0000350**	0.0000553**
	(2.35)	(2.25)	(3.88)	(3.57)	(3.27)	(5.46)
Constant	4.574**	4.580**	4.526**	4.592**	4.591**	4.649**
	(123.70)	(132.12)	(131.47)	(132.90)	(131.88)	(126.88)
Day of the week FE	Yes	Yes	Yes	Yes	Yes	No
Individual FE	No	No	Yes	No	No	Yes
N	6832	6832	6832	6832	6832	5830
R^2	0.101	0.161	0.451	0.168	0.172	0.392

Notes: : The table reports estimates for worker log production using model (18). The models include as controls holiday dummy, repeated employment dummy, and employment in several departments dummy. Product demand is measured in thousands, and experience is measured in 30 days. Since the model is estimated over a logarithmic transformation of the total production, the estimates represent production percentage change. * $p < 0.10$,** $p < 0.05$

ishing.[43] Workforce-size estimates show that worker average production decreases as the number of workers increases also at a diminishing rate, a result that implies that the production environment is not intrinsically competitive.[44] A comparison of these work environment coefficients across the different models shows that their magnitude and sign remain similar and significant under all specifications. This observation indicates that such factors are crucial for determining worker's productivity, and thus I incorporate them directly in the worker's structural effort problem.

Fact 2: Job experience is a crucial factor in determining production for gig workers, not for permanent workers.

Column (2) in Table 2 captures the effect of worker job proficiency by incorporating multiplicative terms of experience and worker type. The multiplicative terms make interpretation of the results non-trivial, as one cannot look at the coefficient of the gig-worker dummy in isolation. Figure 4a facilitates interpretation by plotting the results (in unit terms).

The results presented in Figure 4a indicate distinct production patterns of permanent workers and gig workers, as the role of experience varies dramatically by type. Gig workers hold an increasing production pattern as their tenure progresses. Specifically, the production quantities of gig workers with low experience indicate that they arrive to the job with a relative disadvantage with respect to their job skills. The results further indicate that gig worker learn the required job skills quickly, and after one months of experience their average production is higher than the average production of permanent workers.

Column (3) in Table 2 expands the analysis by adding individuals' fixed effect that controls for workers' (observed and unobserved) heterogeneity.[45] Figure 4b demonstrates that while the relationship between production and job experience has been preserved for gig workers, the same relationship among permanent workers is in fact decreasing, especially at higher levels of experience. Interestingly, the figure also indicates that under this individual-specific analysis, gig workers are more productive than permanent workers for all (observed) experience levels.

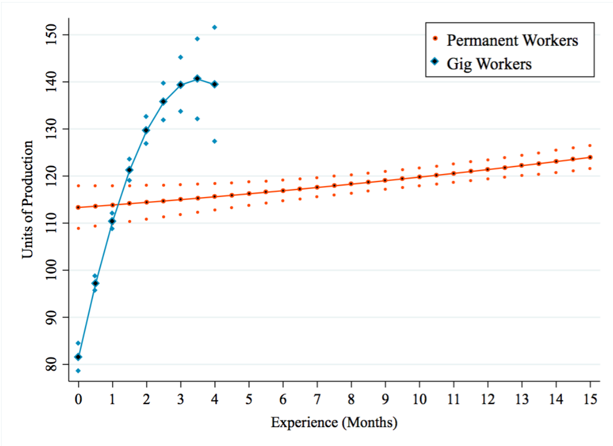
Fact 3: The stringency of hiring standards is different between gig and permanent workers.

Figure 4a indicates that permanent workers are on average significantly more productive than gig workers immediately after joining the firm, all else being equal. The explanation for this observation lies in the different hiring processes of these workers. Since permanent workers are

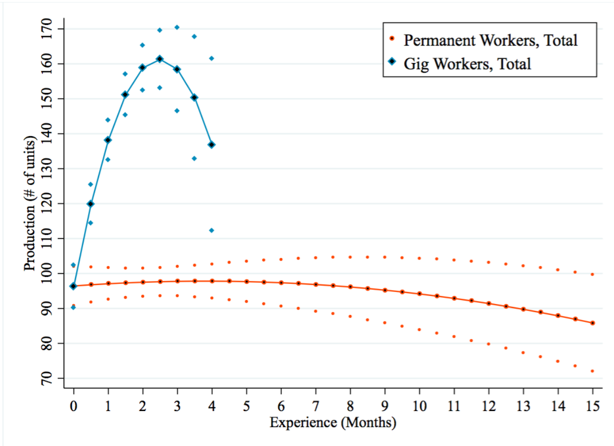
[43] To illustrate, increasing daily demand by 1,000 units increases production by 6.4% on average, whereas an increase of 5,000 units in product demand would lead to a lower productivity increase of 4.6%.

[44] To illustrate, increasing the labor force size in a given day by 10 workers (gig workers or non-gig workers) reduces worker productivity by 6.5% on average while increasing the workforce by 50 workers is associated with an average productivity decrease of 3.6%.

[45] This modeling approach comes at the expense of adding individual-level attributes (particularly gig-worker dummy) that are absorbed into the individual-specific factor.

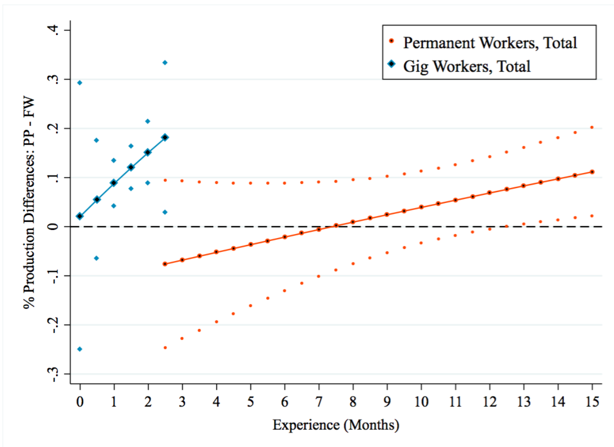


(a) Production, Without Ind. FE



(b) Production, With Ind. FE

(c) Pay Incentive Response, Without Ind. FE



(d) Pay Incentive Response, With Ind. FE

Figure 4: Production and Incentive Response by Experience and Worker Type

Notes: All figures are associated with the results presented in Table 2 derived under different specifications of regression model in Eq. 18. The points in Figures 4a and 4b are associated with Columns (2) and (3) of Table 2 respectively, and represent the predicated total production of workers for various experience levels by worker types and its 95% confidence intervals. The points in each Figures 4c and 4d are associated with Columns (5) and (6) Table 2 respectively, and represent the percentage difference in production under flat wage relative to bonus incentive pay regimes for various experience levels by worker types and its 95% confidence intervals.

hired on a long-term contract and their relationship with the firm is binding, the firm conducts a rigorous screening process and thoroughly examines the job fit. Gig workers, in contrast, are hired to "fill the gap" with less emphasis on job match and a less-stringent selection process. Even if recruiters want to hire only the most able and suitable gig workers, leniency and compromise are inevitable because many are hired around high-demand times.

Fact 4: Gig workers are more responsive to performance-based incentives than permanent workers.

Thus far, I have ignored the influence of incentives on productivity in the analyses. Column (4) of Table 2 starts by examining the aggregate response to incentives by controlling for a performance-pay dummy variable that equals one for days that the incentive pay schedule is in place. The identification strategy relies on the change in the payment scheme as exogenous, because its implementation timing was unanticipated and revealed to workers only several days before it occurred. Moreover, the change in the pay scheme was short lived, as the performance-pay system was in place for only a short period and thus sorting and selection effects that occur when more productive workers self-select into the firm are irrelevant.[46] Lastly, the change in the pay structure was orthogonal to the management method, a fact that further assists in overcoming potential endogeneity concerns.

