

NBER WORKING PAPER SERIES

TOWARD AN UNDERSTANDING OF CORPORATE SOCIAL RESPONSIBILITY:  
THEORY AND FIELD EXPERIMENTAL EVIDENCE

Daniel Hedblom  
Brent R. Hickman  
John A. List

Working Paper 26222  
<http://www.nber.org/papers/w26222>

National Bureau of Economic Research  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
September 2019

We thank Stephane Bonhomme, Aaron Bodoh-Creed, Steven Levitt, Emir Kamenica, James McKinnon, Ivan Canay, Vadim Marmer, and Christopher Cotton for helpful comments. Special thanks to Joseph Seidel for excellent research assistance. The authors also wish to thank seminar participants at The U. of Chicago, U. of Wisconsin-Madison, UCL, CREST, Queen's U., the Economic Science Association North American meeting, Econometric Society North American Summer Meetings, the U. of Chicago Advances with Field Experiments conference, Penn State, Carnegie Mellon, and the Olin Business School at Washington U in St. Louis. We also acknowledge Danielle Hickman, Leah Vanevenhoven, and Bloom Bakeshop for insightful conversations on the interpretation of our results. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

**DISCLOSURE STATEMENT:** As per NBER's submission requirements, we wish to disclose the following information:

- 1) This research was executed entirely with internal funding provided by the University of Chicago and did not use outside sources of monetary support.
- 2) Neither Hedblom, nor Hickman, nor List received financial support from any interested parties in the execution of this research.
- 3) List held a position as Chief Economist at Uber Technologies during the data collection phase of this research. As explained in Section 2 of the current manuscript, Hedblom, Hickman, and List incorporated a firm, HHL Solutions, LLC, a non-profit consultancy, as part of the experimental research design. The two clients served by HHL Solutions, LLC included Uber Technologies, and the Becker-Friedman Institute for Research in Economics at the University of Chicago. Specifically, these two entities were the intended recipients of the data products resulting from test subjects' real-effort work producing data entry services. List did not receive any financial compensation specific to this research from Uber Technologies.
- 4) No outside parties had rights to review this manuscript prior to its circulation.
- 5) This research was conducted under Institutional Review Board Approval IRB15-1094 at the University of Chicago.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2019 by Daniel Hedblom, Brent R. Hickman, and John A. List. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Toward an Understanding of Corporate Social Responsibility: Theory and Field Experimental Evidence

Daniel Hedblom, Brent R. Hickman, and John A. List

NBER Working Paper No. 26222

September 2019

JEL No. C14,C93,J3,J33,J44,L21,M52

**ABSTRACT**

We develop theory and a tightly-linked field experiment to explore the supply side implications of corporate social responsibility (CSR). Our natural field experiment, in which we created our own firm and hired actual workers, generates a rich data set on worker behavior and responses to both pecuniary and CSR incentives. Making use of a novel identification framework, we use these data to estimate a structural principal-agent model. This approach permits us to compare and contrast treatment and selection effects of both CSR and financial incentives. Using data from more than 1100 job seekers, we find strong evidence that when a firm advertises work as socially-oriented, it attracts employees who are more productive, produce higher quality work, and have more highly valued leisure time. In terms of enhancing the labor pool, for example, CSR increases the number of applicants by 25 percent, an impact comparable to the effect of a 36 percent increase in wages. We also find an economically important complementarity between CSR and wage offers, highlighting the import of using both to hire and motivate workers. Beyond lending insights into the supply side of CSR, our research design serves as a framework for causal inference on other forms of non-pecuniary incentives and amenities in the workplace, or any other domain more generally.

Daniel Hedblom  
Department of Economics  
The University of Chicago  
1126 E. 59th St.  
Chicago, IL 60637  
hedblomdaniel@gmail.com

John A. List  
Department of Economics  
University of Chicago  
1126 East 59th  
Chicago, IL 60637  
and NBER  
jlist@uchicago.edu

Brent R. Hickman  
Olin Business School  
Washington University in Saint Louis  
One Brookings Drive, Campus Box #1133  
Saint Louis, MO 63130  
USA  
hickmanbr@gmail.com

A data appendix is available at <http://www.nber.org/data-appendix/w26222>

# 1 Introduction

Among social scientists, the virtues of corporate social responsibility (CSR) have been debated at least since [Friedman \(1970\)](#) famously described it as a “fundamentally subversive doctrine,” declaring that the only social responsibility of a firm is to maximize profits while avoiding deception or fraud. Nevertheless, today CSR is ubiquitous as most large firms have entire branches dedicated to ensuring socially responsible practices and/or planning and executing charitable activities. Indeed, hundreds of millions of dollars annually are budgeted on such programs. One economic explanation for the prevalence of CSR is that making the world a better place is not necessarily at odds with profit maximization. Proponents have argued that investments in CSR actually increase profits, but there is no consensus in the empirical literature on whether they have a positive effect on the firm’s bottom line. Some studies report a positive effect ([Waddock and Graves \(1997\)](#)) and others find mixed, negative, or no effects ([Barnett and Salomon \(2012\)](#); [Godfrey et al. \(2009\)](#); [Servaes and Tamayo \(2013\)](#)).

A possible reason for these mixed findings is that most empirical studies of CSR tend to focus primarily on the demand side of the market ([Du et al. \(2011\)](#); [Elfenbein et al. \(2012\)](#); [McDonnell and King \(2013\)](#); [Servaes and Tamayo \(2013\)](#)). The demand side explanation posits that firms use CSR as a marketing tool to signal ethical standards to consumers who value them. While plausible, such efforts ignore the effect of CSR on the supply side of the market, or its impact on employee behavior. For example, it is possible that CSR provides non-pecuniary incentives to (i) advantageously select workers to seek employment at the firm, and (ii) once employed CSR can induce higher productivity, work quality, and/or job satisfaction among employees. Some recent studies investigate supply-side effects of CSR, but these tend to focus on employee retention or wage requirements (see [Carnahan et al. \(2015\)](#); [Burbano \(2016\)](#)).

We take this literature in a new direction by using a natural field experiment that is tightly linked to a structural model to identify how important features of a firm’s production process interact with its CSR activities. The central premise is that workers potentially respond to variation in CSR, allowing us to estimate important behavioral parameters of the model. A firm’s existing workers are willing to supply labor and produce high-quality output, in part, due to their social preferences. We refer to this as the *treatment effect* of CSR. Yet, heterogeneous preferences for CSR

also cause workers to vary by their propensity to select into different jobs. A worker who prefers to work for a CSR firm will more readily accept a given wage offer from this firm than from a non-CSR firm offering the same wage. Thus, CSR may shape the composition of the firm’s worker pool, which we refer to as the *selection effect*. Our field experiment allows us to separate such treatment and selection effects.

We begin by presenting a structural model that provides direction into the exogenous variation necessary to quantify unobservable worker characteristics, such as productivity, worker quality, and time preferences. Our model and experimental design draw upon a novel identification framework proposed by [D’Haultfoeuille and Février \(2015\)](#) and [Torgovitsky \(2015\)](#) (henceforth, DFT) who show how discrete instruments (exogenous wage offers in our case) can be used to identify idiosyncratic time preferences. We build upon this framework in two important ways. First, we allow for transitory labor-supply shocks in settings where workers repeatedly make labor-leisure decisions across different periods of time. Second, while DFT require exclusion restrictions that imply no self-selection on wage offer, our experimental design allows us to relax this constraint by expansion of the observable set. Through use of a two-stage randomization during the hiring process, we are able to allow for selection on worker unobservables across wage contracts, while using observed variation in application rates to adjust for this selective entry, *ex post*. Allowing for selection on wage contracts allows us to compare pecuniary and non-pecuniary incentives in worker hiring. This methodological contribution highlights how field experimental variation can both relax identification assumptions and broaden empirical insights.

Our field experiment consists of randomized treatments in two stages: (1) a hiring stage and (2) a real-effort work stage. Both stages generate data on worker behavior when faced with different wage contracts and non-pecuniary incentives in the form of information about the firm’s investments in CSR. The field experiment is tightly linked to the theoretical model to ensure that we are able to generate the appropriate instruments and variation in observables to permit parameters of interest to be accurately estimated with minimal *a priori* assumptions on the data-generating process.

