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

Preliminary Draft

DO NOT DISTRIBUTE

Michal Hodor†

click here for latest version

August 12, 2020

Abstract

As many manufacturing companies aspire to becoming the online, on-demand supplier of the product or service they bring to the market, the natural question is how a firm should optimally operate such a production system. This paper addresses this question, and claims that to efficiently implement it, a firm needs to employ a bonus performance pay scheme and hire both gig and permanent workers to have the necessary flexibility to deal with immediate production response and demand uncertainties. I develop a comprehensive structural framework that includes both the firm and its workers, and then apply an equilibrium analysis to solve it using a recent dataset from an online, global, mid-size manufacturer that produces customized apparel and accessories. The main findings indicate that permanent workers' productivity levels change only slightly as a result of pay incentives, while gig workers' productivity levels are significantly higher under performance-based pay regimes. I then embed these results into the firm's problem, and solve a dynamic problem, finding both the optimal compensation method and the optimal labor-force composition. The findings indicate that the decision regarding which of these tools to use is not straightforward, as they could be completing or competing solutions, depending on several factors, including workers' recruiting costs, demand variations, and forecasting precision.

†University of Pennsylvania, Economics Department (mhodor@sas.upenn.edu)

1 Introduction

Major changes have occurred in the global economy, as the practice of online shopping continues to grow and suppliers increasingly offer on-demand production.[1,2] Accordingly, many manufacturing companies aspire to becoming the online, on-demand market supplier of their product.[3]

The advantages that come with customized on-demand manufacturing are two-fold: customers obtain exactly what they want, and the firm only has to produce as much as it sells. That means little waste and no excess inventory or over-production in times of high-expected demand.[4] However, operating an on-demand production system poses enormous challenges in planning and committing to a workforce capacity. Such systems are unlike traditional production systems in which firms operate assembly lines that work to produce large quantities of products kept in storage facilities until delivery. In customized on-demand manufacturing, an adjustable manufacturing process aims to produce customized items based on real-time data from consumers with the minimum possible lead time.[5] Hence, the production system constantly moves from off-peak periods when only a handful of workers is needed to peak times when demand surges and immediate response is required.

This naturally raises the question of how a firm should optimally operate a customized on-demand production system with the goal of maximizing profits. Specifically, what modifications are required to labor management practices for a firm to successfully implement such a production process. I propose two instruments that the firm can integrate into its labor management practices. The first is a flexible-incentive pay system that offers a performance-based compensation scheme. This component varies labor input on the intensive margin, with the underlying assumption that incentive pay elicits higher production (see Oyer & Schaefer, 2010, for review). The second is an adjustable labor force input, which can be achieved by hiring on-demand workers (hereafter, "gig workers"),[6] in addition to a stable core of permanent labor force. This component provides flexibility at the extensive margin of labor. In practice, one can think about these two instruments as demonstrating a trade-off between labor quantity and effort quality, and the decision over which of them to implement is an empirical one.

I investigate how these labor instrument can be used to achieve optimal operation of an on-demand customized production process using a novel data from an online, global, mid-size manu-

[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]The on-demand production trend expresses itself in various behaviors, including the way people watch TV shows and movies, rent cars, or summon taxis.

[3]The fashion industry has numerous examples of suppliers who sell online customized apparel and accessories, and this paper focuses on a manufacturer from this industry.

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

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

[6]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.

facturer that produces customized apparel and accessories items (hereafter, "the firm").[7] The data contains a rich set of information for all individuals that worked for this manufacturer in 2015 and in 2018, including detailed objective measures of daily production and its quality, daily wages, as well as aggregate information on the daily product demand and the exact pay structure.[8] Apart from serving as a representative example for the type of emerging on-demand customized production firms in the economy, it also applies unique labor management practices that are especially suitable for this paper's goal. Specifically, to successfully fulfill its product demand fluctuations under the minimum possible lead time, the firm applied both the intensive and extensive margin instruments previously introduced. The majority of employees are gig workers, hired on a seasonal basis for short-contract terms before times of anticipated product-demand surges.[9] Additionally, the firm deviated from a flat-wage pay structure to a bonus performance-pay scheme twice for short periods along the span of the data. This abrupt pay change was orthogonal to the management method, and thus provides exogenous variation that assist in overcoming potential identification problems, including confounding effects caused by changes in other labor management practices, and sorting and selection effects that occur when more productive workers self-select into the firm.[10] The data thus provides a unique framework with the necessary details for the analysis carried in this paper.

I start by presenting reduced-form evidence that examine the evolution of production patterns and response to incentives over time in the firm, separately for gig and permanent workers. The findings show striking fundamental differences between these workers types. While permanent workers hold close to constant production quantities over various levels of job experience, gig workers demonstrate a parabolic production pattern over time spent in the firm. When controlling for worker-specific heterogeneity this analysis shows that gig workers are significantly more productive than permanent workers for almost all job experience levels. Furthermore, gig workers exhibit statistically and economically significant response to incentive that amount to between 12% to 17% productivity gains on average under a performance-based payment scheme. At the same time, permanent workers average production response to incentive is not statistically different than zero for almost all job experience levels when accounting for worker-specific heterogeneity.

These empirical findings emphasize the complexity in the decision over which of the labor instruments previously introduced to implement. On the one hand, the firm bears the costs of recruiting workers, which vary by worker type (permanent or gig) due to their availability, screen-

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

[8] 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 factories.

[9] Gig workers are hired in addition to a core of permanent employees.

[10] Salop and Salop (1976) introduced the idea that firms can use compensation scheme as a self-selection device; 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; 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; and Oyer and Schaefer (2005) argued that inducing self-selection is one of the leading potential explanations for the recent rise in stock-based pay for lower-level employees.

ing process, and job proficiency. On the other hand, there is additional pay that arises from a performance-contingent contract, which could also vary with worker type, because each type may respond differently to monetary incentives. Although in practice both instruments can be used to immediately adjust production levels to variations in product demand, evidence indicates that the extensive margin solution is a common human resource strategy (Foote & Folta, 2002; Grossman, 1998), while the intensive margin solution is rather rare.

To conduct a comprehensive cost-benefit analysis that considers the implementation of these labor instruments, I develop a structural framework of a firm that operates an on-demand customized production system. I apply an equilibrium analysis to solve this framework in two integrated stages that include both the firm and its workers. In the first stage, workers solve for the optimal daily-effort decision under differing compensation schemes and varying daily workloads.[11] The worker's problem captures features that could explain the fundamental production differences between permanent and gig workers, as documented in the reduced-form analysis. Workers are heterogeneous in their total factor productivity, personal motivation,[12] and effort cost. In the second stage, the firm solves a dynamic problem with heterogeneous labor input and finds both the optimal compensation method (flat wage or bonus incentive pay) and the optimal labor-force composition (composed of gig and permanent workers). Workers' optimal effort decisions and their on-the-job learning processes are integrated into the firm's problem through the incentive compatibility constraint, in which workers vary in terms of their temporary or permanent status and job experience. [13] The firm's dynamic problem further accounts for product demand uncertainty, inventory limitation,[14] and short-demand lead times – all essential features that characterize firms that produce on-demand customized products, and more generally, features that are common among global export-oriented factories.

I exploit the substantial variations in product-demand in the presence of massive hiring of gig workers and exogenous changes in the pay scheme to estimate the equilibrium framework and characterize the optimal labor operation of a firm that produce on-demand customized products. I apply different estimation methods in each step of the equilibrium model. At the first stage, I use an indirect inference estimation procedure to solve for a worker's daily effort decision. At the second stage, I apply a simulated maximum-likelihood procedure to uncover the costs associated with hiring gig and permanent workers, the off-laying cost of permanent workers, and the natural separation rate of permanent workers. This analysis is done under the assumption that gig workers

[11]Throughout this paper daily workload is defined as the product demand divided by the number of workers in a given day.

[12]This notion is motivated by Amabile (1993) who introduced worker's intrinsic motivation into models of labor supply.

[13]Taking into account workers' on-the-job learning process is particularly important in an analysis that aims to capture the cost-benefit tradeoff of hiring gig workers, because they are usually untrained and hired for a short-term period.

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

are hired for a predetermined periods, and thus no layoff costs or separation rate from the firms are associated with these workers. Also, in both stages I am taking as given the incentive pay structure offered by the firm and assume that the firm and workers behave optimally subject to the structure.

The results from the worker's problem indicate that the sources of the distinct production patterns of gig and permanent workers are personal motivation and effort cost. Specifically, I find that gig workers hold 50% higher personal motivation than permanent workers, a difference which significantly significant. The personal motivation 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 being task-oriented, desire for independence, or simply because they like to undertake certain activities.[15]. Further, I find statically and economically difference in the curvature of the effort cost function among workers by their types, whereby gig workers hold quadratic effort cost curvature and permanent workers' effort cost curvature is close to linear.

The solution of the firm's problem quantifies the hiring costs associated with each of the workers' types. The estimates indicate that the cost of hiring gig workers are significantly lower than the hiring cost associated with permanent workers. This results coincide with the notion that permanent workers are going through a more rigorous screening process since they are hired based on a long-term contract, whereas gig workers, in contrast, are hired to "fill in the gap" with less emphasis on job match.

Rigorous estimation of the structural equilibrium framework establishes a comprehensive understanding of the choices made by both the workers and the firm and thus serve as a solid benchmarks to characterize their optimal behavior beyond the given setup. I implement this idea by applying two types of policy analyses. First, I use the estimates of the structural parameters to find the optimal combination of pay scheme and labor-force composition, so that the incentive pay benefits would offset the costs to a firm in an on-demand environment with high-demand uncertainty. I find that it is only the combination of the two labor instruments that provides the necessary flexibility to optimally operate an on-demand customized production process. Specifically, back-of-the-envelope calculations show that the firm could reduce the labor cost during peak seasons by 22% by integrating both the hiring of gig workers and the implementation of bonus incentive pay.

Additionally, I use the estimated structural parameters to design an optimal pay structure through a counterfactual analysis. Specifically, I apply two conceptually different approaches in this analysis. First, I adopt a utilitarian approach, which focus on the firm's profitability and seek to find the optimal bonus incentive structure that minimize the firm's labor cost given the output demand uncertainty. The second approach follows a central planer perspective, as it aims to find

[15]Psychologists have argued that individuals undertake many activities without expecting for an extrinsic reward (De Charms, 2013; Deci, 1971). Deci, Olafsen, and Ryan (2017) offered a review of studies that 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 and Tirole (2003); Besley and Ghatak (2005); Francois (2000); Glazer (2004).

the optimal bonus incentive structure under the constraints by which workers' utility is greater than or equal to the one gained in the original settings.

This study sheds new light on how different compensation and hiring schemes can be used to enhance firms' production and profitability both generally, and particularly in an online on-demand customized setting. Even though these types of firms are prevalent today, there is no previous study that has investigated this type of emerging firms and explored the optimal labor-management decisions under their strict and limiting production polices.

2 Related Literature

This paper relates to several strands of literature concerning the relationship between a worker and a firm. Broadly, it relates to the agency theory. Similar to studies in this field, the current study seeks to develop the optimal contract that binds the principal (firm) and the agent (worker) while taking into account their unmatched interests. However, unlike the common practice of using abstract theory and structure that are less applicable to practical problems (see Laffont & Martimort, 2009, for review), this paper constructs an empirically tested optimal contract based on real firm data. In particular, the suggested contract is not only an anecdotal illustration of the employee-employer relationship, but rather it is the result of a thorough examination of the consequences of contract execution on both the worker and the firm.

This study more closely relates to work that has attempted to solve firms' labor-decision problem in labor economics. However, models in this literature usually have had little to say about the composition of the optimal workforce and instead focused on how many workers should be hired under the assumption of a flat, market-clearing wage (Hamermesh, 1991). Models that departed from the fixed-wage assumption adopted the efficiency wage hypothesis, which claims that the wage should be higher than the market-clearing wage in order to encourage workers to increase their productivity (and potentially reduce turnover). Other models followed the tournament theory, which suggests that 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, however, 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.

Personnel-scheduling models in the operational management literature 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 both permanent and temporary workers, and Bard (2004b) who considered a similar problem in service organizations while incorporating differences in employee skill levels and demand uncertainty. Most relevant to this current paper is a study 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 flexibility to meet the time variation in customers' demand, and labor supply-side uncertainty. Their paper and the current one target a problem of a similar nature, however, the former is highly theoretical, with applications and solutions based on simulations and mathematical programming less relevant to practical problems, while the latter applies an empirical and data-driven approach. Moreover, the framework presented in this paper introduces an additional and pertinent aspect to the labor problem, which refers to the labor pay structure.