The coefficient of the performance pay dummy variable in Column (4) of Table 2 indicates that incentives induced workers to increase production by 13% on average, a result that stands in line with similar estimates in the literature. In order to identify the distinct production response of workers to incentive pay by type and additionally disentangle the role of experience in explaining this relationship, Column (5) of Table 2 incorporates interactions between worker type, pay schedule, and job experience. To ease interpretation, Figure 4c visually illustrates the percentage change in total production under performance pay (PP) relative to flat wage (FW) by worker type along various job experience levels. The estimates indicate a striking difference between the way gig workers and permanent workers respond to the incentive pay scheme. Specifically, gig workers with only few days on the job produce on average as much as 28% more under a performance-based pay scheme than under a flat wage scheme. This large productivity response decreases to around 20% after 30 days of experience and to 8% after 60 days of experience. At the same time, among permanent workers the average production response to incentives is not statistically different than zero for job experience of 10 months or less. At approximately 12 months of job experience – the average time a permanent worker spends in the firm – the production response to incentives stands at 10%. This declining response to incentives at higher levels of job experience could be interpreted that the power of incentives on worker's productivity reduces over time. In other words, high production levels are not sustainable and eventually workers revert back to

[46] The performance pay system was in place for three weeks during 2018, which serves as the main data source for all of the analyses presented in the paper. During 2015, a performance-pay system was in place for only two weeks, and this year's data is used in the structural model validation.

productions level associated with a flat wag.[47]

Since the productivity of many workers was observed under both flat wage and performance pay systems, I continue by incorporating individual fixed effects into the model. The results, presented in Column (6) of Table 2 and Figure 4d, provide interesting findings. For permanent workers, both the analysis that captures the average effect and the analysis that examines an individual-specific production response yield similar results. Nevertheless, the key difference that emerges from a comparison of these estimation procedures is the higher standard errors associated with the individual-fixed-effect model that make the production response to incentives of permanent workers not statistically different than zero for all job experience levels.

For gig workers, a comparison between the average and individual-specific responses to performance-based pay is illuminating. In contrast to the average effect, which is shown to mostly monotonically decrease, the worker-specific response to incentives presents an increasing patten. Specifically, the production response to incentive pay increases with job experience and reaches an approximate 15% higher production under performance-based pay relative to flat pay at 60 days of job experience.

Fact 5: Gig and permanent workers hold different intrinsic behavioral motives and job perception.

The findings thus far give rise to a conceptual difference in the job perceptions of permanent and gig workers and demonstrate an intrinsic behavioral difference between them. In particular, the results illustrate gig workers as holding high motivation to fulfill their hired purpose, with a desire to exploit the most possible gain from the temporary position. At the same time, the results imply that permanent workers seek stability and thus remain at sustainable production levels. The notion of personal motivation can explain these differences, whereby gig workers exert higher effort (or refrain from shirking) not only because they get paid, but also for non-pecuniary reasons, such as being task-oriented, having a desire for independence, or simply because they like to undertake certain activities.

The large productivity response to incentive pay scheme of gig workers can be further explained in two other ways. First, it could be a result of the *expectations anchoring* effect. Many of the gig workers joined the firm near the end of 2018 for the Christmas season, knowing that the firm could initiate the incentive pay system soon. This knowledge created expectations and a behavioral-conditioning effect whereby the workers needed the incentive to produce more than the lowest productivity level required. That is, the incentive structure echoed the fact that production above the expected lower bound required an uncompensated effort, and that gig workers would receive the same pay for producing below or above the bonus threshold. Second, it could also be explained by the marginal gains from the gig workers' incentives that are conceptually different

[47] The declining incentive response pattern could be attributed to a novelty effect. Perhaps the new pay system led to higher productivity at first because it drew worker interest and attention, as a new technology does, but over time, the effect gradually diminishes.

from those of the permanent workers. In particular, when considering the expected job duration, the marginal wage gain of \$1 for a gig worker who works a total of two weeks is dramatically higher than that of a permanent worker who has worked at the firm for a year. Understanding that every additional unit of production could lead to a large relative increase in wages is thus a potential channel that could motivate gig workers to increase production.

7 Estimation Methods

7.1 Indirect Inference

I use an indirect inference to estimate the structural parameters of the worker's problem (Gourieroux, Monfort, & Renault, 1993). In practice, I focus on the empirical findings presented in Section 4 in addition to models over the performance-incentive wage gains and daily quantities of defective production as auxiliary models to tightly link the structural parameters to the empirical findings.[48] This methodology is chosen over others, because instead of selecting a set of moments (as is typically done with the simulated moments method), it captures particular aspects of the data identified as relevant and meaningful.

As such, the idea is to repeat simulations to find the data-generating parameters that yield (on average) regression estimates equal to the actual estimates obtained from the data. First, I solve for the worker's daily effort decision for a vector of possible values (a guess) of structural parameters given the realization of the idiosyncratic production shock, and a set of daily and personal observables. In the second step, based on the effort solution and the set of initial parameters, I calculate the optimal production, low-quality production, and wages for each worker. This step generates a simulated data set that corresponds with the initial set of parameters. I then use the simulated data to estimate the auxiliary model coefficients and obtain a vector of auxiliary parameters, $\psi_{\text{sim}}(\Theta_W)$. The choice of the auxiliary parameters allows for rather transparent identification of the model's structural parameters.

The optimal set of parameters $\hat{\Theta}_W$ is the one that minimizes the distance between auxiliary parameters estimated on the actual data and auxiliary parameters estimated on the simulated data,

$$\hat{\Theta}_W = \arg\min_{\Theta} \Big(\hat{\psi}_{\text{data}} - \psi_{\text{sim}}(\Theta)\Big) \mathrm{W} \Big(\hat{\psi}_{\text{data}} - \psi_{\text{sim}}(\Theta)\Big)',$$

where W is a symmetric and positive semi-definite weighting matrix.[49]

[48] These models are guided by the identification assumptions previously outlined.

[49] W is set to be equal to the inverse of the variance-covariance matrices of the auxiliary models. Cross models covariances are set equal to zero.

7.2 Simulated Maximum Likelihood

7.2.1 Solution Method

Given that the solution of the firm's optimization problem is not analytic, I solve the model numerically by applying a simulated maximum likelihood procedure.[50] The solution consists of the values of $\mathbb{E}(V_{t+1}(\Omega_{t+1})|d_k(t) = 1, \Omega_t))$ for all k and elements of Ω_t (hereafter, *Emax*), in which I use backward recursion beginning with the last decision period.[51]

Two complications arise in solving the model numerically. First, in any period, the firm faces multiple choices as defined by all the possible combinations of the decision vector $\{P_t^N, P_t^L, z_t\}$. This choice set is composed of $2 \times (1 + \max P_t)$ alternatives wherein the first term represents the choice whether to implement a performance-based pay scheme and the second reflects the decision of whether or how many permanent workers the firm decides to hire or fire. For example, if the firm limits the number of hired permanent workers to 35, then there are 72 possible choices. Solving the dynamic-multinomial choice problem for each choice creates a significant computational burden.

Additional computational issues arise from the state-space size. Specifically, the size of the state space makes a full solution of the problem computationally intractable because the *Emax* functions must be calculated for all state values at each t. This is because in order to keep track of the workers' weekly tenure category, it is necessary to keep track of the complete sequence of tenure paths, and thus the state space is composed of all the possible number of permanent workers and the possible combinations of tenure categories. The state space of the problem, although finite, is huge, as it grows exponentially with the number of state variables. This notion makes a full solution of the dynamic-programming problem infeasible, leaving aside the iterative process necessary for estimation. One way to reduce the size of the state space in a way consistent with the data (and the evidence established in the stylized facts) is to limit the tenure categories. I do so by considering three tenure categories of permanent workers. Although limiting attention to only three categories eases the computational burden, the dimensionality problem remains.[52] To further solve this issue, I adopt the approximation and interpolation method developed in Keane and Wolpin (1994, 1997, 2001) in which the *Emax* functions are evaluated at a random subset of state points, and the values are used to fit a global polynomial approximation in the state variables.[53]

[50] Keane, Todd, and Wolpin (2011) provided a survey of structural estimation methods of a discrete choice dynamic-programming model in a multinomial choice setting including the simulated maximum likelihood procedure.

[51] The *Emax* functions appear on the right-hand side of (1).

[52] To illustrate, when considering 20 permanent workers, the number of possible tenure categories sequences is 209.

[53] As in Keane and Wolpin (1994) I performed a Monte Carlo integration over the ζ to calculate the expected value of the maximum of the alternative-specific value functions at those state points. I used 1,500 state points for the estimation of the *Emax* approximations and 50 draws for the numerical integrations. I tested several other alternatives and the *Emax* approximations did not appear to be sensitive to change in these parameters. I used a polynomial regression of the fifth order in the *Emax* approximation. The R-squares were above 0.99 in all model periods.