In the hiring stage of the experiment, we recruit subjects via online advertisements placed on marketplaces for jobs in 12 cities throughout the US. We randomly vary the offered wage rate and how much information is revealed about our firm’s involvement in CSR. Our subjects are recruited from an actual marketplace at wage

rates common to adult semi-skilled labor (\$11/hr to \$15/hr) to avoid sample selection problems when subjects are recruited from a very specific population (Levitt and List 2007). Application rates in each treatment group allow us to investigate and compare how pecuniary incentives (different wage rates) and non-pecuniary incentives (information about CSR) affect the probability that a given individual signals acceptance of the firm’s offer by applying for the job. Beyond that, the hiring stage provides an instrument for structural identification of labor-supply costs and selection effects.

In the second stage of the experiment, subjects who are hired engage in real data-entry work through a custom-designed, web-based task system. In order for employees to work they must log into our website, which then continuously records data in the background about labor supply, productivity, and output quality. The website also provides a means of varying task framing to pin down treatment effects. Importantly, both those who are recruited for a CSR job and those who are recruited neutrally receive the *same* within-worker variation in task framing prompts. This aspect of our research design allows us to separate the influence of treatment and selection effects on outcomes of interest.

Our research design yields several interesting insights. First, we find that both wages and CSR have important effects on application rates. An increase in the wage offer from \$11/hour to \$15/hour increases application rates by about 32%, while advertising the firm as a CSR firm increases application rates by 24%, or three quarters as much. What remains unclear is whether we are merely observing increased response rates by the same set of workers, or whether the reduced-form results signal composition effects that alter the mixture of worker traits within the firm’s labor pool.

We therefore use our field experimental data to estimate a structural model of unobserved worker heterogeneity. We study the relative impacts of worker selection on both wage offer and non-pecuniary job perks. This is an important feature of our methodology, as our structural methods use a theoretical model of worker behavior to quantify unobservable, idiosyncratic characteristics, while controlling for transitory phenomena such as learning effects and labor-supply shocks. We also disentangle the impacts of work-stage treatments from hiring-stage selection on idiosyncratic labor supply, productivity, and work quality. In this manner, our field experiment is designed to establish an ideal data-generating process for estimating the primitives of a rich model of worker production and labor supply. As a consequence, our structural

econometric model yields results that cannot be derived directly from raw data, and is particularly useful for conducting counterfactual analysis and for generalizing the results beyond the experimental setting.

Several insights emerge from the structural estimates. First, we construct stochastic dominance tests on the distributions of estimated worker characteristics across different subgroups in our sample. This sheds light on worker selection in terms of wage offers and CSR advertisement at the recruitment stage. We find evidence of significant composition shifts on three dimensions of worker heterogeneity: productivity, work quality, and labor-supply costs. The analysis highlights that using either high wage offers or CSR (or both) as a recruitment tool induces non-trivial advantageous selection on workers' baseline productivity and accuracy. Consistent with labor search theory, higher wage offers induce workers with a higher shadow value of time to apply. Interestingly, a CSR hiring strategy operates similarly: CSR-recruited workers tend to have higher utility costs of supplying time to the firm.

Second, we find that a CSR advertisement campaign during the work stage induces a separate advantageous treatment effect on existing workers (i.e., holding their unobserved characteristics fixed). It induces workers to increase their output per unit of time, and also to voluntarily reduce the amount of non-productive down-time while they are on shift (i.e., getting paid by the firm for their time). Since treated workers produce output faster and take fewer breaks, one might worry that these gains arise at the expense of work quality, but we find no evidence to that effect. These results imply that an internal CSR advertisement campaign can have unambiguously advantageous impacts on the firm's bottom line through altering existing workers' behavior.

Finally, through a counterfactual simulation based on our structural estimates we find that selection effects can meaningfully improve the firm's bottom line. For various levels of quality control on the firm's output, per-unit production costs among \$15-wage recruits is nearly the same as per-unit production costs among \$11-wage recruits. This is because the former group produce more output per hour at a higher quality level, requiring fewer costly redundancy measures. High-wage recruits also supply more output to the firm each day through increased labor supply and productivity. This reduces the number of workers needed to maintain a given output level, which may further improve the firm's bottom line if there is a non-trivial human resources cost tied to each worker. These results are of independent interest, though they relate

to the CSR question by demonstrating that a profit-motivated firm may wish to adopt sustainable hiring/compensation practices that directly benefit its workers.

In addition, CSR selection implies not only that workers produce more output across various levels of quality control, but their per-unit production costs are roughly 25% lower than their neutral-recruited counterparts as well. We estimate that the scale of a firm needed to profitably justify a \$1,000,000 annual CSR budget at 411 workers or more. Holding this CSR budget fixed, a firm which continuously employs more than 411 data-entry workers would be able to strictly profit from a \$1,000,000 annual CSR investment, through gains on selected worker productivity and work quality alone. This result could explain why larger firms tend to invest more in CSR activities: with a larger employee base over which to spread costs and reap productivity benefits, the per-worker profitability of a dollar invested in CSR may be higher as well.

The paper proceeds as follows. Section 2 reviews relevant literature and describes the field experimental design. Section 3 outlines the theoretical model while Section 4 provides a brief overview of identification and estimation, with technical details relegated to Appendix A. Section 5.1 presents descriptive statistics and preliminary results on worker application rates from the hiring stage. Sections 5.2–5.4 present results from the structural model, and Section 6 presents a counterfactual simulation exercise to explore implications for the firm’s cost structure. Section 7 concludes.

## 2 Literature Review and Experimental Design

To the best of our knowledge, the only previous studies using field experiments to investigate the impact of CSR on labor market fundamentals are due to Burbano (2015, 2016). Using experiments conducted on mTurk and Elance, Burbano (2016) estimates reduced-form effects of CSR messages on labor output and salary requirements. Our study links a structural model with a field experiment to identify and estimate the interplay between unobservable worker characteristics and CSR activities on the part of the firm, especially through worker selection at the recruiting stage.<sup>1</sup>

---

<sup>1</sup>Burbano (2015) uses a field experiment to explore the effect of CSR on virtual workers, and finds that CSR increases virtual worker’s willingness to do extra, unrequired work. In related work, List and Momeni (2017) explore the reduced-form effects of CSR on employing cheating. They find that CSR induces misbehaviors at work for some employees.

Our paper also contributes to the broader literature on the effects of CSR on firm performance ([Greening and Turban \(2000\)](#)). A recent paper by [Bertrand et al. \(2018\)](#) argues that large firms are able to use corporate philanthropy as a tax-exempt form of political lobbying by donating to charities connected to elected officials. To complement this work, our study also explores how CSR affects the bottom line, but through its impact on the firm’s labor force and cost structure.

We also contribute to the recent strand of literature within economics using experiments to generate data to identify structural models. Previous examples include charitable giving ([DellaVigna et al. \(2012\)](#)), voting behavior ([DellaVigna et al. \(2017\)](#)), gift exchange ([DellaVigna et al. \(2019\)](#)), disappointment aversion ([Gill and Prowse \(2012\)](#)), and childhood education ([Cotton et al. \(2019\)](#)). We contribute to this literature by providing an in-depth labor market application with sorting and treatment effects in the workplace. Finally, we contribute to the literature on equalizing differences ([Roback \(1982\)](#); [Rosen \(1986\)](#)). Our structural analysis of selection effects constitutes a new way of estimating compensating differentials or equalizing differences across jobs that vary non-pecuniary incentives. Thus, the methodology in this paper has applicability in areas beyond CSR. One could readily use a similar method to estimate the effect of virtually any workplace characteristic—*e.g.*, flexible hours or workplace competition—over which there are heterogeneous worker preferences.