A crucial aspect of the labor-force decision is the worker's productivity level. Studies in labor economics and operational management literatures include various factors that affect firms' hiring decisions, such as education and age, however they usually overlook the crucial feature of how employees' production level evolves over time. In most models, there is an implied assumption that employees reach their highest sustainable production capacity immediately upon joining the firm. In practice, however, employees' production abilities are changing over time, both in the direction of increasing productivity via an on-the-job learning process or decreasing productivity due to fatigue. This aspect is particularly important when hiring gig workers, because they are untrained workers hired for a short-term work period when production distortions are presumably at their peak. Stratman, Roth, and Gilland (2004) identified this gap and built a theoretical model that takes into account dynamics of workforce skill levels in the case of a workforce composed of permanent and temporary workers. At the core of their analysis stands a premise that temporary workers have relatively less skill and therefore higher average production times, higher average defect rates, and lower rates of learning. These are strong assumptions that do not necessarily hold in practice, and in particular, will be shown not to hold in the data used in the current paper. Hence, this paper relaxes these assumptions and develops a model that captures workers on-the-job learning patterns as exhibited in the real data.

Since this paper's modeling framework incorporates a bonus pay structure, the linkage between worker productivity and performance-based payment methods completes the description of the relationship between a worker and a firm relevant to this paper. At the core of economists' attention are two main questions relating to this concept: how workers respond to a given set of incentives and the optimal set of incentives that an employer should provide. Many studies in personnel and labor economics have answered the first question and established a positive relationship between incentives and productivity (see Oyer & Schaefer, 2010, for review). The second questions has been mainly subject to theoretical research, as discussed above. The current paper answers both questions in innovative and pertinent settings. It is the first to investigate the differential production response to incentives of gig and permanent workers (who are subject to an identical payment scheme) in the same workplace. Moreover, it provides a structural framework that enables a through exploration of the optimal set of incentives that an employer should provide in an environment of on-demand customized production.

With few exceptions, the majority of the empirical papers that have examined the relationship between incentives and productivity (all of which consider a homogeneous labor force) use reduced-form methods. The papers most similar to this one are Shearer (2004) and Bellemare and Shearer (2011), which also used a structural framework to identify the underlying mechanism behind workers' responses to incentives. The current paper goes beyond these studies by considering a heterogeneous labor force composed of both gig and permanent workers, and furthermore modeling the firm's dynamic pay-structure decisions in an equilibrium framework, with the goal of linking productivity and profitability over time. Even though it is clear that firms maximize discounted profits and not productivity, many studies tend to ignore this non-trivial linkage. Freeman and Kleiner (2005) illustrated how incentives and profits are not necessarily positively related. They found that the abolition of performance pay reduced productivity but increased profits as quality rose in the absence of production incentives. By using a structural modeling approach to study the underlying mechanisms behind performance incentives for both the worker and the firm, this paper conducts a thorough assessment of the linkage between productivity and profitability, which takes into output quality in addition to other essential features that characterize this linkage.

While labor and personal economists have focused on studying setups of a traditional labor force, gig workers have been the focus of researchers in operations management. Specifically, such studies have examined firms' operational and pricing decisions for service jobs of a self-scheduling nature performed by independent providers, such as ride-hailing drivers. To illustrate, Chen (2016) showed that dynamic wage schemes leads drivers to work more hours with analyzed data from 25 million rides. Hall, Horton, and Knoepfle (2018) used differences in timing and city size, as well as exogenous fare changes, to identify the fare impact on drivers' hourly earnings. The current paper answers questions of a similar nature with three major differences. First, the job environment considered is in manufacturing, not service. Second, workers are not self-scheduling but rather assigned to shifts by their manager and last, the workforce arrangement is such that both permanent workers and gig workers are present at all times. Although 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; Grossman, 1998), there has been little academic research on the management practices of gig workers in a manufacturing environment.

In this paper, I seek to fill the abovementioned gaps by bridging the literatures of labor economics, personnel economics, and operational management. I take a unique viewpoint by applying an empirical equilibrium approach for a firm that operates in an on-demand production environment. This novel framework incorporates intertwined decisions that have been highlighted in various literatures – i.e., workforce size decisions, workforce composition decisions, and decisions on the pay-structure nature. Coupling these decisions is especially timely, as gig-type jobs are becoming more prevalent. However, these jobs pose both a theoretical challenge with a need to expand the traditional firm view to account for relevant factors adopted by current firms, and a practical challenge of how firms can optimally operate under high product-demand variability in which tasks are time-sensitive.

3 Data and Settings

3.1 General

The recent micro-data used in this paper is from an online, global, mid-size manufacturer that produces customized apparel and accessories items. The data contain a rich set of information on all individuals that worked for the firm in 2015 and in 2018.[16] During these years, the management adopted new compensation practices. Specifically, around the months of February 2015 and December 2018, the management deviated from its usual pay structure – a flat hourly wage – and established a performance-pay scheme with the goal of inducing productivity in times of demand peak. The data from 2018 is utilized as the main data source throughout the paper, as it includes a larger workforce and provides detailed information about the quality of production and production score (which will be discussed in detail later). The data from 2015 includes fewer workers and partial quality measures and was thus used to perform an out-of-sample validation test of the model assumptions and to compare the model's 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 Days	328.68	23.35	14.67
Total Production Adj	122.02	97.50	5.12
Low-Quality Production Adj	2.45	3.41	-2.88
Production Score Adj	155.41	138.67	1.65
Number of Workers	44	216	

Notes: Production measures are adjusted to 8 hours of work.

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 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.

[16] Months February through December.

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[17] 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 the worker actually attended work.[18] Table 1 shows that during the year of 2018, 216 (83%) of the 260 workers who worked in the assembly department were gig workers. As expected, the average number of days of experience among gig-workers, which stands on 23 days, is significantly lower than that of permanent workers who are employed on average for 1 year.

The employment records include information about the workers' gender, and for for most employees, their ages.[19] 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 the age of workers 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. Looking at the descriptive characteristics creates an image of a typical gig worker – to illustrate, a gig worker is a young woman, presumably a student on vacation, who wishes to fill spaces in her schedule with non-binding work while school is not in session.

3.3 Production

The detailed 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. 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.[20] 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). Hence, as each

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

[18] 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.

[19] 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.

[20] 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.

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 plant's attendance system. I used this data to adjust workers' production by the number of hours worked, and thus created a uniform and comparable production measure.

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.[21] 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. Since 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 type and implies a dramatic difference whereby permanent workers are more productive than gig workers. Although this conclusion may seem intuitive, as it is presumably assumed that permanent workers are more skilled than gig workers, it is in fact misleading. While gig workers are usually hired before or during times of demand surge, permanent workers work throughout the year. Moreover, workers' productivity is not only a function of the workload, but more closely defined by their experience. As such, gig and permanent workers are working at times under intrinsically different conditions and merely comparing the average production yields an incomplete picture. In order to reliably compare the productivity levels of permanent and gig workers, a more comprehensive framework is required, as is subsequently performed in this paper. Also, Table 1 shows that permanent workers work less on average than gig workers, however, the shift length is determined by the manager and not the worker, so this is not a concerning outcome.

3.4 Demand

The daily demand variable represents the sum of all orders placed online in a given day. Fig. 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

[21] The scoring menu was established during 2018 and finalized and presented to workers several weeks before implementation of the incentive scheme. For this reason the analysis that includes the production score is restricted to the months of October, November and December.

days of production, starting on the day the order is placed and ending on the day the item is shipped. Moreover, the firm guarantees its consumers the minimum possible shipment lead time, depending on where the item is being shipped. 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. This consistent observation helps to simplify the demand forecast constraint at the firm's problem, which takes this particular seasonal demand pattern with additional daily idiosyncratic shock.

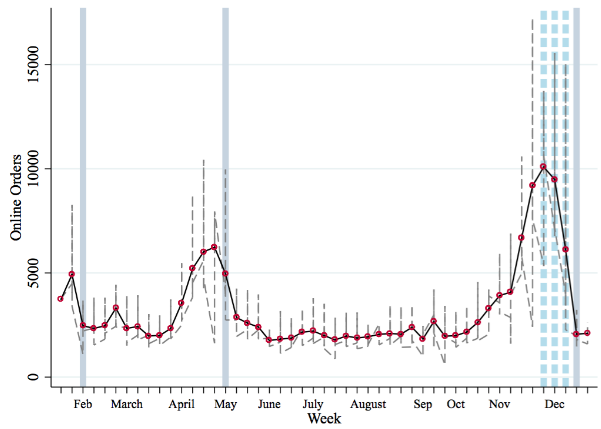
3.5 Labor-Force Size

Figures 1b and 1c presents a time series of average production per worker and average number of workers by type over the weeks of 2018, respectively. A simple examination of these figures indicate that production is positively related to surges in demand and negatively related to 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 times of high demand many new and inexperienced gig workers join the labor-force pool. Thus, if job skills and experience are significant elements in determining production, workload might then remain relatively high even in the face of mass hiring. At the same time, the performance-pay schedule also changes workers' production incentives irrespective of the number of workers and demand.

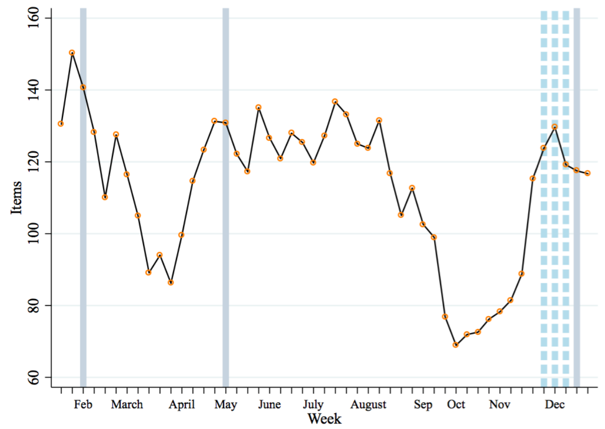
Figure 1d sheds light on the relationship between experience and job proficiency by showing a time series of the share of defective items out of total production and the number of new workers over the weeks of 2018. New workers do not undergo an official training program; in practice, they have on-the-job 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 their respective roles and levels of importance when searching for a firm's optimal labor-management behavior is one of the goals of this paper.

3.6 Pay Incentives

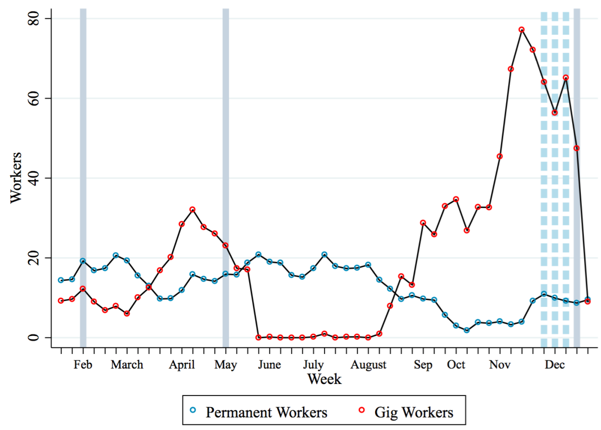
During the time under consideration, the firm switches between two pay regimes: a flat wage and a performance-based wage. During periods in which the flat-wage scheme is in effect, workers are



(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 dashed gray line in Fig. 1a represents the daily orders and the solid black line represents the average of these orders over weeks. Production per worker and defective production per worker are adjusted for nine 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.

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 pay on top of their minimum fixed hourly rate, while production below the threshold reverts back to the flat-wage regime. Figure A.1 presents this exact wage structure, as it was offered during the year 2018.

There are several reasons to adhere to such an performance-based incentive pay with a minimum guarantee, both from the worker and firm perspective. 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 guarantees that they cannot be worse off under the performance-based scheme, only better. In particular, this incentive setup completely eliminates the risk associated with pure performance-pay structure and instead gives workers 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 firm performance, as changes in incentives may reflect a wider package of changes in the firm's management practices. Second, the literature emphasizes a concomitant change in workforce composition when introducing performance pay, because of high-ability workers attracted to this form of pay.[22] Last, there is an issue with the output quality generated under performance-pay schemes, as such schemes might motivate workers to speed up production by compromising quality. The data set used in this paper overcomes all of these challenges. First, the information come 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 have been offered for 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 keeping workers attentive to quality during production. This system prevented workers from exploiting the performance-pay system and guaranteed its efficacy in increasing high-quality production.

[22] Salop and Salop (1976) introduced the idea that firms can use compensation scheme as a self-selection device; 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;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; and Oyer and Schaefer (2005) argued that inducing self-selection is one of the leading potential explanations for the recent rise in stock-based pay for lower-level employees.

The literature discusses 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 it may lead workers to reduce cooperation with their co-workers (Baron & Pfeffer, 1994; Bewley, 1999; Lazear, 1989), workers sabotaging the performance of others, or workers being directly worse off in utility terms as a result of their being structurally averse to pay inequality (Charness & Rabin, 2002; Fehr & Schmidt, 1999). I believe that these concerns are not pertinent under the setup studied in this paper. The task of assembly is completely autonomous – 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.