7.2.2 Estimation Method

The solution to the firm's minimization problem serves as input into estimating the model's parameters. In the decision model previously presented, the observed outcomes at each period include: the pay scheme, the number of permanent workers that are hired, and the number of laid-off permanent workers. Let the outcome vector at t be denoted by $d_k(t)$, so that the likelihood of these observed outcomes is defined by

$$\mathcal{L}(\Theta_F) = \prod_{t=1}^{T} \mathcal{P}\left(d_k(T), d_k(T-1), ..., d_k(1) \middle| \Omega_1\right),$$

where Ω_1 is the initial state space at the beginning of the year that consists only of the number of permanent workers in the previous period and the time these workers spent in the firm at that period, and Θ_F is the vector of the firm's problem parameters. Given the assumption of serial independence of the demand shock across periods, I can write the likelihood as a product of outcome probabilities conditional on the corresponding state space. I use the kernel-smoothed frequency simulator proposed McFadden (1989), namely,

$$\hat{\mathcal{P}}\left(d_k(t) = 1 | \Omega_t, \zeta_t, \varepsilon_t\right) = \frac{1}{S} \sum_{s=1}^{S} \frac{\exp\left\{\left[V_t^k - \max\left(V_t^j\right)\right] / \tau\right\}}{\sum_i \exp\left\{\left[V_t^i - \max\left(V_t^j\right)\right] / \tau\right\}},$$

where the j superscript denotes the vector of value functions over all alternatives, and τ is the smoothing parameter.[54,55]

The set of the firm's model parameters enters the likelihood through the choice probabilities computed from the dynamic-programming problem solution. The parameters of the workers problem and particularly the parameters that govern the performance-based pay, enter through the incentive-comparability constraint. The estimation procedure iterates between the solution of the dynamic program and the calculation of the likelihood.

[54] Ideally, I would have liked to calculate the probability of $d_k(t) = 1$, by which the firm is observed to choose alternative k at week t numerically by randomly taking $1, ..., S$ draws from the distribution of the product demand shock ζ, and worker's production shock ε, and determine the fraction of times that the value function for that alternative is the largest among all feasible alternatives. However, there is a practical problem in implementing this approach, because even for large S, the likelihood is not smooth in the parameters. See Keane and Wolpin (1997), Eckstein and Wolpin (1999), and Todd and Wolpin (2006) for further applications.

[55] I set the smoothing parameter equal to 700, which provided sufficient smoothing given the magnitudes of the value functions.

8 Results and Model Fit

8.1 The Worker's Problem

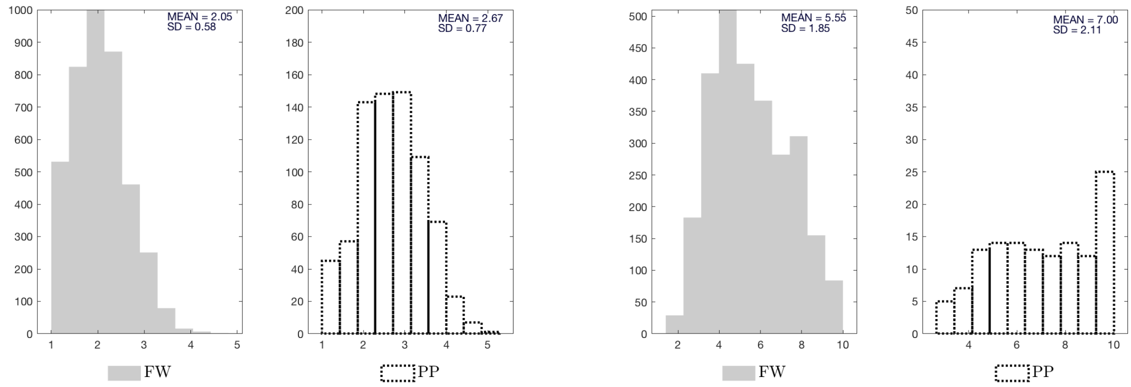
8.1.1 Estimation Results

Table 3 presents estimates of the worker's structural model. The results reveal the sources governing workers' productivity and show that while the personal motivation and total factor productivity of gig workers is more than double than that of permanent workers, their effort- cost curvature is higher than that of permanent workers. To illustrate the interactions between these estimates, consider the model equations. When examining workers' effort-cost function in (3), the effort-cost curvature estimate indicates that gig workers' cost function is more convex, meaning that for the same effort level (all other things equal), gig workers' experience higher effort costs than permanent workers. However, gig worker have higher personal motivation that compensates for their higher exertion of effort. Such tradeoffs are coupled with the advantage permanent workers hold due to their higher experience levels that result in lower-effort costs due to the way the effort function is constructed. In this context, the experience-elasticity estimate, as shown in the production function in (2), reveals that permanent workers have a salient production advantage relative to gig workers resulting from their high experience levels. Thus, taking together the parameters' estimates, one can conclude that for an average gig worker and a permanent worker who wish to reach the same production level (for a given a production shock), gig workers' higher total factor productivity offsets their dis-advantage resulting from their low experience levels.

Table 3: Labor Supply: Estimates of Structural Parameters

Parameters Descriptions	Symbol	Estimate
Total Factor Productivity		
Gig Worker	α_g	32.92
Permanent Workers	α_p	12.02
Personal Motivation		
Gig Worker	η_g	140.18
Permanent Worker	η_p	60.83
Effort Cost Convexity		
Gig Worker	γ_g	4.01
Permanent Worker	γ_p	3.21
Experience Elasticity	δ	0.089
Effort Effect on Low-quality Production	ϕ_E	0.71
Experience Effect on Low-quality Production	ϕ_X	0.22
Variance of Idiosyncratic Production Shock	σ_ε	0.365

The latent factor of effort also demonstrates the source of production differences between gig and permanent workers. Figure 5 present histograms of the optimal effort solution the model



(a) Gig Workers

(b) Permanent Workers

Figure 5: Histograms of the Optimal Effort Solution

yields according to workers' type and pay scheme. For permanent workers with a flat wage, the optimal effort level is highly variable and concentrated strongly in the middle levels with an average of 5.55. This result is attributable to permanent workers that are hired throughout the year, and thus, work during periods with both a low and high workload. For the same workers, the optimal effort with performance pay indicates that they exploit the bonus pay system and benefit from it, as their effort levels are higher and largely concentrated at the highest level of 10 units. Unlike permanent workers, gig workers' optimal effort with performance pay is concentrated at low levels with an average of 2.05, and the switch to performance-based pay shifts their effort to higher levels with an average of 2.67 and a greater variance. Given that gig workers are hired during times of demand surges, these patterns support the notion of expectation anchoring that was proposed in Section 4 (Fact 5), by which the bonus incentives reduce gig workers' production with a flat wage because considering the possibility of pay incentives echoes the fact that production above the minimum amount required is associated with uncompensated effort.

The estimate in Table 3 reveals the value and sign of the parameters that govern production quality. The results show that a higher effort a worker exerts, the higher low-quality item she will produce. This result can be interpreted to mean that higher effort causes a rush through the production process with less attention to production quality. The results also show that workers with higher experience tend to produce more low-quality items, although this relationship is weaker than the effort-quality link.

8.1.2 Model Fit

Table 4 reports production averages generated by the model next to their empirical counterparts. As evident from these moments, the model performs well in predicting workers' productivity along different dimensions of worker types and pay scheme. Specifically, the total average of production,

Table 4: Labor Supply: Production Average in the Data and Model

	Data	Model Solution
$\bar{Y}^{z}_{i,\nu,d}$	107.41	107.18
$\bar{Y}^{z}_{i,\nu=p,d}$	116.38	115.61
$\bar{Y}^{z}_{i,\nu=g,d}$	102.23	102.31
$\bar{Y}^{z=PP}_{i,\nu,d}$	119.67	125.62
$\bar{Y}^{z=FW}_{i,\nu,d}$	105.87	104.87
$\bar{Y}^{z=PP}_{i,\nu=p,d}$	125.87	121.23
$\bar{Y}^{z=PP}_{i,\nu=g,d}$	118.60	126.38
$\bar{Y}^{z=FW}_{i,\nu=p,d}$	115.94	115.36
$\bar{Y}^{z=FW}_{i,\nu=g,d}$	99.33	98.07

as well as the production averages under a flat wage (both total and the ones calculated separately by worker types) is shown to be almost identical. The model slightly overestimates the production averages under performance pay. This can as can be seen, for example, in the total production under this pay scheme in which there is a difference of 7 items between the model and the data. The breakdown of production averages separately for gig and permanent workers under the bonus pay reveal that over-prediction originated in the gig workers group. That is, according to the data, the average daily production of this group is 118 items while the model predicts an average production of 128 items. Given that the production variance under this pay regime is equal to 49, and considering that the predictions could range from 30 to 250, these average differences appear to be negligible.