## 2.1 Experimental Design

We design our natural field experiment (see [Harrison and List \(2004\)](#)) in a manner that combines market realism and expands on ideas from DFT to identify the distribution of unobserved worker characteristics while allowing for selection effects from variation in both wages and CSR. The experiment uses randomized treatments administered in two stages: (1) a hiring stage and (2) a real-effort work stage. In the first stage, we recruited and hired subjects while varying wage and non-pecuniary treatment conditions. The second stage had subjects performing data-entry tasks through a web-based worker portal that produces detailed measures of outputs and labor inputs during each day of the sample period. The experiment was designed with several important conditions in mind. First, exogenous variation is necessary in both stages for identification, as will become clear in the next section. Second, regardless of how workers were recruited, *all* are given the same treatment-control variation during the work stage. This aspect of our design is crucial for separating selection effects



and treatment effects. Third, we need to achieve a work environment as natural as possible, as the subjects investigate the firm and earn money in a job they chose to apply for on their own. Fourth, the work task must be uniform across treatments to allow identification of labor-supply costs, which include an idiosyncratic component and a common cost function.

In order to satisfy these conditions, we started an actual consultancy firm, *HHL Solutions, LLC*. The firm specializes in data collection and data management services for various clients, including for-profit firms (e.g., Uber Technologies, Inc.), and non-profit firms (e.g., The University of Chicago). A website for the firm was also designed and published online (see Online Appendix B for a screenshot). The website did not play a direct part in the experiment, but was set up in case potential subjects chose to search for additional information about the firm. Running our own firm rather than partnering with an existing firm or using a crowdsourcing market (e.g., mTurk) gave us the flexibility and control to design the user experience and workload to balance the needs of producing scientifically viable data and serving client needs.<sup>2</sup> Our empirical analysis requires detailed observations with a work task that is simple enough to enable a clean data-generating process, but realistic enough to be meaningful as an economic case study. Creating our own firm also provided control over the size of the workforce, which would not have been possible in a typical industry partnership.

## 2.2 Hiring Stage

Employment advertisements were posted on [www.Craigslist.com](http://www.Craigslist.com) in 12 major cities throughout the US. The particular cities were chosen based on size, geography, and the relative activity of the associated Craigslist page. The initial ad was a terse announcement of a short-term employment opportunity for data-entry workers. It mentioned that wages were to be set somewhere in the range of \$11-\$15. Interested individuals were instructed to request more information by replying to the ad via email. The set of all individuals who responded to the initial ad comprise our pool of *potential applicants*, and they were subsequently randomized into treatment groups.

---

<sup>2</sup>Two further reasons dictated that we start our own firm rather than use internet crowdsourcing. First, subjects recruited via mTurk may tend to pay less attention to experimental materials and may have a different attitude towards pecuniary and non-pecuniary trade-offs than the general worker pool in a developed country, which is our population of interest (see [Chandler et al. \(2014\)](#); [Goodman et al. \(2013\)](#); [Paolacci and Chandler \(2014\)](#)). Second, work-stage variation and output monitoring require a large degree of flexibility in the design of our online task system.

The first stage treatments were administered via email. Importantly, all subjects, regardless of treatment, were exposed to the same original job advertisement on Craigslist. This permits us to randomize subjects to treatments in a controlled manner. In contrast, if we were to use more than one version of the initial ad, there would be no way of controlling which ad each subject was exposed to.<sup>3</sup> The timing of the first stage of the experiment is summarized in figure 1. We used two different treatments in the first stage: for subject  $i$  there were (1) two possible wage levels,  $W_i \in \{\$11, \$15\}$ , and (2) two slightly different texts containing information about the firm. The CSR information treatment for subject  $i = 1, 2, \dots, I$  is represented by a dummy variable  $X_{0i}$  equalling 1 if the subject was informed of CSR in the job description letter, and 0 otherwise. The wage and CSR treatments were crossed, resulting in 4 treatment cells summarized in Table 1.

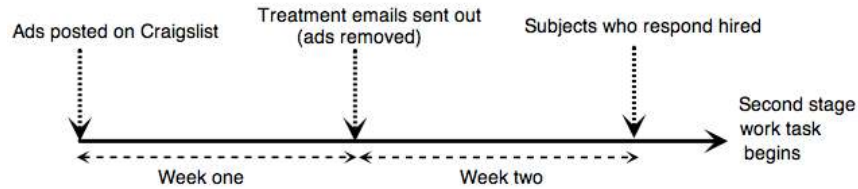


Figure 1: Hiring Stage Timing

	$X_{0i} = 0$	$X_{0i} = 1$
$W_i = \$11$	11 / Neutral	11 / CSR
$W_i = \$15$	15 / Neutral	15 / CSR

Table 1: Hiring Stage Treatment Cells

We chose the wage levels of \$11 and \$15 in part because we wanted to be able to attract a reasonably representative set of data-entry workers. According to market research firm Payscale.com, at the time we ran the experiment \$11/hr and \$15/hr corresponded to the 25<sup>th</sup> and 75<sup>th</sup> percentile wages for data entry operators in the United

<sup>3</sup>This element of the design is similar to Flory et al. (2014) and Leibbrandt and List (2014) to study gender differences in attitude towards competition and wage negotiation. A similar "two-step" randomization procedure was used by Ashraf et al. (2010) to separate screening and sunk-cost effects in a field experiment on healthcare product pricing.

States.<sup>4</sup> The pecuniary and non-pecuniary incentive variation in our hiring stage was contained in the randomized email we sent in response to initial queries. This email encapsulated the difference between our "Neutral" and "CSR" recruitment-stage treatments in the following way. In the two treatment cells with  $X_{0i} = 0$ , the letter contained information on the company, the task, the (randomized) offered wage, and directions for how to formally apply for the job. In the two treatment cells with  $X_{0i} = 1$  it contained all the same text with an addition to one paragraph that explained *HHL Inc.*'s CSR activity of charging low prices to non-profit clients engaged in good causes (which, to avoid deception, was in fact true, as discussed below).

Figure 13 in the online supplement contains a side-by-side comparison of the hiring-stage recruitment letters for a worker who was randomly assigned a high wage offer. The only difference between a neutral recruitment letter and a CSR recruitment letter is the addition of the following four sentences in the third paragraph: "*Some of [our clients] work in the nonprofit sector with various charitable causes. For example, with projects aimed at improving access to education for underprivileged children. We believe that these organizations are making the world a better place and we want to help them doing so. Due to the charitable nature of their activity, we only charge these clients at cost.*" The letters also contain our wage variation: the fourth block of text below the salutation reads "The hourly wage rate is \$ $j$ ." with  $j \in \{11, 15\}$

Note that none of our communication with test subjects involved deception, as our consultancy had at least one of each actual type of for-profit/non-profit clients. For example, data entry services were performed on behalf of Uber Technologies, Inc. and the Becker-Friedman Institute at the University of Chicago (for research independent of this paper) which conducts ongoing studies on childhood human capital investment, with emphasis on alleviating inequality of educational opportunity.

After viewing the information letter, subjects' behavior provides us with our first outcome variable: the application decision. By computing the share of the initial pool of subjects (the ones that responded to the original Craigslist ad) who apply, we can estimate how the wage rate (pecuniary incentive) and the information about CSR (non-pecuniary incentive) affect the probability of applying. The application decision is an important outcome variable as it constitutes the extensive margin of our treatments. In the hiring stage, to ensure that we generated a representative sample of employees across the treatment cells while staying within our budget, we

---

<sup>4</sup> [http://www.payscale.comresearchUSJob=Data\\_Entry\\_OperatorHourly\\_Rate](http://www.payscale.comresearchUSJob=Data_Entry_OperatorHourly_Rate) accessed 04/14/2017

hired a random subset of  $I$  applicants. In terms of the model, we obtain a vector  $(\mathbf{X}_0, \mathbf{W}) = (X_{01}, \dots, X_{0I}, W_1, \dots, W_I)$  with the CSR treatment dummy and wage offer for each agent.<sup>5</sup> Our model of unobserved worker characteristics includes four dimensions of heterogeneity: two separate aspects of worker productivity, work quality, and preferences for leisure which dictate labor-supply decisions. Our hiring-stage variation is crucial for pinning down how the firm’s CSR activities create selection effects that shape the distribution of these characteristics within its applicant pool.