To conclude, the data used in this paper is particularly unique and interesting. The detailed production data coming from a firm that operates an on-demand customized manufacturing process under massive changes in labor force size and short-term changes in the monetary incentive structure generate the necessary variations to identify distinctive incentive responses of permanent and gig workers, and establishes a comprehensive understanding of the optimal labor management practices of firms in our changing economy.

4 Stylized Facts

The goal of this section is to lay the foundation for building a structural model. I examine the evolution of production patterns and response to incentives over job experience in the firm separately for gig and permanent workers. In particular, I provide strong evidence that illustrates the differential role of experience in explaining production and incentive responses among permanent workers and gig workers, and additionally, I establish the performance-based scheme as an important instrument in a working environment composed of both permanent and gig workers. The conclusions drawn in this section have formed five facts that guide the model construction and provide its underlying assumptions.

The estimation model in Eq. 1 seek to explore the relationship between the 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, which include worker type (permanent or gig), job experience, and gender:

$$\begin{aligned}
\log(Y_{id}) = \gamma_0 &+ \beta_{\text{Gig}}\text{Gig}_i + \beta_{\text{PP}}\text{PP}_d + \beta_{\text{F}}\text{Female}_i + \beta_{\text{Exp.}}\text{Exp.}_{id} + \beta_{\text{Exp.}^2}\text{Exp.}^2_{id} \\
&+ \beta_{\text{Exp.}\times\text{Gig}}\left(\text{Exp.}\times\text{Gig}\right)_{id} + \beta_{\text{Exp.}^2\times\text{Gig}}\left(\text{Exp.}^2\times\text{Gig}\right)_{id} \\
&+ \beta_{\text{Exp.}\times\text{PP}}\left(\text{Exp.}\times\text{PP}\right)_{id} + \beta_{\text{Exp.}^2\times\text{PP}}\left(\text{Exp.}^2\times\text{PP}\right)_{id} \\
&+ \beta_{\text{PP}\times\text{Gig}}\left(\text{PP}\times\text{Gig}\right)_{id} + \beta_{\text{PP}\times\text{Exp.}\times\text{Gig}}\left(\text{PP}\times\text{Exp.}\times\text{Gig}\right)_{id} + \beta_{\text{PP}\times\text{Exp.}^2\times\text{Gig}}\left(\text{PP}\times\text{Exp.}^2\times\text{Gig}\right)_{id} \\
&+ \alpha_1\text{Demand}_d + \alpha_2\text{Demand}^2_d + \beta_1\left(\#\text{ of Workers}\right)_d + \beta_2\left(\#\text{ of Workers}\right)^2_d \\
&+ \delta_i + C_{id} + \varepsilon_{id},
\end{aligned} \tag{1}$$

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	-0.156**	-0.362**		-0.329**	-0.295**	
	(0.0108)	(0.0212)		(0.0215)	(0.0231)	
Performance Pay=1				0.138**	-0.261**	-0.113
				(0.0159)	(0.119)	(0.109)
Performance Pay=1 × Gig Worker=1					0.530**	0.168*
					(0.128)	(0.0769)
Performance Pay=1 × Gig Worker=1 × Experience					-0.146**	0.0776
					(0.0583)	(0.0827)
Performance Pay=1 × Gig Worker=1 × Experience^2					0.0120	-0.0175
					(0.0147)	(0.0196)
Gig Worker=1 × Experience		0.326**	0.449**	0.287**	0.279**	0.352**
		(0.0193)	(0.0235)	(0.0198)	(0.0221)	(0.0381)
Gig Worker=1 × Experience^2		-0.0633**	-0.0929**	-0.0568**	-0.0526**	-0.0854**
		(0.00566)	(0.00706)	(0.00568)	(0.00642)	(0.00214)
Performance Pay=1 × Experience					0.0380**	0.0152
					(0.0138)	(0.0132)
Performance Pay=1 × Experience^2					-0.000544**	-0.000111
					(0.000266)	(0.000259)
Female=1	0.0850**	0.0578**		0.0624**	0.0660**	
	(0.0139)	(0.0138)		(0.0138)	(0.0138)	
Experience		-0.00301	0.00500	-0.00232	-0.00250	-0.0137
		(0.00228)	(0.00780)	(0.00227)	(0.00234)	(0.00715)
Experience^2		0.000131**	-0.000976**	0.000114**	0.000137**	-0.000660**
		(0.0000544)	(0.000316)	(0.0000542)	(0.0000573)	(0.000235)
Product Demand	0.0687**	0.0700**	0.0441**	0.0890**	0.0872**	0.0386**
	(0.00740)	(0.00723)	(0.00660)	(0.00753)	(0.00752)	(0.0132)
Product Demand^2	-0.00221**	-0.00262**	-0.00118**	-0.00368**	-0.00360**	-0.000498
	(0.000524)	(0.000512)	(0.000460)	(0.000524)	(0.000523)	(0.000944)
Number of Workers	-0.00732**	-0.00785**	-0.00647**	-0.0104**	-0.00999**	-0.00788**
	(0.00113)	(0.00111)	(0.000983)	(0.00114)	(0.00114)	(0.00268)
Number of Workers^2	0.0000367**	0.0000348**	0.0000394**	0.0000482**	0.0000456**	0.0000537*
	(0.0000101)	(0.00000985)	(0.00000870)	(0.00000992)	(0.00000993)	(0.0000226)
Constant	4.646**	4.714**	4.608**	4.721**	4.648**	4.637**
	(0.0277)	(0.0304)	(0.0313)	(0.0302)	(0.0321)	(0.0611)
Day of the week FE	Yes	Yes	Yes	Yes	Yes	No
Individual FE	No	No	Yes	No	No	Yes
N	8179	8179	8150	8179	8179	5830
R^2	0.0929	0.137	0.408	0.144	0.153	0.393

Notes: : The table reports estimates for worker production using Eq 1. The models include as controls day of the week the fixed effect, holiday dummy, repeated employment dummy, and employment in several departments dummy . Product demand is measured in thousands, and experience is measured in 30 days.* $p < 0.10$,** $p < 0.05$

where Gig_i denotes an indicator for whether worker i is a gig worker, and PP_i denotes an indicator for days in which the incentive performance-pay system was instituted. The variable Exp_{id} represents the number of days of job experience worker i has accumulated up to day d. The model incorporates worker experience, Exp_{id}, both linearly and as a square term, in a three-way interaction pattern to capture how the relationship between worker type and pay scheme differ 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, which is 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. δ_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, another variable that indicates the worker's repeated employment status, 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, however, the model in Eq. 1 takes the most comprehensive form, while the results in the table present various specifications of this model. Also, the model is estimated over a logarithmic transformation of the total production, so the estimates represent production percentage change.

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

Column (1) in Table 2 explores the relationship between production and worker type when controlling for gender and work environment as represented by linear and quadratic terms for product demand and the number of workers in a given day. The coefficient of the dummy variable for gig workers indicates that on average, gig workers are 15% less productive than permanent workers, ceteris paribus.

Examination of the work environment variables presents interesting findings. Worker productivity responds positively to an increase in the workload (represented by product demand); however, the increasing productivity response to higher product-demand levels is diminishing. Specifically, 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%. Workforce size estimates show that the production environment is not intrinsically competitive, because worker average production decreases as the number of workers increases, all else equal. This evidence is interpreted to mean that distributing the work among more workers reduces productivity. This effect also exhibits a diminishing pattern, as 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%.

The estimates in Column (1) of Table 2 indicate that worker type plays a crucial rule in determining productivity. This model, however, does not address the effect of experience on the productivity differences between worker types. As gig workers and permanent workers are inherently different and presumably hold different perceptions and goals, overlooking the differential experience effect when comparing their productivity levels could yield biased estimates.

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

Columns (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 2a facilitates interpretation by plotting the results (in unit terms). Moreover, this figure (and others in Figure 2) goes beyond the analysis presented in the table and illustrates worker conditional productivity for high-quality production in addition to total production quantities.

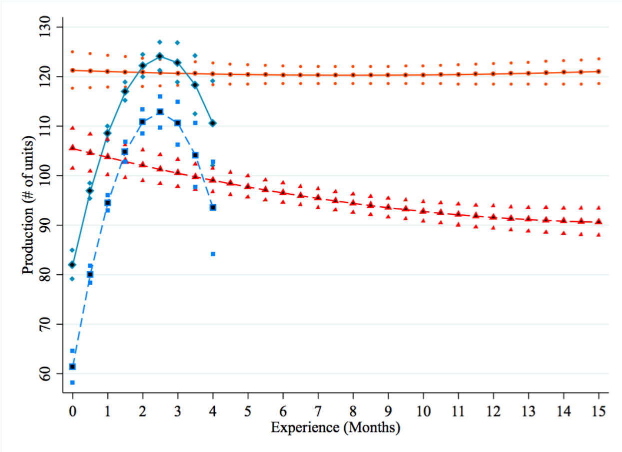
The results presented in Figure 2a indicate distinct production patterns of permanent workers and gig workers, as the role of experience varies dramatically by type. Gig workers hold a parabolic production curve as their tenure progresses. Specifically, the relationship between production and experience is positive until approximately two months of experience, in which the average production of gig workers is higher than the average production of permanent workers. At levels of job experience higher than 2 months, higher job experience is associated with lower productivity. The parabolic pattern persists for the production of high-quality products, with the main difference that gig workers are more productive than permanent workers for a wider range of job experience levels. Unlike gig workers, the productivity levels of permanent workers remain almost unchanged over time when examining total production and decreases over time when focusing on production of high-quality output. Therefore, for these workers experience has little effect on productivity.

Columns (3) in Table 2 expands the analysis by adding individuals' fixed effect that controls for workers' (observed and unobserved) heterogeneity.[23] Figure 2b 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 when controlling for workers' (observed and unobserved) heterogeneity, gig workers are more productive than permanent workers for all (observed) experience levels.

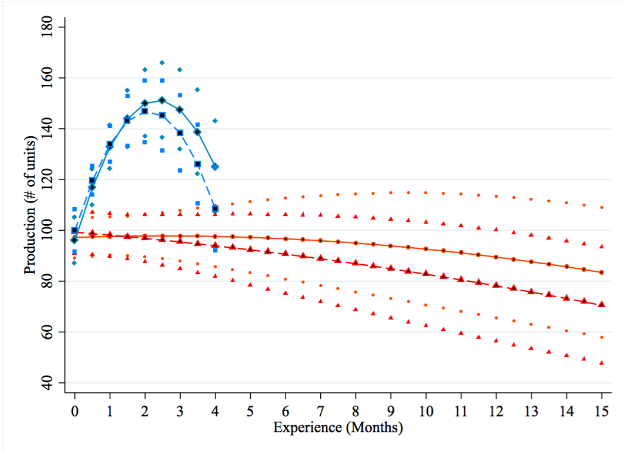
Fact 3: The stringency of hiring standards are different for gig and permanent workers

Figure 2a 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

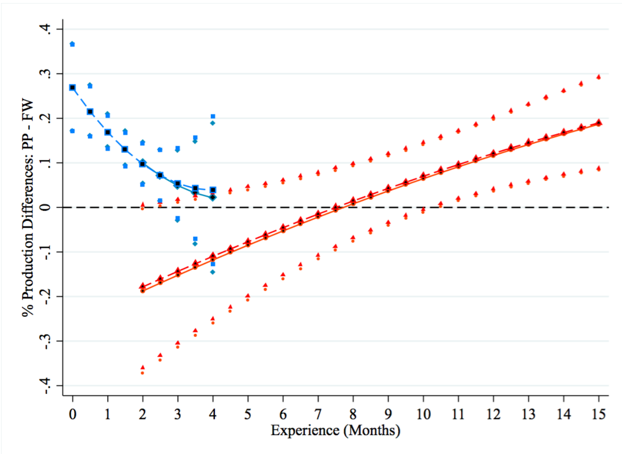
[23]This modeling approach comes at the expense of adding individual-level attributes (particularly gig-worker dummy) that are absorbed into the individual-specific factor.