To further assess the model's validity, I compare the production distribution as generated by the model solution. The results, presented in Figure 6, compare the densities generated by the data with those predicted by the model's solution. Visual inspection of the results show that through the lens of the model, workers' production quantities match not only production averages, but also the distribution of production along the possible range of production quantities. This is compelling evidence of the strength of the estimated parameters and the model construction.

8.2 The Firm's Problem

8.2.1 Estimation Results

Table 5 presents estimates of the labor-demand side structural parameters. The results indicate that hiring costs of gig workers are lower than the hiring and layoff costs associated with permanent workers, and that the recruiting costs of permanent workers are higher than their layoff costs. Allegedly, based on these estimates, one could infer that it is more profitable to hire permanent workers; however, such an assessment overlooks several important aspects of the firm's decision.

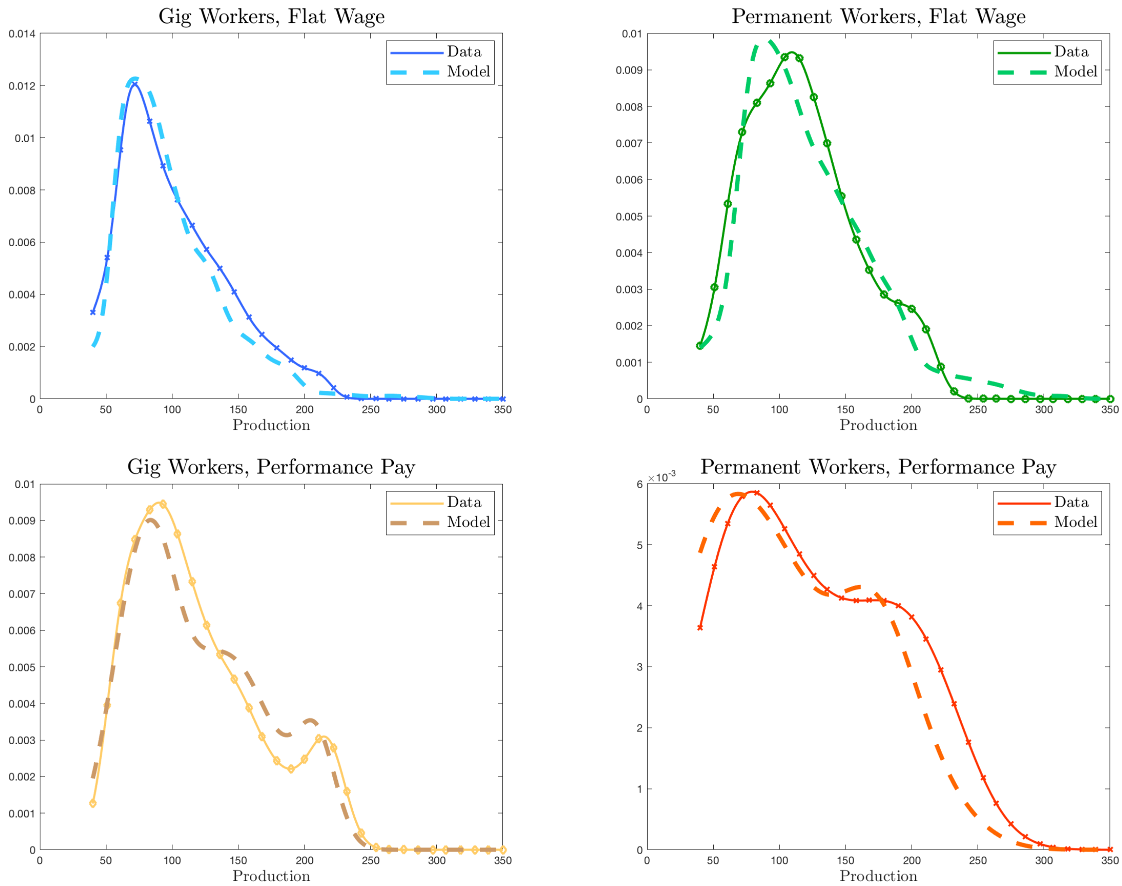


Figure 6: Model Fit: Production Densities in the Data and in the Model

To demonstrate these aspects, consider an average gig and permanent workers. Based on the average number of days each worker is working in the firm I estimate that

$$\text{Daily hiring cost(Permanent)} < \text{Daily hiring cost(Gig)},$$

and at the same time,

$$\text{Daily Pay(Permanent)}_z \geq \text{Daily Pay(Gig)}_z,$$

for $z = \{FW, PP\}$. The pay difference is larger than the daily hiring cost difference, and thus taking the abovementioned inequalities together, the seeming conclusion is that hiring gig workers would be more profitable for the firm. Yet, this comparison ignores the fact that permanent workers are more productive than gig workers (under both pay schemes) and more importantly, the firm is operating under strict production timelines and product demand uncertainty. This "simple" comparison demonstrates the complexity of the firm's problem. In the following section I conduct a thorough analysis that takes into account all the tradeoffs in the firm's problem and explore its optimal behavior given all the results obtained thus far.

Table 5: Labor Demand: Estimates of Structural Parameters

Parameters Description	Symbol	Estimate
Separation rate	μ	0.067
Recruiting permanent worker	R^P	4296.37
Laying off permanent worker	L^P	138.46
Recruiting gig worker	R^G	318.42

9 Counterfactuals

The structural equilibrium-framework estimation establishes a comprehensive understanding of the choices made by both the workers and the firm, and thus serves as a solid benchmark to characterize their optimal behavior beyond the given setup. Knowing that the firm experimented with the performance-based pay parameters during my observation period, as well as with the time periods implementing the incentive pay, I do not impose that those choices were made optimally when estimating the model. Rather, I investigate whether the firm can improve its current choices by performing simulations based on the estimated model in which I characterize equilibrium choices made by workers and the firm under different scenarios.

9.1 Original Incentives Structure

The objective of this analysis is to solve for the optimal labor-force composition schedule throughout the year that will minimize the firm's labor costs under two assumptions. First, I maintain the same incentive structure parameters as offered by the firm in practice, as it is presented in Figure 7a. Second, different than the original settings by which the firm offered this incentive pay structure – i.e., only for short period of several weeks – I assume that the incentive pay was offered more frequently throughout the year, especially around times of expected demand surges. In particular, the dashed green vertical lines in Figure 7b indicate the weeks the incentive pay scheme was offered.

Interestingly, although these assumptions depart only slightly from the original settings, letting the firm solve for the optimal labor-force composition yields a reduction of more than 8% in labor costs. Figure 7b shows the optimal workforce composition and hiring schedule given this setup, and a positive correlation is apparent between the implementation of the incentive pay scheme and the hiring of gig workers. Moreover, the number of gig workers the firm hires is higher in weeks with higher expected product demand. These findings reflect the obtained results, according to which the hiring costs of gig workers are dramatically lower than those of permanent workers and more importantly, gig workers increase their productivity significantly in response to the performance-based incentives pay.

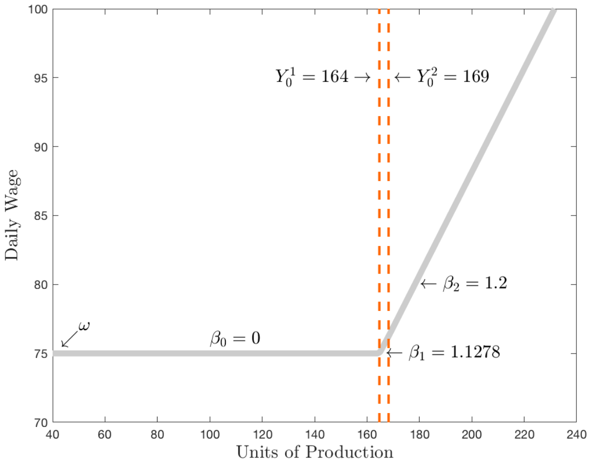
Additionally, the optimal workforce schedule in Figure 7b indicates that the number of permanent workers increases between the beginning of the year until the second peak season (around Mother's Day) while it decreases gradually afterwards. This outcome results from the constraint that enables the firm to implement the incentive pay structure only around seasons with high expected demand. The firm increased the number of permanent workers in response to the increase in product demand around Mother's Day, which actually starts several weeks before the actual holiday. After this period, there was a sharp decrease in product demand, which then remained stable and at a relatively low level. Therefore, the firm did not need to hire additional permanent or gig workers, and permanent workers left mostly due to the natural separation rate. Also, since the firm's value function internalizes the future value associated with permanent workers, this outcome could reflect the positive relationship between defective production and job experience, a result obtained in the worker's problem.