### 2.3 Work Stage

Subjects were informed that they would be working for 10 days, and that they were free to work any number of hours during each day of the project period. Concerning our model, one day corresponds to a period  $t$  so that  $t = 1, 2, \dots, T = 10$  in the experiment. Each subject worked from home through our custom-designed online task system. They were each provided with a username and a password for logging into the system at any time. This allowed us to track the number of hours worked each day, the quantity of output produced, and the quality of output.

Upon logging in, subjects saw a dialog box explaining the task system itself and some information about the client associated with the work for the particular day. After reading this information, subjects were taken to the main screen. Here, subjects were faced with a list of tasks that could be completed in any order. A task was initiated by clicking on the associated link, taking the subject to a data entry screen. On the data entry screen, the subject is presented with a snapshot from *Google Streetview* depicting a street in a major city in the US. The snapshot was sampled from a large database of over 3,200 images collected prior to the experiment. Below the snapshot was a web form with a number of text boxes and drop-down menus for entering data. The subject’s task was to visually process the image and correctly complete the form by entering information for each variable listed.

There were 6 meta-data variables about the picture itself and 12 variables related to the street visible in the picture. Some of the variables require the subject to make an assessment of the quality of the street or the buildings visible in the picture

---

<sup>5</sup>In our random hiring procedure, we over-sampled high-wage-offer recruits at 97 vs 73. I.e., we randomly sampled 33% more high-wage recruits than low-wage recruits in anticipation of our selection-correction procedure which requires us to exclude a subset of high-wage workers for part of the estimator (see discussion in Section 4.2 and Appendices A.3 and A.7). Our sampling of CSR recruits and neutral recruits was roughly even at 51.8% and 48.2%, respectively.

and assign a rating on a Likert scale. Training materials were distributed to each worker explaining how to map visual queues into Likert-scale values. Other variables required the subject to count the number of occurrences of certain objects in the picture (e.g., potholes and broken windows). Once all the information is recorded, the subject submits the task by clicking on a button, taking him or her back to the main screen. The completed tasks were then marked as “completed” in the middle of the main screen. Time was automatically tracked while the subject was logged into the system. The subject could view a running total of paid working time (within the current day) on the main screen at any time. The task was chosen so both our private and non-profit clients could gain valuable information relevant to their operations, while avoiding deception of our test-subjects. Online Appendix B contains screenshots of the main web-page and the task pages where they actually coded information about the *Google Streetview* images.

### 2.3.1 Work Stage Treatments

Work-stage CSR treatment is denoted by  $X_{1it} = 1$  if CSR treatment is present on day  $t$  and  $X_{1it} = 0$  otherwise. For all workers, regardless of recruitment status, within-person variation in treatments consisted of information about the client associated with the task of a given day, and *HHL*’s dealings with that client. This information was presented to subjects when logging into the online work system and in information boxes on the main screen after tasks were completed. On days when the CSR treatment was present, there was a short, color-coded prompt on the screen indicating that today’s project was for a non-profit client organization that works with improving access to education for underprivileged children. Moreover, the prompt indicated that our firm wants to help these clients to make the world a better place, so we charge them for our services only at a rate which will cover our production cost. On neutral days a similar color-coded prompt appeared in the same location on their screen, indicating that the day’s project was for a for-profit client, and thanking them for their efforts (see Figures 14–17 in Online Appendix B). In order to avoid confusion, the screen also indicated that the employee’s own wage rate is the same, regardless.

Each test subject was given the opportunity to work for a set of 10 contiguous days.<sup>6</sup> The variable  $X_{1it}$  served as our instrument for CSR during the work stage.

---

<sup>6</sup>All employment spells occurred during September and October of 2016 in 6 waves, with each

There were two treatment schedules: one where CSR prompts occurred on even-number days only (i.e.,  $X_{1it} = 1$  for  $t = 2, 4, 6, 8, 10$ ) and one where CSR prompts occurred on odd-number days only (i.e.,  $X_{1it} = 1$  for  $t = 1, 3, 5, 7, 9$ ). Once employees were hired, half of all subjects selected from each of the four bins in Table 1 were assigned to an odd-day treatment schedule and the other half were assigned to an even-day schedule.

## 2.4 Observables

The model (described below) is partitioned into two parts: labor supply and output quantity/quality, conditional on being at work. Subjects are allowed to decide for themselves how many hours to work each day. Our time tracking system automatically logs a worker out (and halts paid time) if he or she does not touch the mouse or keyboard for longer than two minutes. This prevents subjects from gaming the system by logging in and earning money without at least paying nominal attention to the computer screen. Still, this setup purposefully provided latitude for heterogeneity across individuals, as well as within-person behavioral differences in terms of concentration level (accuracy) and “feet dragging” (productivity).

Daily labor supply,  $H_{it}$ , is recorded as the total number of hours subject  $i$  spent logged into the system during day  $t$ . In order to be paid, the worker must be logged into the website using her assigned username and password, which allows us to link all data collected during her web session back to her. After workers log in, our interactive web portal records and timestamps each page view, and compiles user responses into a database. Each *Google Streetview* image also had a unique identifier, so that we can count the total number of unique images  $Q_i$  that worker  $i$  processed, and compare responses by different workers on the same image. We are also able to use timestamps on page views to assign a chronological index  $q_i$  ordering each unit of worker  $i$ 's output from 1 to  $Q_i$ . Each worker processed a given image exactly once. For each completed unit of worker  $i$ 's output, we have two measures of time expenditure:  $\tau_{q_i}$  is the total amount of time that worker  $i$  spent on the data entry page, and  $D_{q_i}$  is the amount of non-productive down-time (i.e., on the home page of the website) after completion of the  $(q_i - 1)^{th}$  unit and before completion of the  $q_i^{th}$  unit of output.

Since workers are paid by the firm for both types of time (as is the case for most wave of test subjects having  $t = 1, \dots, 10$  occurring on the same days. Roughly half of all waves began on a Tuesday, and the rest began on a Thursday.

other naturally-occurring work settings), taking these two measures separately allows us to quantify multiple dimensions of worker productivity: *active productivity*, or time inputs required to produce output while a worker is on-task, and *passive productivity*, or the amount of a worker’s time spent between productive tasks. A possible analog of passive productivity in traditional work settings would be the amount of paid time an office worker spends surfing the internet or socializing with co-workers on non-work-related topics, while active productivity would be the amount of time it takes the worker to complete a report, conditional on being actively focused on it.

Measuring work quality, or data entry accuracy rate, denoted  $A_{qi}$ , is a little more involved. For each image there were 19 variables for workers to code on the image processing page. Some of these responses were on a Likert scale, exhibiting an element of subjective judgment, so we distributed to our workers a training handout containing standardized explanations and graphic examples of what constitutes an appropriate score (e.g., a “Street Quality” score of 3 as opposed to 4 or 5). For variables that involved trans-coding of text, correctness was judged on the basis of spelling but not capitalization. In order to determine correctness for each response we used an industry-standard consensus criterion. Specifically, each unique image was coded by multiple workers, and for each variable we specify the correct answer as the modal response across all workers for the same variable-image pair. A given worker’s response to each variable is assigned a value of 1 if it matches the modal response for that same image-variable pair, and 0 otherwise.

This choice reflects common industry practice for data-entry workers in contexts marked by an element of subjectivity. For example, Pearson, the College Board, and ACT, Inc. are leaders of a large industry on standardized testing. Currently, Pearson is contracted by the U.S. State Department to administer and score the foreign service exam, which includes an essay portion. After training workers on the criteria for awarding a given integer score from a set scale to an essay response, Pearson assigns 3 independent workers to score the same essay in double-blind fashion. Whenever 2 or more workers agree on a common score (i.e., the mode) it is declared as the test-taker’s official score. Pearson and ACT track their workers’ ability to select the modal response in order to gauge and monitor the quality of their work.

Since *HHL Solutions* played the double role of (i) producing usable data entry output for actual clientele and (ii) producing worker behavior data for the current study, we designed the production process with a great deal more redundancy than if

( $i$ ) were the only goal. A minimum of 7 different workers coded data for each unique image, with the median and mean being 17 and 21.5 workers per image, respectively. This high degree of redundancy ensured we could compute reliable measures of the modal responses for each image-variable pair, which was essential for deriving a valid measure of work quality.