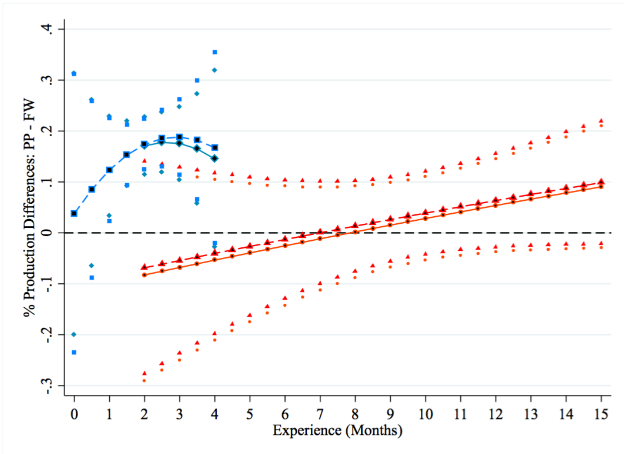
(a) Production, Without Ind. FE



(b) Production, With Ind. FE



(c) Pay Incentive Response, Without Ind. FE



(d) Pay Incentive Response, With Ind. FE

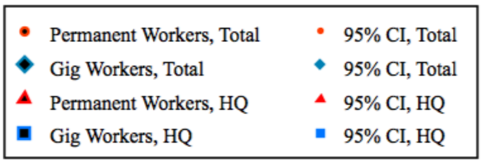


Figure 2: Production and Incentive Response by Experience and Worker Type

Notes: All figures are associated with the results presented in Table 2 derived under different specifications of regression model in Eq. 1. The points in Figures 2a and 2b are associated with Columns (2) and (3) of Table 2 respectively, and represent the predicated total production of workers for various experience levels by workers' types, and its 95% confidence intervals. The points in each Figures 2c and 2d are associated with Columns (5) and (6) of Table 2respectively, and represent the percentage difference in production under flat wage relative to bonus incentive pay regimes for various experience levels by workers' types, and its 95% confidence intervals.

lies in the different hiring processes of these workers. Since permanent workers are 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 the analyses have ignored the influence of incentives on productivity. 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 abrupt, as the performance-pay system was in place for only short period,[24] thus sorting and selection effects that occur when more productive workers self-select into the firm are irrelevant. Lastly, the change in the pay structure was orthogonal to the management method, a fact that further assist 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 2c 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. While the average productivity difference between performance-based pay and a flat wage scheme for gig workers decreases over time, the same difference among permanent workers actually increases over time. Specifically, gig workers with only a 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 higher than zero for job experience of eight months or below, and it is most apparent at higher job experience levels. The average job experience of permanent workers stands at approximately

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

12 months by which production response to incentives stands at 10%.

Since the productivity of many of the workers is observed under both flat wage and performance pay systems, I continue by exploring each worker's response to the performance-based incentive pay scheme by incorporating individual fixed effect factor. The results, presented in Column (6) of Table 2 and Figure 2d, provide strong evidence of the effect of incentive pay on worker's specific productivity.

For permanent workers, both the analysis that captures average effect and the one that examine individual-specific production response yields similar results. Nevertheless, the key difference that emerges from a comparison of these estimation procedures is the higher standard errors associated with the individual-specific specification 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 monotonically decrease, the worker-specific response to incentives presents a parabolic shape. Specifically, the production response to incentive pay increases with job experience and reaches a peak of approximately 20% higher production under performance-based pay relative to flat pay at 60 days of job experience, after which it decreases to 15% at 120 days of experience. This declining response to incentives at higher levels job experience could be interpreted to mean that the power of incentives on worker's productivity reduces over time. That is, high production levels are not sustainable and eventually workers become exhausted and revert back to productions level associated with a flat wage. Additionally, 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.

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 demonstrates an intrinsic behavioral difference among them. In particular, the results illustrate gig workers as holding high motivation to fulfill the purpose they were hired for, with a desire to exploit the most possible gain from their temporary position. At the same time, permanent workers are less-adventurous, seekers of stability who thus remain at sustainable production levels, as their positions are secure and their contracts are not task specific. The notion of personal motivation[25] can explain these differences, whereby gig workers exert higher effort (or refrain from shirking), not only because they get paid, but 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

[25] Amabile (1993) introduced worker's intrinsic motivation into models of labor supply.

in two other ways. First, it could be a result of the *expectations anchoring* effect. Many of the gig workers joined the firm towards 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 they needed the incentive to produce more than the lowest productivity level required. That is, the incentive structure echoes the fact that production above the expected lower bound requires uncompensated effort, and they will receive the same pay for producing below or above the bonus threshold. Second, gig workers' significant productivity response to incentive pay could be explained by their marginal gains from incentives, which is conceptually different from that of the permanent workers. When considering the expected job duration, the marginal wage gain of $1 for a gig worker who works for a total of two weeks is dramatically higher than that of a permanent worker who has already 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 motivates gig workers to increase production.

5 Model

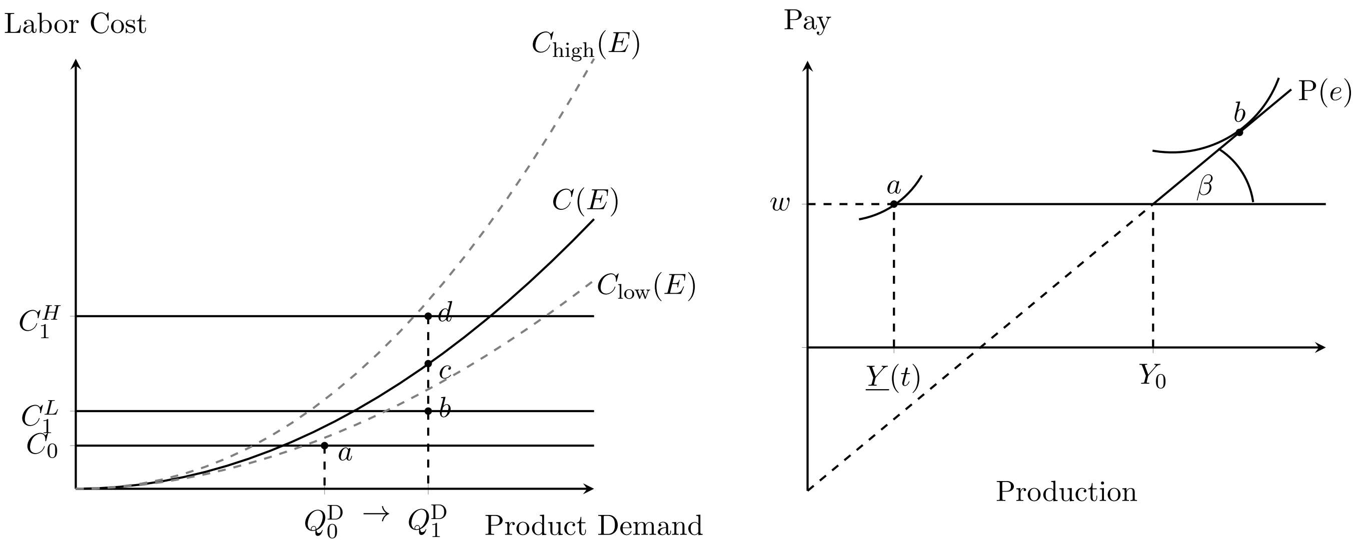
The focus of this analysis is a firm that wishes to satisfy its output demand optimally with the objective of minimizing costs. The firm may choose to administer flat-wage contracts or to incentivize workers to increase their efforts by implementing a performance-based pay system. In the latter scenario, the firm views workers' efforts endogenously and can potentially substitute labor quantity and effort. Therefore, in order to build a comprehensive framework of this problem, both the worker's decision and the firm's decision should be closely analyzed. This section presents the model behind this framework and explains it both heuristically and structurally. Specifically, the first subsection presents the problem in an abstract fashion and outlines its basic principles and concepts. The second subsection presents the structural model and describes the specifications of the problem in detail.

5.1 Simple Model

5.1.1 The Firm's Problem

Figure 3a illustrates the firms problem. The firm starts by facing demand level Q_0^{D}, which is associated with C_0 labor costs under flat wages represented by 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 the new demand level optimally, the firm faces three potential heuristic functions for its new labor costs, which 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



(a) The Firm's Problem

(b) The Worker's Problem

Figure 3: Simple Economic Model

to meet the new demand level Q_1^{D}, while 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 and d. In the first scenario, it is optimal for the firm to be in b, which means that 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 Figure 3a shows, 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 the effort-cost function is low, as represented by $C_{\text{low}}(E)$, the performance-based scheme becomes 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 a 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 as low as possible. As this simple illustration shows, there are many moving parts, and the goal of the structural equilibrium model is to disentangle these forces and find the firm's optimal decision.

5.1.2 The Worker's Problem

A worker needs to 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 $\underline{Y}(t)$. As illustrated in Figure 3b, a worker will maximize her utility at point a and earn a wage of w. Note that this production-level threshold 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.

There are many possible performance-based wage structures. Figure 3b presents a generalized form of a performance-pay scheme with a minimum wage guarantee. The output production threshold, Y_0, is the threshold after which the performance-pay 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 w, which is the same as the pay under the flat-wage scheme. Any production level between $\underline{Y}(t)$ and Y_0 is not associated with financial benefits. Thus, if a worker chooses to produce an amount in this range, the firm earns all the rent.

The worker's effort decision under such a wage structure depends on her personal production capabilities. A worker capable of producing above Y_0 finds it optimal to be anywhere on the upward-sloping line as long as the effort-cost associated with such production is smaller than the gained benefit.[26] A worker that is incapable of producing above Y_0, and chooses to produce anywhere between $\underline{Y}(t)$ and Y_0, has chosen a sub-optimal effort level, because the effort exerted to produce above $\underline{Y}(t)$ is uncompensated. Hence, in this simplified model, optimal effort decisions when performance-based pay is instituted occur at either point a, as under the flat-wage scheme, or at any point on the upward-sloping line, for example, point b. In practice, the data tells us that workers optimize by producing at a level between $\underline{Y}(t)$ and Y_0. These production levels are rationalized by considering workers' inherent motivation differences, an idea introduced in the previous section and will be incorporated directly into the worker's model below.

[26]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.

5.2 Structural Model

5.2.1 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. First, I stresses the fundamental difference between labor and capital. 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 the market-clearing wage and rental rate, respectively. 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 effort the worker actually exerts. Not only are there legal restrictions, but the firm must also usually obtain its 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 occasional. For this reason, I focus on the worker's effort, considering it as 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" simple 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 how much effort workers exert, instead of hours worked, as a new labor-intensive margin.

When thinking about a performance-pay contract from the firm's 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 of this statement is that firms maximize discounted profits, not productivity, and performance-contingent contracts may increase productivity, but may not increase profits. The two main factors that could cause a negative relationship between productivity and profits in the face of performance-based contracts are quality reduction and the distribution of earning gains. Performance-based pay could motivate workers to speed up production by compromising quality, an idea that could explain the well-known phenomenon of "teaching to the test." This issue is of particular concern when output is measurable, but quality is not, and thus workers may increase measured production at the expense of unmeasured quality.[27] For this reason, output quality is integrated as a factor of production into the "new" firm's problem, and as the "traditional" firm's problem can answer questions such as how increasing capital affects labor output, so does the new framework with respect to output quality.

[27] 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. Holmstrm 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.

Additionally, the distribution of earning gains between the firm and the workers is a key aspect in linking productivity and profitability. Assuming that performance-based pay leads to an increase in productivity and consequently to earning gains, the firm is passing along some of these gains to its workforce in the form of the additional compensation paid for their additional production. The parameter that governs this allocation is the incentive rate (denoted by β in Figure 3b). 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 question of which exact performance-pay contract to implement is a challenging one. The new firm's problem incorporates a feature that allows for a performance-based compensation method, and a closer look at the incentive structure is presented in the counterfactual analysis, which constructs the optimal contract.

Last, the new type of firm is characterized by unorthodox labor arrangements – specifically, the hiring of gig workers. Even though it is a prevalent practice among manufacturers today to hire gig workers to perform a specific task for a project or a season, it is rarely examined in the economics literature. As previously shown, gig workers and permanent workers are inherently different in their learning abilities, motivation to produce, and response to incentives. Therefore, decisions regarding the number of workers hired, and the duration and nature of their contracts (gig or permanent), go hand-in-hand with the decision on a remuneration scheme. The new firm's problem incorporates these key aspects by considering not only a decision on the labor-force size, but also on a decision regarding its composition. To the best of my knowledge, this is the first research that applies such an analysis in an empirical setup.

5.2.2 Model Details

The firm faces a labor cost minimization problem in a calendar year. At 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 labor-related decisions made at week t depend on the labor recruitment history up to this point, and particularly on the job experience workers have accumulated. Additionally, worker type plays a crucial role in the firm's decision, because the hiring nature is different. While gig workers are hired for a predetermined short-term period, permanent worker are hired based on a long-term contract with undetermined length.

The pay-scheme type decision is binary – PP or FW – and thus the problem does not impose optimality of the wage structure. That is, at the first phase, 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 at first is thus to solve the problem and identify the parameters that govern the firm's labor and pay scheme decisions. Doing so allows me to construct a reliable model that describes workers' response to the existing incentives and isolate it from other factors such as the working environment and personal attributes. Only after the model is validated and estimated can one turn to computing the optimal incentive structure. An analysis that solves for the optimal contract structure is presented in the counterfactual section.