9.2 Profit Maximizing Approach

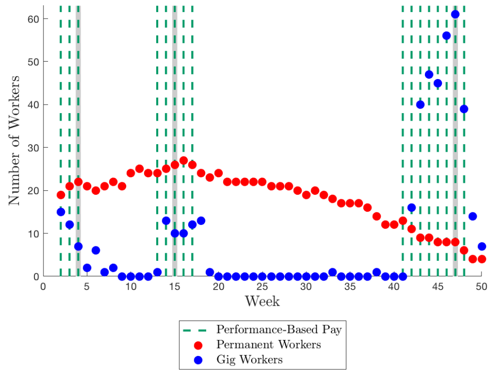
Next, I depart from the original set of incentives offered by the firm. Specifically, I optimize the firm's labor-cost minimization problem over the set of parameters that define the pay structure. Figure 7c presents the optimal pay scheme that resulted from this analysis. It is apparent that the obtained optimal set of parameters and the original set of parameters offered by the firm exhibit different features. First, in the original incentive pay scheme, the threshold that defines the entry of the performance-based pay scheme is located at a high production level. A second threshold that is associated with higher piece rate is set only a few of production units above the entry level. In the obtained optimal performance-based pay structure, however, the thresholds are located at two different edges of the production distribution. The lower-optimal threshold is in fact immaterial, as it is set at the production level defined as the minimum production required to remain employed, and the higher threshold is set at a quantity slightly higher than the original one.

Additionally, there is a clear difference in the piece rates offered in the optimal pay structure and in the original pay offered. In the original pay, there is a slight increase in the piece rates between the two thresholds, which is consistent with their proximity, whereas the optimal pay structure exhibits a sharp increase in the piece rates between the two thresholds. In particular, the first piece rate is low and offers workers a small monetary gain for every unit of production, which is equal to approximately 5% daily wage gains for an average worker. The second piece rate is set to be higher than the initial piece rate; furthermore it is significantly higher than the one originally offered. Lastly, the original and optimal pay structures also differ in their wage benchmark, such that the optimal pay structure defines workers' minimum wage at a lower level relative to the original pay structure. This reduction in the daily wage, which worsens workers' initial wage conditions, is slightly offset by offering the piece rate starting at a low level of production.

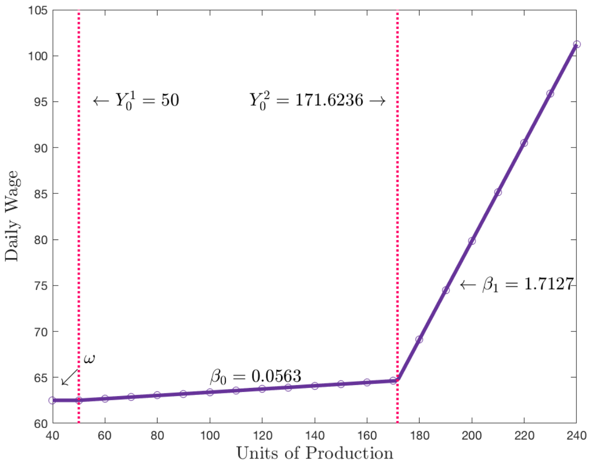
Figure 7d presents the optimal workforce schedule and pay-scheme choices throughout the year given the optimal pay structure. One can see that the firm finds it optimal to implement the



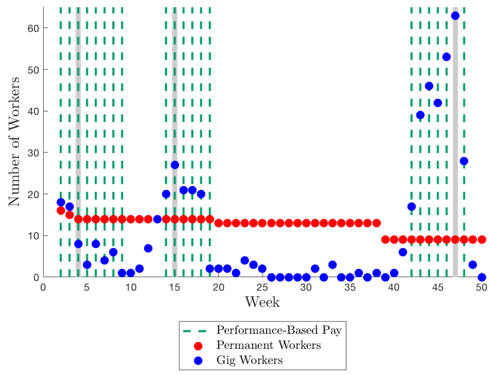
(a) Original Incentive Pay Structure



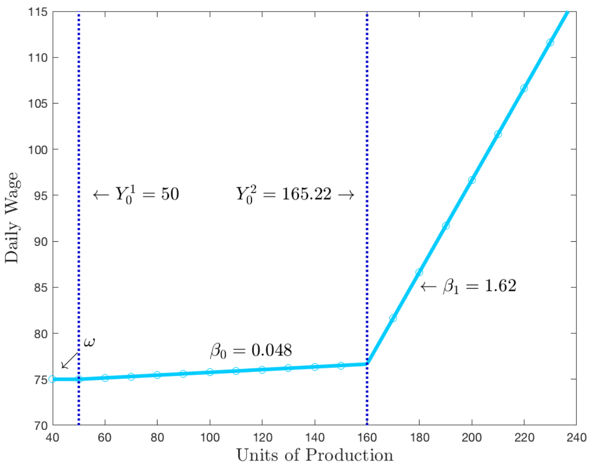
(b) Optimal Workforce Schedule given the Original Pay structure around Peak Seasons



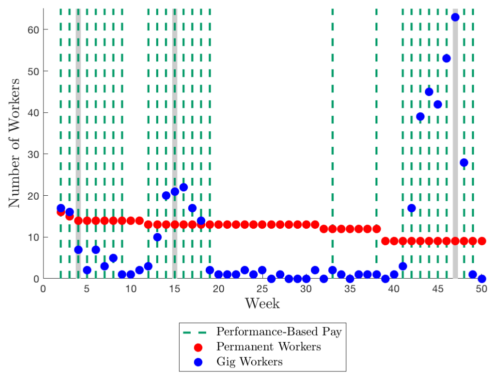
(c) Profit Maximizing Approach Incentive Pay Structure



(d) Optimal Pay and Workforce Schedule given a Profit Maximizing Approach Pay Structure



(e) Central Planner Approach Incentive Pay Structure



(f) Optimal Pay and Workforce Schedule given a Central Planner Approach Pay Structure

Figure 7: Counterfactuals Analyses

Notes: 7b represents results from a counterfactual analysis that uses the original incentive structure, but varies the weeks in which it is implemented and the workforce composition.

incentive pay structure around weeks with an anticipated increase in product demand. Different than the settings in 7b, though, the firm finds it optimal to implement the incentive pay scheme for longer periods, especially at the beginning of the year. This change leads to a different staffing schedule of permanent workers throughout the year. Specifically, the group of permanent workers remains almost unchanged during most of the year, while gig workers provide flexibility at the extensive margins during weeks of high-expected demand, as well as weeks with stable lower demand. Toward the end of the year, the firm finds it optimal to lay off a small group of permanent workers and instead hires many gig workers. Interestingly, this hiring pattern was observed in the solution for the counterfactual analysis presented in 7b, as well as in the original data, as shown in Figure 1.

Combining all these features, it is apparent that in the obtained optimal pay structure, the firm extracts the most production out of its workforce, but at the same time it also provides generous compensation at higher levels of production quantities. This structure match the firm's objective in this current setting by which it wishes to minimize costs under uncertain conditions. In particular, consistent with this goal, the firm's optimal choices under these settings yield a 26% reduction in labor costs relative to the original setup. However, this outcome comes at the expense of workers' utility, which is reduced by 8% on average.

9.3 Central Planner Approach

The previous results raise the question of whether an incentive pay structure exists that yields a Pareto improvement relative to the original settings. To answer this question, I optimized the firm's cost minimization problem over the set of parameters that define the incentive pay structure under workers' participation constraint. In particular, this constraint states that workers' expected utility under incentive pay is at least as high as the expected utility gained under the original settings.