Table 11 in Online Appendix B displays descriptive statistics on responses and accuracy rates for each individual variable across a total of 62,138 worker-image observations. A majority of the variables in the web form—*road work visible*, *graffiti visible*, *trees/shrubs visible*, *for-sale signs visible*, *broken street signs visible*, *people covering faces*, *shoes hanging from wires*, *street number*, *month*, *year*, *city*, *state*—had very high mean accuracy rates across workers and images, ranging between 92% and 99%. The final seven variables—*building quality*, *quality of visible cars*, *litter*, *picture quality*, *street quality*, *number of visible potholes*, *street name*—posed more of a challenge for workers, with average correctness across all worker-image pairs ranging from 85% to 47%. Broadly, the difference between the two sets of variables appears tied to the nature and difficulty of the task. For example, 5 of the 7 were Likert-scale variables where accuracy required careful thought and concentration to digest our training materials and translate them into judgment calls. Judging what constitutes a pothole in the pavement also required a higher level of attention. Workers also appeared more adept at trans-coding street numbers than text strings for street names without mistakes.

For each worker  $i$  and each unit of output  $q_i$ , the accuracy rate  $A_{q_i} \in \{\frac{0}{7}, \frac{1}{7}, \dots, \frac{7}{7}\}$  was coded as the mean across the variable-specific correctness outcomes for the final seven variables. Given that our primary research question concerns worker heterogeneity, we define our accuracy index to focus on more difficult tasks in order to illuminate aspects of job performance that most set workers apart in a vertical sense.  $A_{q_i}$  had a mean and standard deviation of 0.586 and 0.226, respectively.<sup>7</sup>

### 3 The Model

Our model includes four dimensions of worker heterogeneity: active productivity, passive productivity, work quality, and leisure preferences (i.e., costs of supplying

---

<sup>7</sup>As a robustness check, we re-executed our analysis on an alternative form of the accuracy index which used all variables except four which had little or no variance. The results were similar.



time to the firm). Productivity indexes the average amount of time it takes a worker to complete one unit of output. Here we discuss how the sources of variation in our observables lead to identification of the structural primitives. Our identification strategy combines a simple principal-agent framework with panel-data methods and served as a guide for our experimental design.

### 3.1 The Baseline Model

Workers are recruited to produce outputs that the firm sells on an open market each period  $t = 1, 2, \dots$ . The firm produces a single good, but not all units of output, indexed by  $q_i$  for worker  $i$ , are created equal. In particular, the firm prescribes certain specifications that outputs should ideally meet, and accurately producing the good requires cognitive exertion and to some extent is at the worker’s discretion. The firm offers a fixed hourly wage contract,  $w$ , for the worker to supply her available time to the firm.<sup>8</sup> Holding work quality fixed, maintaining peak productivity requires worker exertion, and is also partly under her control. The firm wishes to maximize the quantity of high-quality outputs produced per unit of paid time. We assume that perfect monitoring is prohibitively costly so that the firm is faced with an agency problem.

Each potential worker is characterized by a privately-known 4-dimensional type,  $(\Theta_{pi}, \Theta_{di}, \Theta_{ai}, \Theta_{li})$ , which indexes their baseline productivity,  $(\Theta_{pi}, \Theta_{di})$ , “work quality” (or accuracy rate in our data entry context),  $\Theta_{ai}$ , and shadow value of leisure time,  $\Theta_{li}$ . The firm is characterized by a vector of fixed, commonly-known characteristics,  $\mathbf{Z}$ —*e.g.*, quality of work environment, disutility intrinsically tied to its production process, *etc.*—but it may choose to incur a fixed cost to engage in CSR activities, which are potentially valued by workers. Let  $X_0 \in \{0, 1\}$  denote the firm’s decision of whether to include CSR activities in its regular operations, let  $X_{1t} = 1$  if the firm actively participates in CSR in period  $t$ ,  $X_{1t} = 0$  otherwise.<sup>9</sup> Let  $(\mathcal{T}_p, \mathcal{T}_d, \mathcal{T}_a, \mathcal{T}_l)$  denote

---

<sup>8</sup>Our model assumes that workers have no set schedule, and are allowed to choose their own hours. This permits our research design to create an instrument to identify heterogeneity in labor supply (*i.e.*, leisure preferences). However, this aspect of the model is not uncommon in actual labor markets: short-run hiring frictions often drive employers to request a temporary increase in labor supply of their fixed-schedule workers in the form of overtime incentives.

<sup>9</sup>One can alternatively think of the time-varying  $X_{1t}$  as the firm engaging in an internal advertising campaign in period  $t$  to remind workers of its CSR-oriented contributions to the greater good.

the treatment effects on productivity, accuracy, and labor supply, respectively, of laboring under condition  $X_{1t} = 1$ .<sup>10</sup> If the worker values the feeling of contributing to the firm’s efforts to make the world a better place, it may partially offset the costs of cognitive exertion and/or labor supply.

Upon receiving a job offer  $(w, X_0, \mathbf{Z})$  from the firm, each worker compares it with her outside options and then decides whether to apply for a job and select herself into the firm’s labor pool. Let  $G_{pdal}(\Theta_{pi}, \Theta_{di}, \Theta_{ai}, \Theta_{li})$  denote the joint CDF of worker characteristics present in the firm’s selected applicant pool, with marginal distributions  $G_p$ ,  $G_d$ ,  $G_a$ , and  $G_l$ , respectively. Our experimental framework is designed to control for the fact that the distribution of worker characteristics  $G_{pdal}$  depends on workers’ self-selection choices as a function of  $w$  and  $X_0$ .

### 3.2 Active Productivity

Worker  $i$ ’s production technology governs the amount of time required for each unit of output. We assume it follows a standard *experience curve* form with permanent worker heterogeneity and transitory noise:

$$\tau(q_i; \Theta_{pi}, X_{1q_i}) = \Theta_{pi} \times \mathcal{T}_p^{X_{1q_i}} \times \tau_1 q_i^{-\delta} \times u_{q_i}^p, \quad q_i = 1, \dots, Q_i. \quad (1)$$

Here,  $\tau_1$  is baseline mean time required for the first unit of output,  $\delta$  represents the learning effect or elasticity of production time with regard to cumulative output, and  $u_{q_i}^p$  is an exogenous, positive, unit-specific shock. In a slight shift of notation,  $X_{1q_i}$  is the CSR treatment in place when  $i$  produced her  $q_i^{th}$  unit of output.

Here,  $\Theta_{pi}$  is a permanent, idiosyncratic scaling of production time, with smaller values representing higher productivity workers. As  $\Theta_{pi}$  or  $\mathcal{T}_p$  decrease, the worker becomes more productive. A CSR firm may attract applicants whose unobserved characteristics are fundamentally different, *before* they commence employment. In what follows we refer to the possibility that  $G_p(\Theta_{pi}|X_0 = 1) \neq G_p(\Theta_{pi}|X_0 = 0)$  as the *selection effect*. The working hypothesis in our empirical application is that profit-maximizing firms view costly CSR activity as an investment in their bottom line, and expect the productivity distribution  $G_p(\Theta_{pi}|X_0 = 0)$  to dominate  $G_p(\Theta_{pi}|X_0 = 1)$ , leading to advantageous selection.

<sup>10</sup>Our research design could be applied to general non-pecuniary investments in workplace quality, since the theory does not specifically require A CSR interpretation. If investment in a non-pecuniary perk is one time and permanent thereafter, then  $X_0 \in \{0, 1\}$ , and  $X_{1t} = X_0 \forall t$ .