Permanent Workers: Costs and Structure

Starting with permanent type, the firm decides on the number of permanent workers it hires at time t based on the number of such workers employed in the previous period P_{t-1}. Additionally, 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 in the branch, defined by r_t. Even though one would preferably like to account for job experience as a continuous measure to capture the subtle relationship between experience and productivity, doing so is impossible due to the "curse of dimensionality." Hence, to make the problem tractable, 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. Knowing P_{t-1} and TC_{t-1} the actual decision the firm needs to make is regarding the number of new permanent workers it chooses to hire, denoted by P_t^N, as well as the number of permanent workers that the firm chooses to lay off, denoted by P_t^L. The problem is solved under the assumption that the firm's layoff decision is 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. The cost of hiring permanent workers differs from the cost of laying a worker off, and thus the firm must consider P_t^N and P_t^L as separate alternative in the choice set.

The firm experiences a natural separation rate of its permanent labor force. A permanent worker might terminate her job for exogenous reasons with a separation rate denoted by μ. Combining all these components, the law of motion of permanent workers at time t is defined by

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

Permanent workers are hired based on a long-term contract with an undetermined length. The firm thus needs to consider both recruiting and layoff costs of such workers. Let R^P be the recruiting costs of permanent workers and L^P the costs associated with their job termination. 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. In practice, the firm needs to decide only on the number of new gig workers it wishes to hire. As the availability of these workers changes frequently, I assume that gig workers are selected from an existing pool of temporary workers at the beginning of every week, denoted by t. Thus, the job experience of gig workers evolves weekly and depends on the workers' actual days of experience in the firm, which are denoted by X_{it}. This feature is guided by empirical evidence that established strong linkage between the production skills and response to incentive of gig workers' and their job experience.

I further assume that gig workers' contract length is one week, so that the number of new workers at the beginning of week t N_t^G, is equal the total number of gig workers at this week G_t,

$$G_t = NG_t.$$

Also, I 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

Wages are an important component of the firm's labor-cost minimization problem. 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 effort is endogenously incorporated into the firm's problem through the incentive compatibility constraint defined as follows:

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

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: $z \in \{PP, FW\}$. 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 its workers' effort response 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 expectation is overtaken by workers' idiosyncratic shock (a detailed explanation about the workers problem is presented below). 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, which are denoted by X_{it}.

Total Labor Costs

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

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

Product Demand

On-demand customized production systems are committed to production in minimum time, and thus the firm operates under a strict demand constraint. Labor management decisions of such firms are usually based on product demand forecasts that rely on seasonal patterns common among such product manufacturers. However, the main challenge of operating this type of system arises from the unknown daily product demand shocks, which also require immediate production response.

To incorporate these features into the problem, I use the demand observed in 2015 as a benchmark level and add to it a fixed percentage increase to reflect the average product demand observed in 2018. Let D_t denoted the weekly product demand resulting from these calculations. I then integrate a weekly demand fluctuations represented by an uncertainty shock ζ_t, such that $\zeta_t \sim \mathcal{N}(0, \sigma_D)$ such that the shocks are assumed to be serially independent where σ_D is also derived from the data observed in the year 2015.

Aggregating this information, the firm's demand 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 multinomial dynamic stochastic discrete choice problem, in which it seeks to minimize labor cost. At each week t, subject to an output demand shock, the firm solves for 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 – that is $z \in \{FW, PP\}$. 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.

Let K be all the possible alternatives defined by the vector (P_t^N, P_t^L, z). 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 alternatives are mutually exclusive; i.e. $\sum_k d(t) = 1$.

The firm's dynamic problem can be written as

$$V_t\left(P_{t-1}, TC_{t-1}\right) = \max_{k \in K(t)} \left[V^k\left(P_t, TR_t\right)\right],$$

$$V_t^k\left(P_{t-1}, TC_{t-1}\right) = \begin{cases} -\mathrm{Cost}_t^k(P_{t-1}, TC_{t-1}) & \\ \quad +\psi\, \mathbb{E}\left[V_{t+1}\left(P_t, TC_t\right) | d_k(t) = 1, P_{t-1}, TR_{t-1}\right] & \text{for } t < T, \\ -\mathrm{Cost}_t^k(P_T, TC_T) & \text{for } t = T, \end{cases}$$

such that,

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

where ψ stands for the discount factor.

5.2.3 The Worker's Effort Decision

As previously discussed, under a flat-wage scheme, one would expect to see that workers set their work habits to meet the firm's minimum standards of performance, $\underline{Y}(t)$, as presented in Figure 3b. In practice, however, the data reveals a pattern whereby workers do not produce a uniform quantity during times when the fixed-wage schedule is in effect, and thus a standard economic model could not predict these variations. In order to explain workers' behavior, one must include the worker's problem features that could generate this deviation, which represents aspects other than monetary gains maximization.

The model incorporates several components that could derive workers' productivity variations: a measure that represents workload tightness (defined as based on the demand and number of workers in a given day), the pay method, as well as observed attributes and personal motivation. The nature of the job examined is such that the production technology of each worker's effort places no externalities on co-workers, hence the productivity of a given worker depends solely on individual effort and no social incentives are assumed. In addition, there are no externalities of a worker's effort on co-workers arising from the compensation scheme either, and thus the pay of a given worker depends only on the worker's own effort.

5.2.4 Model Details

On each given day, d, a risk-neutral worker, i of type $\nu \in \{\text{Gig}, \text{Permanent}\}$ wishes to maximize her utility and decide on the effort she exerts, $E_{i\nu d}$. This decision depends on the days of experience she has accumulated in the job thus far, $X_{i\nu d}$, which evolves daily: $X_{i\nu d} = X_{i\nu d-1} + 1$. Also, the worker takes as given the pay scheme offered by the firm and experiences a daily idiosyncratic productivity shock. The 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 it to between zero and one.

Utility

The worker's utility is defined as follows

$$U_{i\nu d}(E, X) = W_{i\nu d}(E, X) - C_{i\nu d}(E, X),$$

where $W_{i\nu d}(E, X)$ denotes the wage a worker receives, and $C_{i\nu d}(E, X)$ denotes the effort cost associated with this wage. Workers are offered a performance-based wage structure of incentives with a guaranteed minimum of

$$W = \max\Big\{w, w + \beta(Y^{\text{HQ}} - Y_0)\Big\}$$

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^{HQ} is the total of high-quality items produced by worker i on day d.[28]

[28] All of these notations match the description in Figure 3b.

Production Function

Production is a function of effort and experience and thus on given day d, they are coupled to determine a total production of

$$Y_{i\nu d}^{\text{Total}} = f_{i\nu d}(E, X),$$

where $f_1, f_2 > 0$, signifying that more experienced workers need to exert less effort to achieve a given level of output. Workers' production technology takes a multiplicative form, with δ denoting the elasticity of experience and α_{nu} denoting the worker type specific total factor productivity,

$$Y_{i\nu d}^{\text{Total}} = \alpha_\nu E_{i\nu d} X_{i\nu d}^{\delta} \mathrm{e}^{\varepsilon_{i\nu d}}$$

such that $0 < \delta < 1$, and $\alpha_\nu > 0$. 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 accounts for the part of the variation in production across workers that is not explained by experience, effort, or personal production shock. In particular, the total factor productivity parameter is constructed to controls for the worker type as follows:

$$\alpha_\nu = \alpha_g \mathbb{I}\Big\{\text{Gig}_i = 1\Big\} + \alpha_p \mathbb{I}\Big\{\text{Permanent}_i = 1\Big\}.$$

Also, the production process incorporates an *iid* worker-day specific production shock assumed to be normally distributed with unknown variance, σ_ε: $\varepsilon_{i\nu d} \sim N(0, \sigma_\varepsilon)$.

Workers are rewarded only for high-quality output, but workers' total production does not reflect output standards, which could be of high or low quality. Items of low quality are those that do not pass the quality-assurance check and thus must undergo a fixing process in order to be sold. Define $\bar{\rho}_{i\nu d}(E, X)$) to be the probability of worker i, of type ν, on day d to generate low- quality items out of the worker's total production. This probability is a function of the worker's effort and experience, and in particular, the probability is parameterized as follows:

$$\rho_{id}(E, X) = \phi_E E_{id} + \phi_X X_{id},$$

which is standardized to remain in the range between zero and one using *logit* transformation.[29]

The question of who experiences and the effort effect-output quality is an empirical one. On the one hand, one would assume that the more experienced a worker, the less low-quality output the worker produces, so that $\bar{\rho}_{id,E} > 0$. On the other hand, under the assumption that a byproduct of high-effort exertion is a rapid and less-accurate production, low-quality probability is positively

[29] Define $\bar{\rho}_{id}(E, X)$ to be the normalization of $\rho_{id}(E, X)$, which takes the following form:

$$\bar{\rho}_{id}(E, X) = \frac{\exp \rho_{id}(E, X)}{1 + \exp \rho_{id}(E, X)} \cdot \Pr(\text{low quality})$$

where Pr(low quality) equals the average of low-quality production observed in the data across all workers and periods.

related to effort, $\bar{\rho} < 0$. These hypotheses are tested in the model by estimating ϕ_E and ϕ_X. The sign and magnitude of these parameters will reveal the strength and relationship between effort, experience, and the probability of producing low- quality output. Having defined this probability, the total high-quality production of worker i of type ν on day d is defined as

$$Y_{i\nu d}^{\text{HQ}} = (1 - \bar{\rho}_{id}(E, X))\, Y_{id}^{\text{Total}}.$$

Effort Cost

Effort entails cost, and the effort cost function takes the following form:

$$C_{i\nu d}(E, X) = \frac{\kappa_d}{X_{i\nu d}} E_{i\nu d}^{\gamma_\nu} - \eta_\nu E_{i\nu d},$$

which implies that effort cost decreases with experience and increases with effort. The effort-cost curvature parameter γ_ν determines the convexity of the effort cost function for each type of worker, with the restriction of $\gamma_\nu > 1$. The parameter η_ν can be interpreted as workers' self-motivation, with the important restriction of $\eta_\nu > 0$.[30] This restriction guarantees that workers supply positive effort levels under the fixed-wage schedule because the marginal cost of effort is negative at zero effort, an outcome which will be further discussed below. Lastly, κ_d represents a day-specific cost evaluation that aims to capture the production intensity at the firm on a given day, defined by the distribution of workload among workers as follows:

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

which is presented in Figure A.2 in the Appendix.

The Worker's Daily Effort Problem

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

$$\max_E \left[\max\left\{w, w + \beta(Y_{i\nu d}^{\text{HQ}}(E, X) - Y_0)\right\}\right] - \left(\frac{\kappa_d}{X_{i\nu d}} E_{i\nu d}^{\gamma_\nu} - \eta_\nu E_{i\nu d}\right)$$

s.t.

$$\begin{aligned}
&0 \le E_{i\nu d} \le 1 \\
&Y_{i\nu d}^{\text{Total}}(E, X) = \alpha_\nu E_{i\nu d} X_{i\nu d}^{\delta} \mathrm{e}^{\varepsilon_{i\nu d}} \\
&Y_{i\nu d}^{\text{HQ}}(E, X) = \left[1 - \rho_{id}(E, X)\right] Y_{i\nu d}^{\text{Total}} \\
&\mathbb{E}\left[Y_{i\nu d}^{\text{HQ}}\right] \ge \underline{Y}(t),
\end{aligned}$$

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.

[30]Modeling worker utility as a function of their personal motivation is motivated by Amabile (1993) who introduced worker's intrinsic motivation into models of labor supply.

5.3 Model Predictions and Implications

5.3.1 Optimal Effort

Worker's objective function is strictly concave with respect to E_{id} at the relevant effort domain of $E \in [0, 1]$ under both the flat wage and the performance-pay regimes. This property guarantees that workers solve for a unique maximum and implies that first-order conditions are meaningful in explaining workers optimal behavior.

Wage: $\beta = 0$

If the firm sets $\beta = 0$, earnings are then fixed and independent of performance and thus it is not a monetary incentive that leads workers to exert effort to produce above $\underline{Y}(t)$, but rather personal motivation. This factor plays a crucial role in the model, and $\eta_\nu > 0$ is in fact a necessary condition for positive production. This idea is presented clearly when solving for the optimal-effort level in the flat-wage scenario, denoted by $E_{id}^{*\mathrm{FW}}$, which yields the following solution:

$$\frac{\kappa_d}{X_{i\nu d}}\gamma\left(E_{i\nu d}^{*\mathrm{FW}}\right)^{(\gamma_\nu - 1)} = \eta_\nu. \tag{2}$$

Equation 2 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 tightness on that given day. Also, the optimal effort under the flat-wage scheme decreases with the workload tightness and increases with experience and personal motivation. The optimal effort is increasing and concave with the effort-cost function curvature, and thus a higher curvature is associated with higher effort, but at a diminishing rate.