The findings indicate that such a pay structure exists, but while workers' participation constraint is binding, the firm's optimal choices under this incentive pay structure, as presented in Figure 7f, reduced labor costs by 12%. Figure 7e presents the optimal pay structure derived from this analysis. One can see that this pay structure has similar features to those obtained under the profit-maximizing approach; however, a major difference between these two pay schemes is the level of workers' base wage. Under the central planner approach, the base wage is set to be higher than under the profit-maximizing approach; specifically, it reverted back to the base wage offered in the original settings. This outcome provides evidence that workers' base wages is the key feature of the payment structure that prevents harming workers' utility even in the face of generous performance incentives. Interestingly, this result relates both to recent discussions on the working conditions of the temporary workforce and to the claim that reliance on outsourcing, contractors, and temporary workers, particularly for non-core activities, has led to lower pay and worse working conditions (Weil, 2014).

Other features of the optimal incentive pay structure obtained in this analysis are similar to those found in the previous analysis including the two incentive thresholds located at two extremes of the production range, the piece rates values, the optimal workforce composition, and the pay schedule choices of the firm. Therefore, the previous discussion regarding these findings is relevant here.

10 Concluding Remarks

Firms with on-demand customized manufacturing systems operate while adhering to strict production timelines, therefore facing a great challenge in workforce capacity planning and commitment. In this paper I consider two labor management practices that provide the necessary flexibility to respond quickly to product demand variations: a performance-based compensation scheme and an adjustable workforce size achieved by hiring temporary gig workers. Using data from a global manufacturer, I estimate an equilibrium framework that examines the optimal combination of these two labor practices in a production environment with uncertainty regarding future product demand and workers' productive capacity. I solved initially for workers' decisions on daily effort when exploiting temporal pay scheme changes. I then embedded workers' optimal effort decisions into the firm's dynamic-cost, discrete-choice minimization problem, solving for both an optimal compensation scheme and an optimal labor-force composition.

I show that although implementing a performance-based compensation scheme and hiring temporary workers are considered as substitute practices pertaining to intensive and extensive labor margins, they are in fact complements. That is, hiring a blended workforce of gig and permanent workers and implementing an incentive pay scheme at the same periods provides the flexibility the firm needs to effectively operate an on-demand customized production system. These findings result from the two main outcomes derived in the model. First, gig workers' production response to incentive pay is large and significant, whereas permanent workers' average production response to incentive pay is not statistically different from zero for almost all job experience levels. Second, the recruiting costs of hiring gig workers are significantly lower than those of permanent workers.

I perform simulations based on the estimated model to investigate whether the firm can improve its choices relative to the original settings. I find that an incentive pay structure exists an incentive pay structure that reduces the firm's labor costs and keeps workers' participation in a constraint binding. Importantly, I further discover that maintaining workers' base wages at the original level is the key feature of the payment structure that prevents harming workers' utility even in the face of generous performance incentives. This outcome relates both to recent discussions on the working conditions of the temporary workforce and to the claim that reliance on outsourcing, contractors, and temporary workers, particularly for non-core activities, has led to lower pay and worse working conditions (Weil, 2014).

Other labor management practices are applicable and may be desirable depending on the operative settings. To illustrate, flexible work arrangements, such as hiring gig workers on an as-needed basis, may not be desirable across all job types. A job that requires teamwork or a major coordinating role may not be suitable for gig workers (Mas & Pallais, 2020). In a scenario whereby the firm inevitably hires an on-demand workforce for a job that requires team collaboration, a performance-based pay structure can be implemented for the team manager, a practice shown to be successful in inducing productivity (Bandiera, Barankay, & Rasul, 2013). Also, if firms cannot accurately measure individual worker contributions, other incentive pay plans may be implemented such as relative rankings of workers or group- based incentives.

My study leaves unexplored some areas that might be of interest for future research. In the model I assume that gig workers are hired in a "spot labor market" that has surplus supply, leading to immediate hiring when demand exists. While this assumption may be reliable in certain circumstances, depending on the market conditions and the job type, examining this friction through the model's lens could broaden our understanding of firms' optimal operation in a production environment with uncertain conditions. Additionally, in the current setting, I assume that gig and permanent workers face the same pay structure – an assumption that follows the original setting of the firm studied in this paper. An exploration of the consequences of non-identical compensation structures for permanent workers and gig workers may be desirable to reveal another layer of workers' different response to incentives. Extending the framework presented in this paper to account for such additional possibilities represents a fruitful venue for future research.

References

Akerlof, G. A. (1982). Labor contracts as partial gift exchange. *The quarterly journal of economics*, *97*(4), 543–569.

Amabile, T. M. (1993). Motivational synergy: Toward new conceptualizations of intrinsic and extrinsic motivation in the workplace. *Human resource management review*, *3*(3), 185–201.

Amabile, T. M., Conti, R., Coon, H., Lazenby, J., & Herron, M. (1996). Assessing the work environment for creativity. *Academy of management journal*, *39*(5), 1154–1184.

Angelici, M., & Profeta, P. (2020). Smart-working: Work flexibility without constraints.

Balbuzanov, I., Gars, J., & Tjernström, E. (2019). Media and motivation: the effect of performance pay on writers and content.

Bandiera, O., Barankay, I., & Rasul, I. (2013). Team incentives: Evidence from a firm level experiment. *Journal of the European Economic Association*, *11*(5), 1079–1114.

Baron, J. N., & Pfeffer, J. (1994). The social psychology of organizations and inequality. *Social Psychology Quarterly*, 190–209.

Beckmann, M. (2016). Working-time autonomy as a management practice. *IZA World of Labor*.

Bellemare, C., & Shearer, B. (2011). On the relevance and composition of gifts within the firm: Evidence from field experiments. *International Economic Review*, *52*(3), 855–882.

Bénabou, R., & Tirole, J. (2002). Self-confidence and personal motivation. *The quarterly journal of economics*, *117*(3), 871–915.

Benabou, R., & Tirole, J. (2003). Intrinsic and extrinsic motivation. *The review of economic studies*, *70*(3), 489–520.

Besley, T., & Ghatak, M. (2005). Competition and incentives with motivated agents. *American economic review*, *95*(3), 616–636.

Bewley, T. (1999). Work motivation. *Federal Reserve Bank of St. Louis Review*, *81*(May/June 1999).

Bloom, N., Liang, J., Roberts, J., & Ying, Z. J. (2015). Does working from home work? evidence from a chinese experiment. *The Quarterly Journal of Economics*, *130*(1), 165–218.

Carroll, G. (2015). Robustness and linear contracts. *American Economic Review*, *105*(2), 536–63.

Chade, H., & Swinkels, J. (2019). *Disentangling moral hazard and adverse selection* (Tech. Rep.). Working Paper, Arizona State University.[450].

Charness, G., & Rabin, M. (2002). Understanding social preferences with simple tests. *The Quarterly Journal of Economics*, *117*(3), 817–869.

Chen, M. K., Rossi, P. E., Chevalier, J. A., & Oehlsen, E. (2019). The value of flexible work: Evidence from uber drivers. *Journal of Political Economy*, *127*(6), 2735–2794.

De Charms, R. (2013). *Personal causation: The internal affective determinants of behavior.* Routledge.

Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of personality and Social Psychology*, *18*(1), 105.

Deci, E. L., Olafsen, A. H., & Ryan, R. M. (2017). Self-determination theory in work organizations: The state of a science. *Annual Review of Organizational Psychology and Organizational Behavior*, *4*, 19–43.

Dong, J., & Ibrahim, R. (2017). *Flexible workers or full-time employees? on staffing systems with a blended workforce* (Tech. Rep.). Working paper.

Eckstein, Z., & Wolpin, K. I. (1999). Why youths drop out of high school: The impact of preferences, opportunities, and abilities. *Econometrica*, *67*(6), 1295–1339.

Eriksson, T., & Kristensen, N. (2014). Wages or fringes? some evidence on trade-offs and sorting. *Journal of Labor Economics*, *32*(4), 899–928.

Fehr, E., & Goette, L. (2007). Do workers work more if wages are high? evidence from a randomized field experiment. *American Economic Review*, *97*(1), 298–317.

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The quarterly journal of economics*, *114*(3), 817–868.

Foarta, D., & Sugaya, T. (2019). Wait-and-see or step in? dynamics of interventions.

Foote, D. A., & Folta, T. B. (2002). Temporary workers as real options. *Human Resource Management Review*, *12*(4), 579–597.

Francois, P. (2000). 'public service motivation'as an argument for government provision. *Journal of Public Economics*, *78*(3), 275–299.

Freeman, R. B., & Kleiner, M. M. (2005). The last american shoe manufacturers: Decreasing productivity and increasing profits in the shift from piece rates to continuous flow production. *Industrial Relations: A Journal of Economy and Society*, *44*(2), 307–330.