### 3.3 Passive Productivity

Our formal model for a worker’s passive productivity is very similar to the model for active productivity, except that we assume workers have no need of learning how to waste time. Specifically, we model downtime on the  $q_i^{th}$  unit of output as

$$D(q_i; \Theta_{di}, X_{1q_i}) = \Theta_{di} \times \mathcal{T}_d^{X_{1q_i}} \times D_0 \times u_{q_i}^d. \quad (2)$$

Here,  $D_0$  is the baseline average down-time and  $u_{q_i}^d$  is an exogenous, transitory shock. As before,  $\Theta_{di}$  represents a permanent, idiosyncratic scaling of one’s propensity to log unproductive down-time. As  $\Theta_{di}$  or  $\mathcal{T}_d$  decrease, the worker becomes more productive by wasting less paid time, holding other factors fixed. Once again, we refer to the possibility that  $G_d(\Theta_{di}|X_0 = 1) \neq G_d(\Theta_{di}|X_0 = 0)$  as the *selection effect*.

### 3.4 Work Quality/Accuracy

Work quality is also measured at the worker-unit level. With an eye toward our data-entry empirical application, we use the terms “work quality” and “accuracy” interchangeably. We define a continuous, latent accuracy index  $A_{q_i}^* \in \mathbb{R}$  as

$$A_{q_i}^* = \Theta_{ai} + X_{1q_i} \mathcal{T}_a + \varepsilon_{q_i}^a, \quad (3)$$

where  $\varepsilon_{q_i}^a$  is an exogenous, iid, unit-specific shock that follows a symmetric distribution  $\Phi(\cdot)$ . The latent accuracy index encapsulates permanent worker characteristics, treatment effects, and transitory output quality shocks. Conceptually, one’s work quality on a particular task becomes perfect as  $A_{q_i}^* \rightarrow \infty$  and perfectly flawed as  $A_{q_i}^* \rightarrow -\infty$ . However, a limited dependent variable problem is present, and we cannot observe the latent continuous accuracy index  $A_{q_i}^*$  directly. Instead, we observe a binary measure<sup>11</sup>

$$A_{q_i} = \begin{cases} 1 & \text{if } A_{q_i}^* > 0 \\ 0 & \text{otherwise.} \end{cases}$$

If  $\Phi$  is the standard normal CDF, we get the familiar probit formulation for accuracy,

$$\begin{aligned} \Pr[A_{q_i} = 1 | \Theta_{ai}] &= \Pr[A_{q_i}^* > 0 | \Theta_{ai}] = \Pr[\varepsilon_{q_i}^a > -\Theta_{ai} - X_{1q_i} \mathcal{T}_a] \\ &= \Phi(\Theta_{ai} + X_{1q_i} \mathcal{T}_a). \end{aligned} \quad (4)$$

---

<sup>11</sup>Our data actually contain a finer accuracy measure (see Section 2.4). Yet, the estimator is based on averaging accuracy across multiple units of output to get treatment effects and worker fixed effects. Since these averages belong to an asymptotically dense set, the method applies whenever  $A_{q_i}$  lives on a finite grid and shocks are uncorrelated across different  $q_i$ ’s. See Appendix A for further details.

Once again, the working hypothesis in our empirical application is that profit-maximizing firms expect a return from their CSR activities in the form of an advantageous shift in the type distribution. We wish to test whether a CSR-induced selection effect causes the accuracy distribution  $G_a(\Theta_{ai}|X_0 = 1)$  to dominate  $G_a(\Theta_{ai}|X_0 = 0)$  within the pool of workers who are willing to accept the firm's employment.

### 3.5 Labor Supply

Having accepted the firm's wage rate offer  $w$ , the worker now decides in each period  $t = 1, 2, \dots$  how many hours of her leisure time to supply to the firm. This decision happens in two stages. Conditional on showing up to work on a given day  $t$ , the worker incurs a utility cost (delineated in dollar units) to supply  $H_{it}$  hours of would-be leisure time to the firm, given by

$$C_{it}(H_{it}; \Theta_{li}, X_{1it}) = \Theta_{li} \times \mathcal{T}_l^{X_{1it}} \times c(H_{it}) \times u_{it}^l, \quad (5)$$

where  $u_{it}^l$  is an exogenous, iid, log-normal shock with  $E[\log(u_{it}^l)] = 0$  and  $Var[\log(u_{it}^l)] = \sigma_l^2$ , and  $c(\cdot)$  is a baseline cost function common to all workers in all periods. Worker type  $\Theta_{li}$  and treatment  $\mathcal{T}_l$  govern the shadow value of time by scaling monetized utility costs. Lower values of either imply a worker's time is less valuable to her, or in other words, it signals a greater eagerness to supply leisure time to the firm at a given wage. We assume standard regularity conditions to ensure a well-behaved decision problem:

**Assumption 1.**  $c'(h) > 0$  and  $c''(h) < 0 \forall h \geq 0$ ;  $c(0) = 0$ ; and  $c'(0) = 1$ .

In words, the cost function is strictly increasing, strictly convex, and it is costless to abstain from working for the firm on a given day if one so chooses. The final assumption acts as a scale normalization for the distribution of  $\Theta_{li}$ .<sup>12</sup> The cost shock  $u_{it}^l$  represents day-to-day variation in a worker's time constraint that is outside her direct control.

Each day worker  $i$  optimizes her objective function  $\max_{h \in \mathbb{R}_+} \{wh - C_{it}(h; \Theta_{li}, \mathcal{T}_l)\}$ . The first-order condition implies the following intensive-margin calculation:

$$w = \Theta_{li} \mathcal{T}_l^{X_{1it}} c'(H_{it}) u_{it}^l. \quad (6)$$

---

<sup>12</sup>Note that any cost model  $C(h) = \Theta_{li}c(h)$  is equivalent to any other cost model  $\tilde{C}(h) = \tilde{\Theta}_{li}\tilde{c}(h)$ , where  $\tilde{\Theta}_{li} = a\Theta_{li}$  and  $\tilde{c}(h) = \frac{c(h)}{a}$  for some constant  $a > 0$ . Since the units of  $\Theta_{li}$  have no inherent meaning, the model can only be identified up to a scale normalization. Thus, we fix the units of total costs by re-scaling  $\Theta_{li} = \tilde{\Theta}_{li} \times \tilde{c}'(0)$  and  $c(h) = \tilde{c}(h)/\tilde{c}'(0)$ , with the result being that  $c'(0) = 1$ .

Since costs are strictly convex, equation (6) implies that labor supply is monotone decreasing in  $\Theta_{li}$ ,  $\mathcal{T}_l$ , and  $u_{it}^l$  (for  $u_{it}^l$  small enough), respectively, holding all else fixed.

Let  $G_H(h)$  denote the CDF of work times, conditional on  $H > 0$ . Since  $c'$  is strictly bounded away from zero, and since  $u^l$  exists on an unbounded support, there is positive probability that a corner solution of  $H_{it} = 0$  will result. That is, on days when cost shocks are very large, workers find it optimal to simply not work. Accordingly, the worker's extensive-margin decision obeys the following rule:

$$H_{it} > 0 \quad \Leftrightarrow \quad w > \Theta_{li}c'(0)u_{it}^l. \quad (7)$$

Note that we have implicitly assumed the worker has observed only the daily shock  $u_{it}^l$  when she makes her extensive-margin decision. Since she must show up to work in order to observe treatment status  $X_{1it}$  it does not enter her extensive-margin decision.

Once again, we say there is a *selection effect* present if the composition of worker types in a firm's labor force is not invariant to its choice  $X_0$ , or  $G_l(\Theta_{li}|X_0 = 1) \neq G_l(\Theta_{li}|X_0 = 0)$ . However, in the labor-supply model the sign of the shift may be ex-ante ambiguous. If the CSR firm is successful at attracting a more productive set of workers who produce higher quality work, their outside employment options may be more valuable, leading to higher  $\Theta_{li}$ . Alternatively, if the CSR-selected workers derive utility from contributing to the firm's socially beneficial mission, their costs of supplying time to the firm may fall.