5.3.2 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,

$$\beta\Big[1 - \rho_{id}\Big]\alpha X_{id}^{\delta} \mathrm{e}^{\varepsilon_{id}} \mathbb{I}_{\mathrm{PP}} + \eta_i = \frac{\kappa_d}{X_{id}}\gamma\left(E_{id}^{*\mathrm{PP}}\right)^{(\gamma - 1)} + \beta\rho_{id,E}\alpha\left(E_{id}^{*\mathrm{PP}}\right) X_{id}^{\delta} \mathrm{e}^{\varepsilon_{id}} \mathbb{I}_{\mathrm{PP}}, \tag{3}$$

where ρ_{id} and $\rho_{id,E}$ 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, and $Y_{id}^{HQ} > Y_0$ when the incentive scheme is in effect. If $\mathbb{I}_{\mathrm{PP}} = 1$, the left-hand side represents the sum of the marginal monetary and non-monetary benefits of production. That is, the marginal monetary benefits of producing above $\underline{Y}(t)$ are 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 personal motivation parameter, η_ν. The right-hand side is the sum of the marginal cost of production and the cost of producing low-quality items above $\underline{Y}(t)$, as workers are not compensated for the production of these items. If $\mathbb{I}_{\text{PP}} = 0$, then the optimal effort solution under the performance-pay and flat-wage schemes coincide.

Effort Cost Function

Worker's effort-cost function and parameters are constrained to satisfy several key features. 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 given that production entails a cost, we might reasonably expect that a utility-maximizing agent will find it optimal to choose $E_{i\nu d} = 0$. The effort-cost production function is constructed to preclude this possibility by setting $\eta_\nu > 0$. Specifically, the marginal cost of effort is given by

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

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

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

This result ensures that employees supply positive effort levels under a fixed wage.

Moreover, two other features of the effort-cost function is that first, it is convex with respect to effort, $C_{id,EE} > 0$, and (all other things being equal) more experienced workers face a lower marginal effort-cost, $C_{id,EX} < 0$.

5.3.3 Participation Constraint

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

Denote workers' alternative utility by $\underline{u}$ and define the participation constraint by equating it to the expected utility under flat wage as follows

$$\mathbb{E}\left[U(E_i^{*\text{FW}}, X_i)\right] \geq \underline{u}.$$

By writing this condition explicitly, one can identify the guaranteed wage level w,

$$w = \underline{u} + \mathbb{E}\left[C_{id}^{*\text{FW}}\right], \tag{4}$$

such that C_{id}^{*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.

5.3.4 Indirect Utility

The indirect utility function under the flat-wage scheme takes the form of

$$V^{*\mathrm{FW}} = w - \mathbb{E}\left[C_{id}^{*\mathrm{FW}}\right],$$

and using the results derived from the participation constraint in Eq. 4, I calculate that workers' maximum attainable utility is equal to their outside option value

$$V^{*\mathrm{FW}} = \underline{u}.$$

The indirect utility function under the performance-pay regime is defined as follows:

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

If the incentive threshold is reached, that is $\mathbb{I}_{\mathrm{PP}} = 1$, this equation can be simplified using the results derived from the participation constraint in Eq. 4. Specifically, one gets that the indirect utility function under the performance-pay scheme can be written as:

$$V^{*\mathrm{PP}} = \underline{u} + \beta\,\mathbb{E}\left[Y^{*\mathrm{HQ,PP}} - Y_0\right] - \mathbb{E}\left[C_{id}^{*\mathrm{PP}} - C_{id}^{*\mathrm{FW}}\right], \tag{5}$$

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

The results in Equation 5 have several implications. First, they show that workers' maximal attainable utility is equal to the monetary benefit of producing above Y_0, minus the cost for effort exerted above what they would otherwise choose under the flat-wage regime (on top of the outside option value, $\underline{u}$). Second, based on Equation 5, one can infer that the maximal attainable utility decreases with the daily workload tightness κ_d, increases with personal motivation η_ν and with the incentive rate, β. Lastly, examination of the components of $V^{*\mathrm{PP}}$ as presented in Eq. 5 distill the forces that stand behind the effort-decision problem when incentives are available. A worker chooses to exert effort 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,

$$\beta\,\mathbb{E}\left[Y^{*\mathrm{HQ,PP}} - Y_0\right] > \mathbb{E}\left[C_{id}^{*\mathrm{PP}} - C_{id}^{*\mathrm{FW}}\right], \tag{6}$$

and in this case, workers are strictly better off under the performance-pay regime

$$V^{*\mathrm{PP}} > V^{*\mathrm{FW}}.$$

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

$$V^{*\mathrm{PP}} = \underline{u} - \mathbb{E}\left[C_{id}^{*\mathrm{PP}} - C_{id}^{*\mathrm{FW}}\right], \tag{7}$$

where the second term of Equation 7 must be equal to zero for the worker to accept this job.

5.3.5 Expected Earnings

The expected earnings in the performance-pay scenario are higher than the expected earnings in the flat-wage scenario only if a worker is capable of reaching Y_0. From the worker's perspective, the earnings when incentives are available should be compensation for the additional cost burden incurred for producing not above Y_0, but rather above $\underline{Y}(t)$. This idea can be viewed in the following performance-based pay expected earning equation, which illustrates the additional pay above w that should be given to compensate a worker for her effort:

$$\mathbb{E}(W|\mathrm{PP}) = w + \beta\, \mathbb{E}\left[Y^{*\mathrm{HQ,PP}} - Y_0\right]. \tag{8}$$

6 Identification

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

$$\begin{aligned}\Theta_W &= (\alpha_p, \alpha_g, \gamma_p, \gamma_g, \eta_p, \eta_g, \delta, \phi_E, \phi_X, \sigma_\varepsilon)\\ \Theta_F &= \left(\mu, R^P, L^P, R^G\right)\end{aligned}$$

where W 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 such a closed-form mapping, in the following section I develop the model's equations and specify the sources of variations in the data that are necessary for identification to ensure that the model's parameters are in fact identified.

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,

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

This solution can be integrated into the production function as follows:

$$Y^*_{i\nu d} = \alpha_\nu \left(\frac{\eta_\nu}{\gamma_\nu}\right)^{\frac{1}{\gamma_\nu-1}} \left(\frac{1}{\kappa_d}\right)^{\frac{1}{\gamma_\nu-1}} X_{id}^{\delta+\frac{1}{\gamma_\nu-1}} \mathrm{e}^{\varepsilon_{id}}.$$

By applying logarithmic transformation, one obtains

$$\log\left(Y^{*FW}_{i\nu d}\right) = \alpha_0 - \left(\frac{1}{\gamma_\nu - 1}\right)\log\left(\kappa_d\right) + \left(\delta + \frac{1}{\gamma_\nu - 1}\right)\log(X_{id}) + \varepsilon_{id}, \tag{9}$$

which can be written separately for each worker type, such that

$$\alpha_0 = \left[\log(\alpha_\nu) + \left(\frac{1}{\gamma_\nu - 1}\right)\log\left(\frac{\eta_\nu}{\gamma_\nu}\right)\right].$$

From Equation 9 one can see that variation in X_{id} identifies the experience elasticity δ, and that the variation in the workload tightness parameter κ_d identifies the curvature of the effort-cost function.

I now turn to examine the probability for low-quality output, which is defined by the parameters ϕ_X and ϕ_E as follows

$$\mathbb{P}(\text{Low Quality})_{id} = \frac{\exp\left(\phi_E E^*_{id} + \phi_X X_{id}\right)}{1 + \exp\left(\phi_E E^*_{id} + \phi_X X_{id}\right)}.$$

This expression can be rewritten in the following odds ration form

$$\frac{\mathbb{P}(\text{Low Quality})_{id}}{\mathbb{P}(\text{High Quality})_{id}} = \exp\left(\phi_E E^*_{id} + \phi_X X_{id}\right). \tag{10}$$

By plugging in the optimal effort under a flat wage (or when $\mathbb{I}_{PP} = 0$) in Equation 10 and then applying a logarithmic to one derives the following result

$$\log\left(\frac{\mathbb{P}(\text{Low Quality})}{\mathbb{P}(\text{High Quality})}\right)_{i\nu d} = \phi_E \left(\frac{\eta_\nu}{\gamma_\nu}\right)^{\frac{1}{\gamma_\nu - 1}} \left(\frac{X_{i\nu d}}{\kappa_d}\right)^{\frac{1}{\gamma_\nu - 1}} + \phi_X X_{i\nu d}. \tag{11}$$

Equation 11 indicates that variation in $X_{i\nu d}$ identifies ϕ_X.

So far the equations have been examined under the assumption that the bonus performance-pay regime was either not offered, $\beta = 0$, or not attained $\mathbb{I}_{PP} = 0$. Now consider the first-order condition when $\mathbb{I}_{PP} = 1$. The first-order condition of the problem can be written as

$$\begin{aligned}
&\beta\left[\frac{1}{1 + \exp\left(\phi_E E^*_{i\nu d} + \phi_X X_{i\nu d}\right)}\right] \alpha_\nu X^{\delta}_{i\nu d} \mathrm{e}^{\varepsilon_{id}} + \eta_\nu \\
&= \frac{\kappa_d}{X_{i\nu d}} \gamma_\nu \left(E^*_{i\nu d}\right)^{(\gamma_\nu - 1)} + \beta\left[\frac{\phi_E \exp\left(\phi_E E^*_{i\nu d} + \phi_X X_{i\nu d}\right)}{\left(1 + \exp\left(\phi_E E^*_{i\nu d} + \phi_X X_{i\nu d}\right)\right)^2}\right] \alpha_\nu \left(E^*_{i\nu d}\right) X^{\delta}_{i\nu d} \mathrm{e}^{\varepsilon_{id}},
\end{aligned} \tag{12}$$

which can be further simplified using the result in Equation 10. Since the optimal effort E^* is a unique solution of the problem, in a case in which $\mathbb{I}_{PP} = 1$, one could solve for it and incorporates it into the production function to obtain the estimated production of workers i of type ν at day d, defined by:

$$Y^{*\text{PP}}_{i\nu d} = \alpha_\nu E^{*\text{PP}}_{i\nu d} X^{\delta}_{id} \mathrm{e}^{\varepsilon_{id}}. \tag{13}$$

The estimation procedure then relies on the parameters to match the estimated production levels with those observed in the real data. Similar procedures are applied on production under the flat wage, the low-quality production quantity, and the performance-incentive wage gains. That is, taken together, Equations. 8, 9, 10, 12 and 13 provide the sets of equation that guarantee identification of the parameters of the worker's models.

To further validate identification of the parameters, I apply a one-dimensional sensitivity analyses on each structural parameter and check if moments selected for the estimation procedure vary substantially with changes in parameter values. Moreover, I apply 10 Monte Carlo simulations based on fictitious pre-known structural parameters values and uncovered these parameters precisely through the estimation procedure.

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 low-quality production as auxiliary models to tightly link the structural parameters to the empirical findings.[31] 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.

In practice, 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 $\Theta_W = (\alpha_p, \alpha_g, \gamma_p, \gamma_g, \eta_p, \eta_g, \delta, \phi_E, \phi_X, \sigma_\varepsilon)$ given the realization of the idiosyncratic production shock, and a set of daily and personal observables. At 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 optimal set of parameters $\hat{\Theta}_W$ is the one that minimize the distance between the auxiliary parameters estimated on the actual data and the auxiliary parameters estimated on the simulated data. Formally, I choose $\hat{\Theta}_W = \left(\hat{\alpha}_p, \hat{\alpha}_g, \hat{\gamma}_p, \hat{\gamma}_g, \hat{\eta}_p, \hat{\eta}_g, \hat{\delta}, \hat{\phi}_E, \hat{\phi}_X, \hat{\sigma}_\varepsilon\right)$ such that

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

where W is a symmetric and positive semi-definite weighting matrix.[32]

The choice of the auxiliary parameters allows for rather transparent identification of the model's structural parameters. All parameters of the auxiliary model contribute on varying levels to the estimation of the structural parameters depending on the relationship between the auxiliary and structural parameters.[33]

---

[31] These models are all guided by the identification procedure previously outlined to represent Eqs. 8, 9, 10, 12 and 13 for both worker's type.