Georgiadis, G., & Powell, M. (2020). A/b contracts.

Glazer, A. (2004). Motivating devoted workers. *International Journal of Industrial Organization*, *22*(3), 427–440.

Gottlieb, D., & Moreira, H. (2015). Simple contracts with adverse selection and moral hazard. *The Wharton School Research Paper*(78).

Gourieroux, C., Monfort, A., & Renault, E. (1993). Indirect inference. *Journal of applied econometrics*, *8*(S1), S85–S118.

Guiteras, R. P., & Jack, B. K. (2018). Productivity in piece-rate labor markets: Evidence from rural malawi. *Journal of Development Economics*, *131*, 42–61.

Hall, J. V., Horton, J. J., & Knoepfle, N. D. T. (2020). Ride-sharing markets re-equilibrate.

Hölmstrom, B. (1979). Moral hazard and observability. *The Bell journal of economics*, 74–91.

Holmstrom, B. (1982). Moral hazard in teams. belljournal of economics, 13, 324-340. *Holmstrom32413Bell Journal of Economics1982*.

Holmstrom, B., & Milgrom, P. (1987). Aggregation and linearity in the provision of intertemporal incentives. *Econometrica: Journal of the Econometric Society*, 303–328.

Holmstrom, B., & Milgrom, P. (1991). Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *JL Econ. & Org.*, *7*, 24.

Hong, F., Hossain, T., List, J. A., & Tanaka, M. (2018). Testing the theory of multitasking: Evidence from a natural field experiment in chinese factories. *International Economic Review*, *59*(2), 511–536.

Houseman, S. N. (2001). Why employers use flexible staffing arrangements: Evidence from an establishment survey. *Ilr Review*, *55*(1), 149–170.

Keane, M. P., Todd, P. E., & Wolpin, K. I. (2011). The structural estimation of behavioral models: Discrete choice dynamic programming methods and applications. In *Handbook of labor economics* (Vol. 4, pp. 331–461). Elsevier.

Keane, M. P., & Wolpin, K. I. (1994). The solution and estimation of discrete choice dynamic programming models by simulation and interpolation: Monte carlo evidence. *The Review of Economics and Statistics*, 648–672.

Keane, M. P., & Wolpin, K. I. (1997). The career decisions of young men. *Journal of political Economy*, *105*(3), 473–522.

Keane, M. P., & Wolpin, K. I. (2001). The effect of parental transfers and borrowing constraints on educational attainment. *International Economic Review*, *42*(4), 1051–1103.

Laffont, J.-J., & Martimort, D. (2009). *The theory of incentives: the principal-agent model.* Princeton university press.

Laffont, J.-J., & Tirole, J. (1988). The dynamics of incentive contracts. *Econometrica: Journal of the Econometric Society*, 1153–1175.

Lazear, E. P. (1989). Pay equality and industrial politics. *Journal of political economy*, *97*(3), 561–580.

Lazear, E. P. (2000). Performance pay and productivity. *American Economic Review*, *90*(5), 1346–1361.

Lazear, E. P. (2004). Output-based pay: incentives, retention or sorting. *research in Labor Economics*, *23*(4), 1–25.

Lazear, E. P., & Rosen, S. (1981). Rank-order tournaments as optimum labor contracts. *Journal of political Economy*, *89*(5), 841–864.

Maestas, N., Mullen, K. J., Powell, D., Von Wachter, T., & Wenger, J. B. (2018). *The value of working conditions in the united states and implications for the structure of wages* (Tech. Rep.). National Bureau of Economic Research.

Mas, A., & Pallais, A. (2017). Valuing alternative work arrangements. *American Economic Review*, *107*(12), 3722–59.

Mas, A., & Pallais, A. (2020). Alternative work arrangements. *Annual Review of Economics*, *12*(1), 631-658. Retrieved from https://doi.org/10.1146/annurev-economics-022020-032512 doi: 10.1146/annurev-economics-022020-032512

McFadden, D. (1989). A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica: Journal of the Econometric Society*, 995–1026.

Mirrlees, J. A. (1976). The optimal structure of incentives and authority within an organization. *The Bell Journal of Economics*, 105–131.

Oyer, P., & Schaefer, S. (2010). *Personnel economics: hiring and incentives* (Tech. Rep.). National Bureau of Economic Research.

Pinker, E. J., & Larson, R. C. (2003). Optimizing the use of contingent labor when demand is uncertain. *European Journal of Operational Research*, *144*(1), 39–55.

Rosen, S. (1986). The theory of equalizing differences. *Handbook of labor economics*, *1*, 641–692.

Salop, J., & Salop, S. (1976). Self-selection and turnover in the labor market. *The Quarterly Journal of Economics*, 619–627.

Shapiro, C., & Stiglitz, J. E. (1984). Equilibrium unemployment as a worker discipline device. *The American Economic Review*, *74*(3), 433–444.

Shearer, B. (2004). Piece rates, fixed wages and incentives: Evidence from a field experiment. *The Review of Economic Studies*, *71*(2), 513–534.

Todd, P. E., & Wolpin, K. I. (2006). Assessing the impact of a school subsidy program in mexico: Using a social experiment to validate a dynamic behavioral model of child schooling and fertility. *American economic review*, *96*(5), 1384–1417.

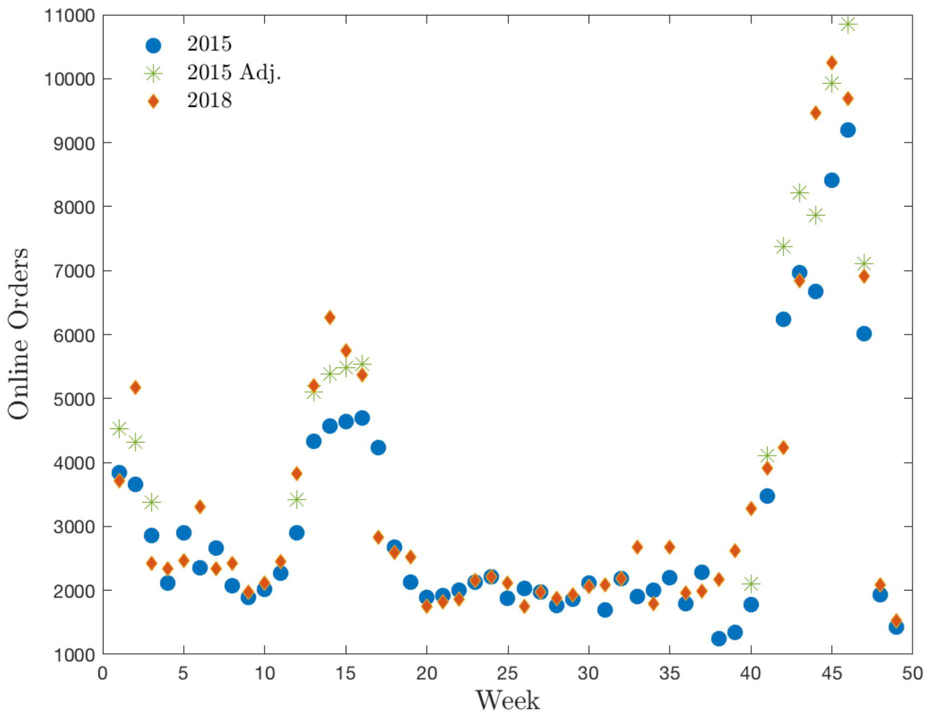
Weil, D. (2014). *The fissured workplace.* Harvard University Press.

Wiswall, M., & Zafar, B. (2018). Preference for the workplace, investment in human capital, and gender. *The Quarterly Journal of Economics*, *133*(1), 457–507.