## 4 Identification and Estimation Overview

Recall from Section 2.4 that for worker  $i = 1, \dots, I$  we have the following observables: recruitment status,  $X_{0i}$ ; wage offer  $W_i \in \{w_1, w_2\}$ ; total cumulative production  $Q_i$ ; and a full panel (of length  $Q_i$ ) of unit-specific outcomes, including production times, accuracy rates, and treatment values,  $\{\tau_{qi}, D_{qi}, A_{qi}, X_{1qi}\}_{qi=1}^{Q_i}$ ; and a panel (of length  $T$ ) of day-specific total work hours, and treatments,  $\{H_{it}, X_{1it}\}_{t=1}^T$ . In addition, we have the following aggregate observables from the recruiting stage,  $\{\mu_{N,1}, \mu_{N,2}, \mu_{C,1}, \mu_{C,2}\}$ , denoting application rates within each of the 4 potential worker pools recruited under conditions  $(X_0 = 0, W = w_1)$ ,  $(X_0 = 0, W = w_2)$ ,  $(X_0 = 1, W = w_1)$ , and  $(X_0 = 1, W = w_2)$ , respectively. The complete model consists of the following structural components to be recovered from the observables:  $\{G_{pdal}(\cdot, \cdot, \cdot, \cdot), \mathcal{T}_p, \mathcal{T}_d, \mathcal{T}_a, \mathcal{T}_l, \delta, \tau_1, \sigma_l, c(\cdot)\}$ .

In the interest of brevity, we provide an intuitive overview here of how raw empirical moments derived from the observables are able to uniquely pin down the model

parameters. We also briefly describe our GMM estimator that implements this identification strategy on our field experiment data. We defer a more complete discussion of identification and estimation in full technical rigor to Appendix A.

## 4.1 Productivity and Work Quality

It is straightforward to see that for the productivity and work quality models identification obtains because they describe processes which produce panel data with multiple within-worker observations, where the unobserved worker characteristics,  $(\Theta_{pi}, \Theta_{di}, \Theta_{ai})$  play the role of a worker fixed effect.

Estimates are obtained via a simple differencing routine as commonly done in panel-data settings. For treatment effects  $\mathcal{T}_j$ ,  $j = p, d, a$ , we difference mean production times/accuracy rates across treatment states ( $X_{1q_i} = 0$  vs  $X_{1q_i} = 1$ ), and then average these differences across individuals. We then average initial production times across workers to obtain  $\tau_1$  and  $D_0$ . To estimate the learning effect  $\delta$  we difference production times across consecutive units of output within individuals, and then average these differences both within and across workers. With these common parameters known, estimating individual fixed effects  $\Theta_{ji}$ ,  $j = p, d, a$ , is essentially a within-worker mean of production times or accuracy rates, conditional on treatment and/or learning. An important caveat is that for the accuracy model there is a necessary (though fairly intuitive) bias correction described in Appendix A.4 due to the interaction of non-linearity in  $\Phi(\cdot)$  and finite-sampling variation in observed accuracy rates  $A_{q_i}$ .

## 4.2 Labor Supply

Our identification strategy for the labor-supply model is more novel, but also more involved. As a starting point, we build upon a recent identification framework developed by [Torgovitsky \(2015\)](#) and [D’Haultfoeuille and Février \(2015\)](#) (henceforth DFT). These authors showed how a version of our model without labor-supply cost shocks could be identified non-parametrically—i.e., without ex-ante restrictions on the functional form of the cost function  $c(\cdot)$  or type distribution  $G_l(\cdot)$ —using exogenous wage contract variation. Identification in DFT hinges on a crucial exclusion restriction: the underlying distributions of leisure preferences within two groups receiving different wage contracts must be the same. This allows the researcher to establish an empirical mapping between a worker’s observed labor supply to the firm and her unobserved value of time  $\Theta_l$ . For completeness, in Appendix A.3.1 we provide

a brief synopsis of the DFT framework and identification result as it relates to our project. We extend this framework in two distinct ways.

First, our model allows for within-worker variation in day-to-day labor supply costs, through the shock term  $u_{it}^l$ . This rationalizes the fact that in the data we see non-trivial variation in individuals' labor-leisure choices. We therefore use within-worker variation in work-time choices to identify the distribution of the shock. However, this process is complicated by an important challenge. If we think of each worker as having seen a random sample of  $T$  labor-supply shocks, then their presence in the model also creates a sample selection problem: as illustrated in equation (7), for any finite value of  $\Theta_l$  the worker's choices will have a mass point at zero. I.e., the sample of labor-supply shocks behind the worker's positive work-hour days is non-random if the worker abstains from work on at least one day. However, within our model it turns out that the distribution of the order statistics of supply shocks (i.e., the  $k^{\text{th}}$  smallest of  $\{u_{i1}^l, \dots, u_{iT}^l\}$ ) *across* workers is a random sample.<sup>13</sup> Moreover, by adopting a parametric assumption on the distribution of  $u_{it}^l$  (log-normality), one can derive explicit expressions for the expectations of these order statistics in order to correct for the sample selection problem.

Second, we extend DFT by relaxing the central exclusion restriction through expansion of the observables. This extension is perhaps the more important of the two from an applied perspective, as it allows us to study selection on wage offer. The additional observables that we add into the mix are hiring-stage application rates. To fix ideas, suppose the firm begins with two equivalent groups of potential employees, and it offers wage  $w_1$  to the first and  $w_2 > w_1$  to the second. Suppose further that not all workers signal their acceptance by formally applying, but the firm sees a 30% increase in application rate from the second pool of workers, relative to the first.

We can infer some things about worker types who would apply under each scenario. E.g., both sets must have the same lower-bound  $\underline{\Theta}_{l1} = \underline{\Theta}_{l2} = \underline{\Theta}_l$ , since workers with very low value of time would be happy to accept either offer. Moreover, the upper-bounds of both sets must satisfy  $\bar{\Theta}_{l1} < \bar{\Theta}_{l2}$ , since any worker who was willing to accept the firm's low offer of  $w_1$  would also be willing to accept  $w_2$  to work at the same firm doing the same job. Finally, we can infer that the influx of 30% more applicants under offer  $w_2$  have time costs in the upper tail interval  $(\bar{\Theta}_{l1}, \bar{\Theta}_{l2}]$ . Since the two groups of potential employees were equivalent to begin with, the additional applications must

---

<sup>13</sup>This result also depends on an independence condition (see Assumption 5 in Appendix A.3).



be accounted for by marginal workers who were unwilling to accept hourly wage  $w_1$  (because their time was too valuable) but were willing to accept wage  $w_2$ .

Suppose now that the firm hires all applicants. After observing workers' labor-leisure decisions for a time, the firm can deduce who these marginal applicants were by compiling a list of the 30% lowest labor-supply employees: these are they whose high value of time prevents them from working more hours under  $w_2$  and would have also prevented them from accepting  $w_1$ . From this it follows that if  $\mathcal{S}_1$  is the set of all low-wage workers, and if  $\mathcal{S}_2$  is the 70% of high-wage workers who log the most hours over time, then these two sets of workers must have underlying types  $\Theta_l$  that follow the same distribution. This second selection correction is explained in full detail in Appendix A.3. Importantly, our identification strategy allows for point identification of *all* workers' idiosyncratic supply costs, including the 30% marginal applicants. Once the subsets  $\mathcal{S}_1$  and  $\mathcal{S}_2$  are computed to satisfy the DFT exclusion restriction, the common cost function  $c(\cdot)$  is nonparametrically identified. With this known, *all* workers'  $\Theta_l$  can be recovered using the FOCs, including the complement of  $\mathcal{S}_2$ .

The labor supply estimator is somewhat involved, but still akin to a basic moment matching scheme. We begin by computing the empirical work-time CDFs from these two sets  $\widehat{G}_{H|\mathcal{S}_1}(\cdot)$  and  $\widehat{G}_{H|\mathcal{S}_2}(\cdot)$  using raw data. We then specify the common cost function  $c(\cdot)$  as a flexible B-Spline function with light shape restrictions to enforce the regularity conditions in Assumption 1. Using equations (6) and (7) it is straightforward (though notationally intensive) to show that one can derive analytic, model-generated analogs to the work-time CDFs,  $\check{G}_{H|\mathcal{S}_1}(h|\mathcal{S}_1; \sigma_l, \boldsymbol{\pi}^c, \mathcal{T}_l, \boldsymbol{\Theta}_l)$  and  $\check{G}_{H|\mathcal{S}_2}(h|\mathcal{S}_2; \sigma_l, \boldsymbol{\pi}^c, \mathcal{T}_l, \boldsymbol{\Theta}_l)$  as functions of the log-normal shock CDF (with variance parameter  $\sigma_l$ ), the cost function  $c(\cdot)$  (with B-Spline weights  $\boldsymbol{\pi}^c$ ), as well as treatment and fixed-effects. We then execute a GMM estimator which chooses values of model parameters to match the means of shock order statistics and to fit the model-generated work-time CDFs as closely as possible to their empirical analogs. Intuitively, mean within-worker labor supply pins down the fixed effects, differences across order-statistics of worker shift length pin down the variance of the shock, and differences across various quantiles of the work-time CDFs pin down the shape of the cost function.