[32] W is estimated in two steps. At the first step, it is set to be equal to the inverse of a diagonal matrix with the parameter's standard errors of the auxiliary model on the main diagonal. The second step calculates the variance-covariance matrix of the simulated auxiliary parameters, ψ_{sim},

$$\hat{\mathrm{W}} = \frac{1}{L}\sum_{l}^{L} \left(\psi^l_{\text{sim}}(\sigma^1_\varepsilon) - \frac{1}{L}\sum_{l}^{L} \psi^l_{\text{sim}}(\sigma^2_\varepsilon)\right) \cdot \left(\psi^l_{\text{sim}}(\sigma^1_\varepsilon) - \frac{1}{L}\sum_{l}^{L} \psi^l_{\text{sim}}(\sigma^2_\varepsilon)\right)'$$

where σ^j_ε, $j = 1, 2$ are different sets of L realizations of the idiosyncratic production shock, and L is equal to 2,000.

[33] To illustrate this idea, Figure A.3 in the Appendix outlines the relationship between a parameter of the structural model and parameters of the auxiliary model with the largest identification contribution, obtained by simulation.

7.2 Method of Simulated Likelihood and Interpolation

I solve the firm's dynamic problem by applying a simulated maximum likelihood procedure. Let the probability that the firm is observed to choose alternative k_t at week t be defined as

$$\mathcal{P}\left(k_t|\Omega_t^d\right) = \mathcal{P}\left(\max_j \left[V_{jt}(\Omega_t)\right]\right),$$

where the probabilities are calculated using the Kernel Smoothed Frequency Simulator proposed McFadden (1989). I then define the likelihood function based on these probabilities as follows

$$\mathcal{P}\left(k_1, ..., k_T|\Omega_1^d\right) = \prod_{t=1}^{T} \mathcal{P}\left(k_t|\Omega_t^d\right).$$

All of the state variables of the firm's dynamic problem are discrete, and the state space dimension is therefore finite. The state space, however, although finite, is huge. This is because in order to keep track of the workers' tenure category each week, it is necessary to keep track of the complete sequence tenure paths. To illustrate, when hiring 20 permanent workers, the number of possible tenure sequences is 209. The state space increases exponentially with the number of permanent workers, which makes a full solution of the dynamic programming problem infeasible, leaving aside the iterative process necessary for estimation. Therefore, to deal with this issue I use simulation and interpolation proposed by Keane and Wolpin (1994)

8 Results and Model Fit

8.1 The Worker's Problem

Table 3 presents the estimates of the worker's problem structural model. These estimates reveal the sources of production differences between gig and permanent workers. First, the personal motivation of gig workers is revealed to be significantly higher than the personal motivation of permanent workers. Second, there is a significant difference in the effort-cost curvature parameter of each worker type. While the effort-cost function of permanent workers is close to linear, the effort-cost curvature of gig workers is higher than quadratic. These results means that for the same effort level (all other things equal), gig workers experience higher effort costs than permanent workers; however, a higher motivation to produce compensates these workers for their higher exertion effort. Also, the results show that while the total factor productivity of gig workers is higher than that of permanent workers, the difference is relatively small.

The estimate in Table 3 reveals the value and sign of the parameters that govern production quality. The results indicate the following relationships between effort and output quality: a higher effort a worker exerts, the less low-quality item she will produce. This result can be interpreted as higher effort means also higher attention to production quality. At the same time, the results

That is, it shows how a change in the value of the structural parameter varies the value of the coefficients generated by the auxiliary model.

Table 3: Labor Supply: Estimates of Structural Parameters

Parameters Description	Symbol	Estimate	SD
Total Factor Productivity			
Gig Worker	α_g	48.26	4.1
Permanent	α_p	40.63	5.18
Personal Motivation			
Gig Worker	η_g	300.023	7.5
Permanent Worker	η_p	194.91	5.7
Effort Cost Convexity			
Gig Worker	γ_g	2.40	0.26
Permanent Worker	γ_p	1.3	0.18
Experience Elasticity	δ	0.15	0.003
Effort Effect on Low-quality Production	ϕ_E	-38.57	5.95
Experience Effect on Low-quality Production	ϕ_X	10.69	1.95
Idiosyncratic production shock	σ_ε	0.73	0.1

show that workers with higher experience tend to produce more low-quality items, although this relationship is weaker than the effort-quality link.

Figure 4: Histograms of the Optimal Effort Solution

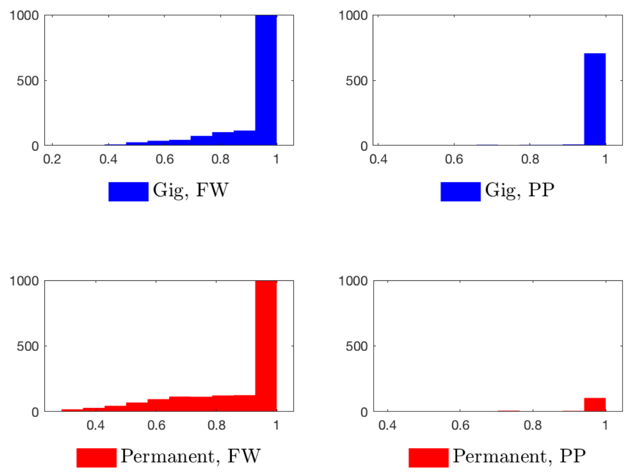


Figure 4 presents histograms of the optimal effort by pay scheme. On average, the optimal effort of gig workers under the bonus performance-pay scheme is higher than the optimal effort of permanent workers under the same pay scheme. Moreover, the optimal effort of gig workers when a flat wage is offered is lower than the optimal effort of permanent workers under the same pay scheme. This interesting result supports the previously discussed expectation-anchoring idea.

Figure 5: Production Densities

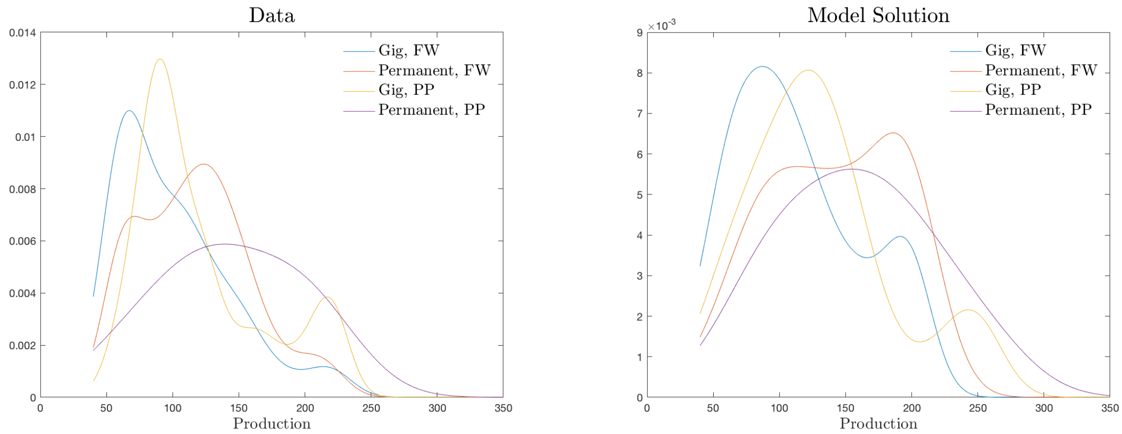


Fig. 5 depicts the production densities under both pay schemes for each of the worker types secretly. A comparison between the figures indicates that the model captures the main features revealed in the data.

8.2 The Firm's Problem

Table 4 shows the labor-demand side estimates of the structural parameters obtained from the firm's problem. The results indicate that the hiring cost of gig workers is lower than the hiring and layoff costs associated with permanent workers. It also shows that the recruiting costs of permanent workers are higher than their layoff costs.

Table 4: Labor Demand: Estimates of Structural Parameters

Parameters Description	Symbol	Estimate	SD
Separation rate	μ	0.07	0.001
Recruiting permanent worker	R^P	296.81	10.88
Laying off permanent worker	L^P	230.57	9.14
Recruiting gig worker	R^G	112.67	8.44

These costs and their relative magnitudes should be examined relative to workers' average productivity and especially in relation to their productivity when instituting the bonus performance-pay scheme. Specifically, since the results from the previous sections have shown that gig workers produce higher amounts under the bonus performance-pay scheme, the fact that their hiring cost is lower than the hiring cost of permanent workers implies that the optimal solution of the firm under demand uncertainty is one that incorporates a combination of these elements: hiring gig workers and instituting the bonus performance-pay scheme. This idea is discussed in the following section that examines counterfactual scenarios of the firm's behavior and seeks to find the optimal incentive structure and hiring schedule that minimize firm labor costs given the demand forecast.

9 Policy Analysis

Back-of-the-envelope calculations show that the firm could reduce labor costs during peak seasons by 22% by both hiring gig workers and offering bonus incentive pay. This striking results raises the question of whether greater improvement can be made. Specifically, since the firm management chose the structure of the bonus pay in a trial and error fashion, it is possible that other pay structures could have generated a more profitable response. To answer this question, I use the estimated labor-cost values from the firm's problem and insert it into an equilibrium framework with alternative bonus incentive contracts. Specifically, I consider two approaches. The first is a utilitarian approach, which focuses on the firm's profitability and seeks to find the optimal bonus incentive structure that minimize the firm's labor costs given the output demand uncertainty. The second approach follows a central planner perspective, as it aims to find the optimal bonus incentive structure under the constraints by which worker utility is greater than or equal to the one gained in the original settings.

To be completed.

10 Concluding Remarks

To be completed.

References

Akerlof, G. A., & Kranton, R. E. (2000). Economics and identity. *The Quarterly Journal of Economics*, *115*(3), 715–753.

Allon, G., Cohen, M., & Sinchaisri, P. (2018). The impact of behavioral and economic drivers on gig economy workers. *Available at SSRN 3274628*.

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.

Anderson, E. (1985). The salesperson as outside agent or employee: a transaction cost analysis. *Marketing science*, *4*(3), 234–254.

Arrow, K. J. (1971). Insurance, risk and resource allocation. *Essays in the theory of risk-bearing*, 134–143.

Baker, K. R. (1976). Workforce allocation in cyclical scheduling problems: A survey. *Journal of the Operational Research Society*, *27*(1), 155–167.

Bandiera, O., Barankay, I., & Rasul, I. (2005). Social preferences and the response to incentives: Evidence from personnel data. *The Quarterly Journal of Economics*, *120*(3), 917–962.

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

Bard, J. F. (2004a). Selecting the appropriate input data set when configuring a permanent workforce. *Computers & Industrial Engineering*, *47*(4), 371–389.

Bard, J. F. (2004b). Staff scheduling in high volume service facilities with downgrading. *Iie Transactions*, *36*(10), 985–997.

Bard, J. F., Binici, C., & desilva, A. H. (2003). Staff scheduling at the united states postal service. *Computers & Operations Research*, *30*(5), 745–771.

Bard, J. F., Morton, D. P., & Wang, Y. M. (2007). Workforce planning at usps mail processing and distribution centers using stochastic optimization. *Annals of Operations Research*, *155*(1), 51–78.

Bard, J. F., & Purnomo, H. W. (2005). Short-term nurse scheduling in response to daily fluctuations in supply and demand. *Health Care Management Science*, *8*(4), 315–324.

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

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.

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).

Billionnet, A. (1999). Integer programming to schedule a hierarchical workforce with variable demands. *European Journal of Operational Research*, *114* (1), 105–114.

Brown, C. (1990). Firms' choice of method of pay. *ILR Review*, *43* (3), 165–S.

Cachon, G. P., Daniels, K. M., & Lobel, R. (2017). The role of surge pricing on a service platform with self-scheduling capacity. *Manufacturing & Service Operations Management*, *19* (3), 368–384.

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

Chen, M. K. (2016). Dynamic pricing in a labor market: Surge pricing and flexible work on the uber platform. In *Proceedings of the 2016 acm conference on economics and computation* (pp. 455–455).

Conlon, E. J., & Parks, J. M. (1990). Effects of monitoring and tradition on compensation arrangements: An experiment with principal-agent dyads. *Academy of Management Journal*, *33* (3), 603–622.

Connelly, C. E., & Gallagher, D. G. (2004). Emerging trends in contingent work research. *Journal of management*, *30* (6), 959–983.

Corominas, A., Lusa, A., & Pastor, R. (2004). Planning annualised hours with a finite set of weekly working hours and joint holidays. *Annals of Operations Research*, *128* (1-4), 217–233.

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.

DeFillippi, R. J., & Arthur, M. B. (1994). The boundaryless career: A competency-based perspective. *Journal of organizational behavior*, *15* (4), 307–324.

Delfgaauw, J., & Dur, R. (2007). Signaling and screening of workers' motivation. *Journal of Economic Behavior & Organization*, *62* (4), 605–624.

Delfgaauw, J., Dur, R., Sol, J., & Verbeke, W. (2013). Tournament incentives in the field: Gender differences in the workplace. *Journal of Labor Economics*, *31* (2), 305–326.

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

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

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.