Appendix

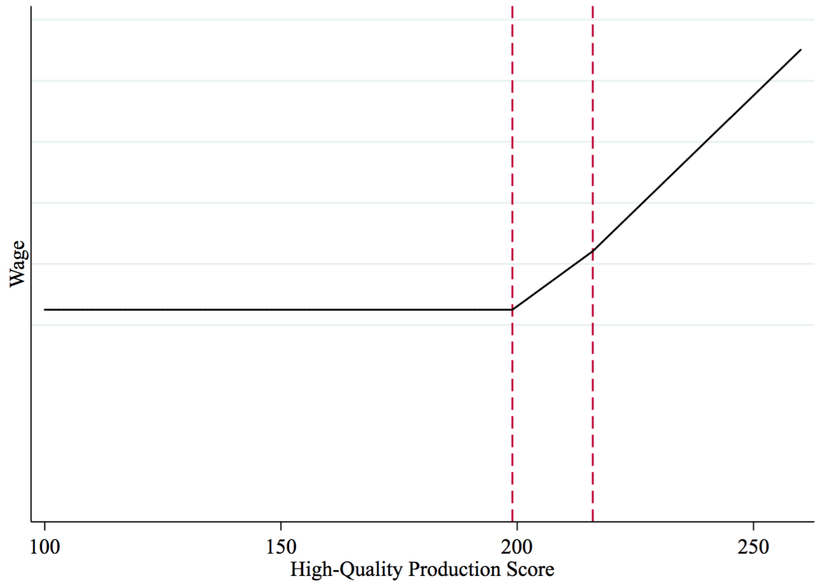
A Figures

Figure A.1: Average Daily Product Demand



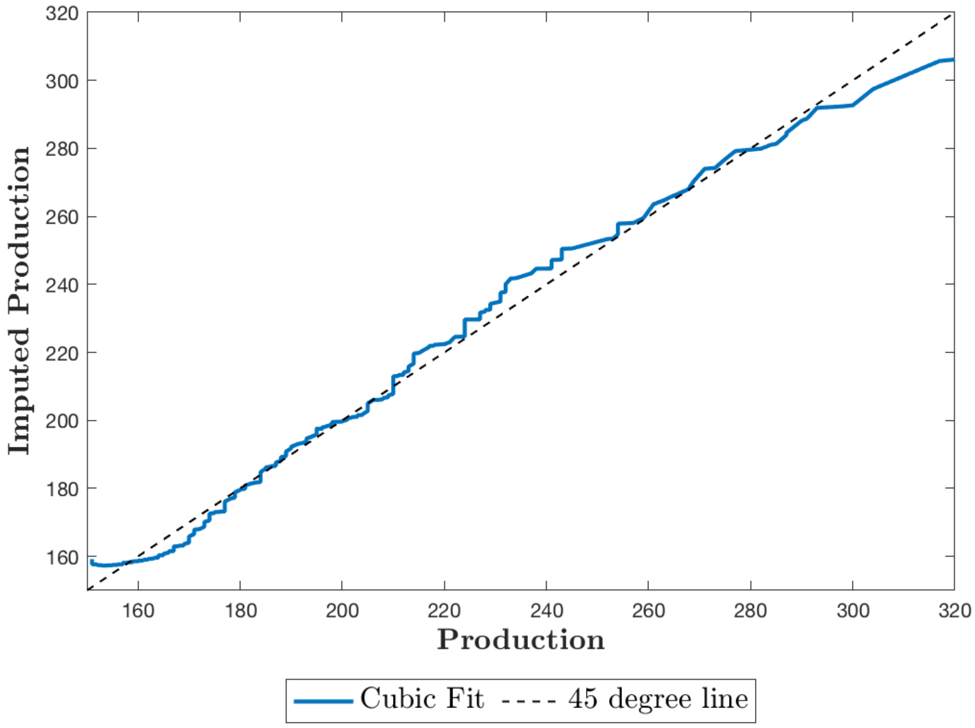
Notes: The data set used to generate this figure is from the years 2015 and 2018. Daily product demand is the sum of all the orders placed online in a given day. The adjusted product demand of the year 2015, which is used as the baseline sources of product demand data in the firm's dynamic cost minimization problem, has been calculated by adding 18% increase during high seasons.

Figure A.2: Wage Structure



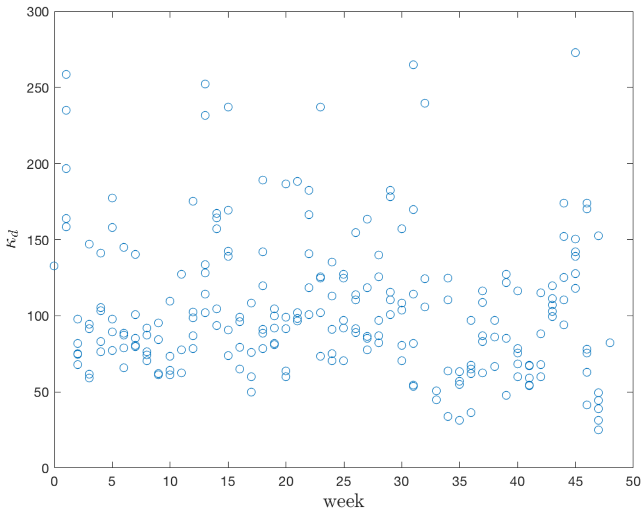
Notes: The figure presents the exact wage structure, as it was offered during the year 2018. The first dashed vertical line represents the incentive regime entry threshold after which workers receive additional pay for production above it. The second vertical line denotes a second threshold within the performance-pay regime associated with a slightly higher pay rate for additional production (shown by the steeper slope of the wage line). This compensation method incorporates a convex structure in which a worker's compensation is higher when exerting higher effort levels.

Figure A.3: Production and Imputed Production



Notes: The figure presents observed production with the imputed production based on production score. The imputed production variable is used to adjust the threshold for entering the performance-based pay regime, which is defined by production score.

Figure A.4: Daily Competitive Climate



B Proof of Concavity of the Worker's Objective Function

Proof. Define $U(E)$ to be the the worker's objective function, as defined by the set of equations in 4. The proof by which this objective function is concave with respect to effort follows from the fact that $\frac{\partial U}{\partial E^2} < 0 \; \forall E \in [1, 10]$

The first derivative of this objective with respect to effort under the most general settings, whereby $\beta > 0$, takes the following form:

$$\begin{aligned}\frac{\partial U}{\partial E} =& \beta \left[\frac{1}{1+\exp\left(\phi_E E_{i\nu d} + \phi_X X_{i\nu d}\right)}\right] \alpha_\nu X_{i\nu d}^{\delta} \mathrm{e}^{\varepsilon_{id}} - \\ & \beta \left[\frac{\phi_E \exp\left(\phi_E E_{i\nu d} + \phi_X X_{i\nu d}\right)}{\left(1+\exp\left(\phi_E E_{i\nu d} + \phi_X X_{i\nu d}\right)\right)^2}\right] \alpha_\nu \Big(E_{i\nu d}\Big) X_{i\nu d}^{\delta} \mathrm{e}^{\varepsilon_{id}} \\ & + \eta_\nu - \frac{\kappa_d}{X_{i\nu d}} \gamma_\nu \Big(E_{i\nu d}\Big)^{(\gamma_\nu - 1)}.\end{aligned}$$

The second derivative of the worker's objective function then takes the following form:

$$\begin{aligned}\frac{\partial U}{\partial E^2} =& -\beta \left[\frac{\phi_E \exp\left(\phi_E E_{i\nu d}^* + \phi_X X_{i\nu d}\right)}{\left(1+\exp\left(\phi_E E_{i\nu d}^* + \phi_X X_{i\nu d}\right)\right)^2}\right] \alpha_\nu X_{i\nu d}^{\delta} \mathrm{e}^{\varepsilon_{id}} \\ & -\beta \left(\left[\frac{\phi_E \exp\left(\phi_E E_{i\nu d}^* + \phi_X X_{i\nu d}\right)}{\left(1+\exp\left(\phi_E E_{i\nu d}^* + \phi_X X_{i\nu d}\right)\right)^2}\right] \alpha_\nu X_{i\nu d}^{\delta} \mathrm{e}^{\varepsilon_{id}} + \left[\frac{\phi_E^2 \exp(\phi_E E_{i\nu d}^* + \phi_X X_{i\nu d})(1-\exp(\phi_E E_{i\nu d}^* + \phi_X X_{i\nu d}))}{\left(1+\exp\left(\phi_E E_{i\nu d}^* + \phi_X X_{i\nu d}\right)\right)^3}\right] \alpha_\nu X_{i\nu d}^{\delta} \mathrm{e}^{\varepsilon_{id}}\right) \\ & - \frac{\kappa_d}{X_{i\nu d}} \gamma_\nu (\gamma_\nu - 1) \Big(E_{i\nu d}^*\Big)^{(\gamma_\nu - 2)}.\end{aligned}$$

This result directly shows that the desired condition is satisfied, that is:

$$\frac{\partial U}{\partial E^2} < 0.$$

□