Gabszewicz, J., & Turrini, A. (2000). Workers' skills, product quality and industry equilibrium. *International journal of industrial organization*, *18*(4), 575–593.

Gans, N., & Zhou, Y.-P. (2002). Managing learning and turnover in employee staffing. *Operations Research*, *50*(6), 991–1006.

Gibbons, R. (2005). Incentives between firms (and within). *Management science*, *51*(1), 2–17.

Gillespie, P. (2017). Intuit: Gig economy is 34% of us workforce. *CNN Money*.

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

Gneezy, U., Niederle, M., & Rustichini, A. (2003). Performance in competitive environments: Gender differences. *The Quarterly Journal of Economics*, *118*(3), 1049–1074.

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

Grossman, R. J. (1998). Short-term workers raise long-term issues. *HR Magazine*, *43*(5), 80–89.

Hall, J. V., Horton, J. J., & Knoepfle, D. T. (2018). *Pricing efficiently in designed markets: Evidence from ride-sharing.* Working Paper, https://john-josephhorton. com/papers/uber price. pdf.

Hamermesh, D. S. (1991). *Labor demand: what do we know? what don't we know?* (Tech. Rep.). National Bureau of Economic Research.

Hamermesh, D. S., & Grant, J. (1979). Econometric studies of labor-labor substitution and their implications for policy. *Journal of Human Resources*, 518–542.

Hojati, M., & Patil, A. S. (2011). An integer linear programming-based heuristic for scheduling heterogeneous, part-time service employees. *European Journal of Operational Research*, *209*(1), 37–50.

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

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. (1994). The firm as an incentive system. *The American Economic Review*, 972–991.

Ibrahim, R. (2018). Managing queueing systems where capacity is random and customers are impatient. *Production and Operations Management*, *27*(2), 234–250.

Jarmon, R., Paulson, A. S., & Rebne, D. (1998). Contractor performance: How good are contingent workers at the professional level? *IEEE Transactions on Engineering Management*, *45*(1), 11–19.

Jensen, M. C., & Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of financial economics*, *3*(4), 305–360.

Jurajda, Š., & Münich, D. (2011). Gender gap in performance under competitive pressure: Admissions to czech universities. *American Economic Review*, *101*(3), 514–18.

Katz, L. F., & Krueger, A. B. (2019). The rise and nature of alternative work arrangements in the united states, 1995–2015. *ILR Review*, *72*(2), 382–416.

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.

Kolisch, R., & Heimerl, C. (2012). An efficient metaheuristic for integrated scheduling and staffing it projects based on a generalized minimum cost flow network. *Naval Research Logistics (NRL)*, *59*(2), 111–127.

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

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. (2018). Compensation and incentives in the workplace. *Journal of Economic Perspectives*, *32*(3), 195–214.

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

Li, Y., Chen, J., & Cai, X. (2007). An integrated staff-sizing approach considering feasibility of scheduling decision. *Annals of Operations Research*, *155*(1), 361–390.

Manyika, J., Lund, S., Bughin, J., Robinson, K., Mischke, J., & Mahajan, D. (2016). Independent work: Choice, necessity, and the gig economy. *McKinsey Global Institute*, *2016*, 1–16.

Marasco, A., & Romano, A. (2011). A mathematical model for the management of a service center. *Mathematical and Computer Modelling*, *53*(9-10), 2005–2014.

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.

Milner, J. M., & Pinker, E. J. (2001). Contingent labor contracting under demand and supply uncertainty. *Management Science*, *47*(8), 1046–1062.

Misra, S., & Nair, H. S. (2011). A structural model of sales-force compensation dynamics: Estimation and field implementation. *Quantitative Marketing and Economics*, *9*(3), 211–257.

Ors, E., Palomino, F., & Peyrache, E. (2013). Performance gender gap: does competition matter? *Journal of Labor Economics*, *31*(3), 443–499.

Oyer, P., & Schaefer, S. (2005). Why do some firms give stock options to all employees?: An empirical examination of alternative theories. *Journal of financial Economics*, *76*(1), 99–

133.

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

Petriglieri, G., Ashford, S. J., & Wrzesniewski, A. (2019). Agony and ecstasy in the gig economy: Cultivating holding environments for precarious and personalized work identities. *Administrative Science Quarterly*, *64*(1), 124–170.

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.

Polivka, A. E., & Nardone, T. (1989). On the definition of contingent work. *Monthly Lab. Rev.*, *112*, 9.

Robles, B. J., & McGee, M. (2016). Exploring online and offline informal work: findings from the enterprising and informal work activities (eiwa) survey.

Ross, S. A. (1973). The economic theory of agency: The principal's problem. *The American economic review*, *63*(2), 134–139.

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.

Stratman, J. K., Roth, A. V., & Gilland, W. G. (2004). The deployment of temporary production workers in assembly operations: a case study of the hidden costs of learning and forgetting. *Journal of Operations Management*, *21*(6), 689–707.

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

Techawiboonwong, A., Yenradee, P., & Das, S. K. (2006). A master scheduling model with skilled and unskilled temporary workers. *International Journal of Production Economics*, *103*(2), 798–809.

Thompson, G. M., & Goodale, J. C. (2006). Variable employee productivity in workforce scheduling. *European journal of operational research*, *170*(2), 376–390.

Valls, V., Pérez, Á., & Quintanilla, S. (2009). Skilled workforce scheduling in service centres. *European Journal of Operational Research*, *193*(3), 791–804.

Van den Bergh, J., Beliën, J., De Bruecker, P., Demeulemeester, E., & De Boeck, L. (2013). Personnel scheduling: A literature review. *European journal of operational research*, *226*(3), 367–385.

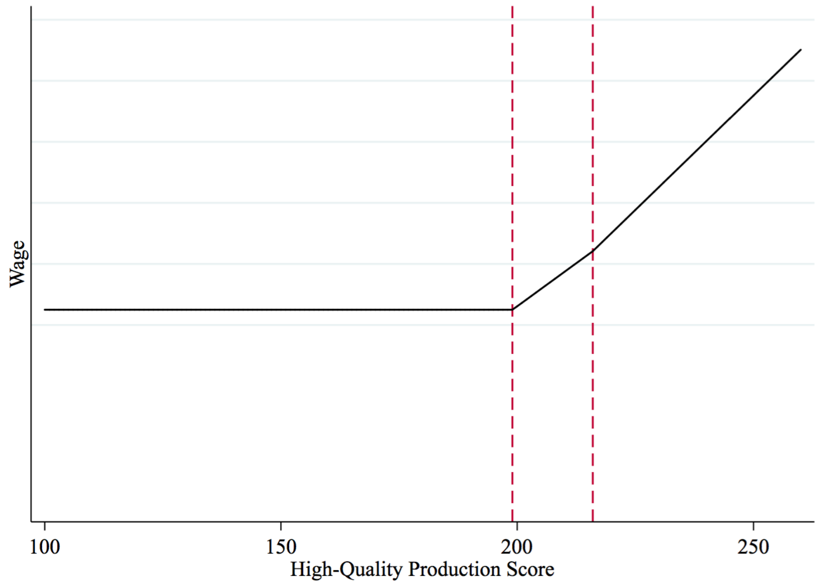
Wild, B., & Schneewei, C. (1993). Manpower capacity planning—a hierarchical approach. *International Journal of Production Economics*, *30*, 95–106.

Wilson, R. (1968). The theory of syndicates. *Econometrica: journal of the Econometric Society*, 119–132.

Appendix

A Figures

Figure A.1: 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.2: Daily Competitive Climate

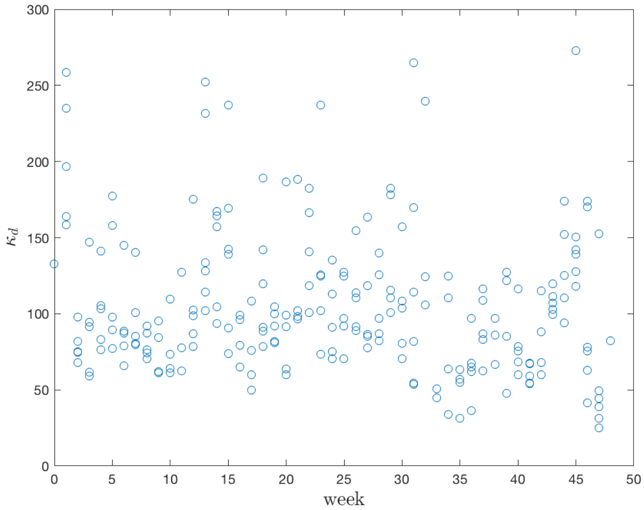
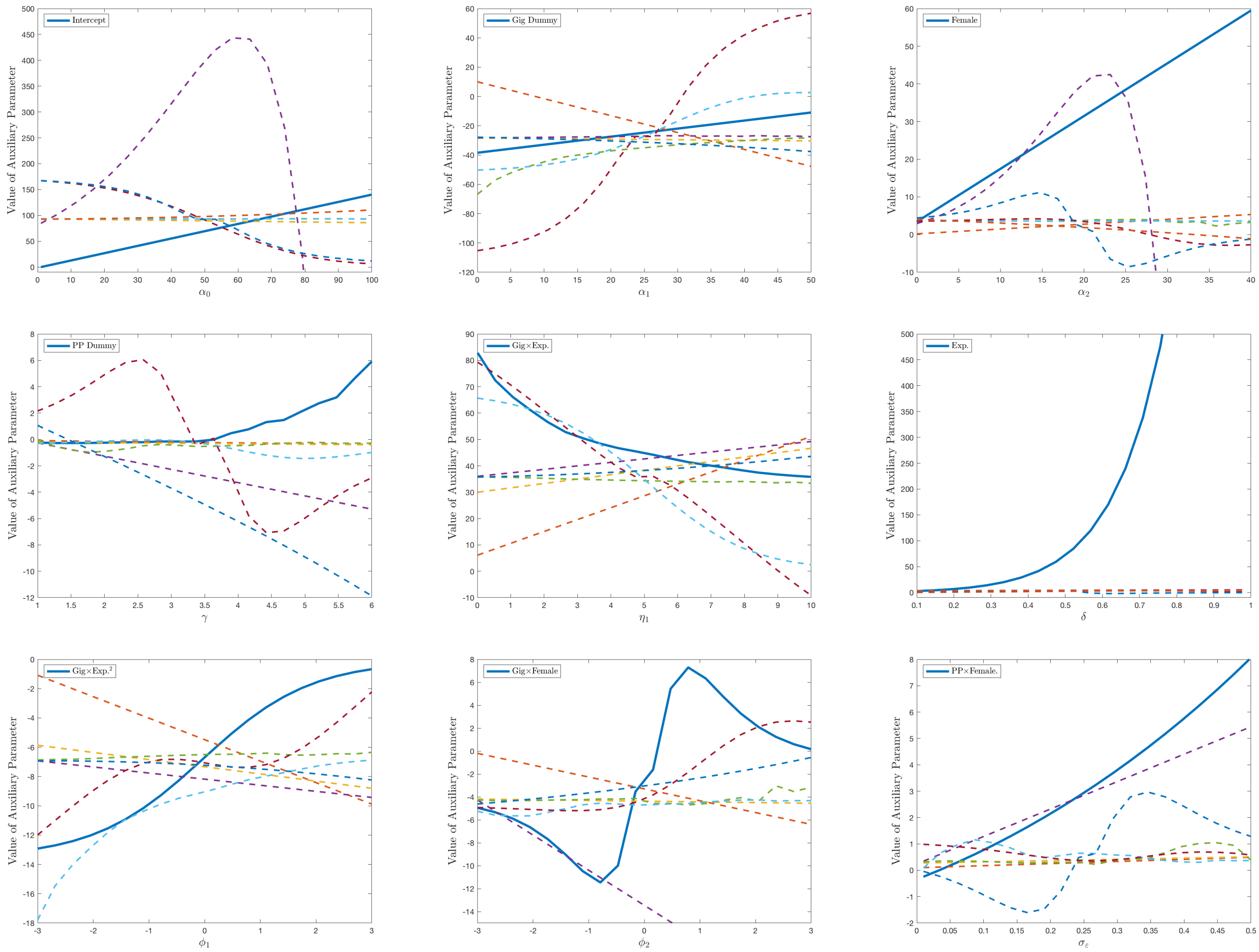


Figure A.3: Identification of Structural Parameters using Perturbation



Notes: The relationship between the parameter of the structural model and parameters of the auxiliary model obtained by simulations. In each plot, the x-axis presents values of the structural parameters, and the y-axis presents the values of the auxiliary parameters. The x-axis states the structural parameter that is being examined. The legend of each figure state which of the parameters of the auxiliary model is plotted. The solid line is computed using fixed values of the other estimated parameters (not the one identified). The dotted lines are computed using random values of the other structural parameters.

B Proofs

Concavity of the Objective Function

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

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

so that

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