One interesting aspect of the estimator is worth mentioning. For each worker in wage group  $j$ , and for a given guess of her fixed effect  $\Theta_{li}$ , one can use equations (6) and (7) to infer her counterfactual labor supply under wage  $j' \neq j$ . This counterfactual prediction is part of our GMM estimator as we construct the model-generated work-



time CDFs. This approach and our flexible parameterization of the cost function implicitly builds the seemingly non-constructive DFT identification logic directly into the estimator. We utilize a block-bootstrap procedure for standard errors which mimics our experimental sampling and allows for correlation in per-unit production shocks. The interested reader is directed to Appendix A for further discussion.

### 4.3 Distributions of Worker Characteristics

A final challenge is that our data structure is an unbalanced panel. Thus,, we are able to make more precise inferences about the unobserved characteristics of workers who log more work days and produce more output. For this reason we use inverse variance weighting in our GMM estimator and in our estimates of the worker characteristics CDFs  $\widehat{G}_{pdal}(\cdot, \cdot, \cdot, \cdot)$  (see Appendix A). In our results section we display various graphs of “weighted CDFs”, where the label refers to inverse variance weighting. Finally, we also employ a bootstrapped variant of the Kolmogorov-Smirnov testing procedure for stochastic dominance in part to address the unbalanced panel problem.

## 5 Empirical Results

Our discussion of results begins with a survey of reduced-form insights from the hiring stage and subsequent work stage. We then explore structural results including treatment effects, unobserved characteristics and selection. We conclude with counterfactual model simulations that capture the impacts of recruitment strategies on a firm’s cost structure, through their influence on the pool of worker traits.

### 5.1 Application Rates and Descriptive Statistics

The field experiment was conducted during the fall of 2016 in Austin, TX; Baltimore, MD; Boston, MA; Dallas, TX; Houston, TX; Indianapolis, IN; Jacksonville, FL; Los Angeles, CA; New York City, NY; Philadelphia, PA; Phoenix, AZ; and San Diego, CA. A total of N=1,103 job description letters (with randomized pecuniary and non-pecuniary variation) were sent out via email to individuals who had expressed preliminary interest in the job. These individuals were randomly assigned to the four treatment cells: (1) \$11 wage rate and neutral (n=266), (2) \$11 wage rate and CSR information (n=309), (3) \$15 wage rate and neutral information (n=277), and (4) \$15 wage rate and CSR information (n=251). Table 2 presents application rates in each treatment cell and differences across cells.

<b>APPLICATION RATES</b>					
	<i>Mean</i>	<i>Estimate</i>		<i>SD</i>	<i>90% CI</i>
Neutral, $w = \$11$ :	$\mu_{N,11}$	0.301	—	0.028	[0.245, 0.356]
Neutral, $w = \$15$ :	$\mu_{N,15}$	0.408	—	0.030	[0.350, 0.466]
CSR, $w = \$11$ :	$\mu_{C,11}$	0.382	—	0.028	[0.327, 0.436]
CSR, $w = \$15$ :	$\mu_{C,15}$	0.494	—	0.032	[0.432, 0.556]

<b>APPLICATION RATE DIFFERENCES</b>					
<b>within wage offer group</b>					
	<i>Mean Difference</i>	<i>Estimate</i>	<i>P-value</i>	<i>SD</i>	<i>90% CI</i>
CSR-\$11 vs. CSR-\$15:	$\mu_{C,15} - \mu_{C,11}$	0.112	< 0.001	0.003	[0.107, 0.117]
Ntr-\$11 vs. Ntr-\$15:	$\mu_{N,15} - \mu_{N,11}$	0.107	< 0.001	0.002	[0.102, 0.112]
\$11 vs. \$15:	$\mu_{15} - \mu_{11}$	0.110	< 0.001	0.002	[0.106, 0.113]
<b>within nonpecuniary recruitment group</b>					
	<i>Mean Difference</i>	<i>Estimate</i>	<i>P-value</i>	<i>SD</i>	<i>90% CI</i>
CSR-\$11 vs Neutral \$11:	$\mu_{C,11} - \mu_{N,11}$	0.081	< 0.001	0.3425	[0.077, 0.086]
CSR-\$15 vs Neutral \$15:	$\mu_{C,15} - \mu_{N,15}$	0.086	< 0.001	0.3425	[0.081, 0.091]
CSR vs Neutral:	$\mu_C - \mu_N$	0.084	< 0.001	0.1171	[0.080, 0.087]

Table 2: Recruitment Stage Results

Not surprisingly, the raw data show that an increase in wage offer from \$11/hr to \$15/hr induced a substantial increase in application rates of 11 percentage points, or roughly a 32% increase in total applications. The data also show that including information about CSR in the job description increases applications by three quarters as much as a \$4 wage offer increase: CSR-recruited workers applied at a rate of 8.4 percentage points higher than neutral recruits, an increase of roughly 24% in total applications. Both shifts are statistically significant, as are the effects broken down by the four stage-1 treatment bins. It is worth mentioning that the observed application rates and wage elasticities are roughly similar to those found in comparable studies (e.g., [Flory et al. \(2014\)](#); [Leibbrandt and List \(2014\)](#)). Thus we see a greater volume of workers self-selecting into the labor pool with CSR, but the question remains whether the composition of characteristics within the worker pool changes as well. This is the focus of our structural model.

A total of 170 subjects were hired for the work-task stage of the experiment; Table 3 presents summary statistics. Each subject was observed during a period of 10 days, which implies 1700 individual-day observations. Most subjects did not

<b>WITHIN-WORKER SUMMARY STATISTICS</b>						
<i>Variable</i>	<i>Median</i>	<i>Mean</i>	<i>SD</i>	Min	Max	#Obs
Days worked	5	5.118	3.075	1	10	170
Avg. Hours worked per day	0.517	1.359	1.914	0.014	12.336	170
Avg. Hours worked per day $ H_{it} > 0$	1.323	1.987	1.992	0.092	13.707	170
Within-person work hour standard deviation	0.796	1.195	1.076	0.042	6.042	170
Avg. Productive Seconds per Output	123.85	139.84	74.70	17.03	445.52	170
Avg. Down-Time Seconds per Output	34.63	41.21	29.31	7.71	242.88	170
Mean Accuracy Rate	0.596	0.590	0.128	0.290	0.857	170

<b>WAGE-OFFER GROUP DIFFERENCES: <math>Group\ Outcome _{\\$15} - Group\ Outcome _{\\$11}</math></b>					
<i>Outcome</i>	<i>Median Difference</i>	<i>Mean Difference</i>	<i>P-Value for <math>H_0: Equal\ Means</math></i>	<i>90% CI</i>	<i>DF</i>
Hours worked per day	0.1636	0.4579	$1.3 \times 10^{-4}$	[0.262, 0.654]	1,698
Productive Seconds Per Output	-15.89	-7.97	$< 10^{-6}$	[-9.20, -6.74]	62,136
Down-Time Seconds Per Output	0.0145	0.3607	0.3222	[-0.239, 0.960]	62,136
Accuracy Rate Per Output	0.0789	0.0715	$< 10^{-6}$	[0.068, 0.075]	62,136

<b>NON-PECUNIARY GROUP DIFFERENCES: <math>Group\ Outcome _{CSR} - Group\ Outcome _{Ntr}</math></b>					
<i>Outcome</i>	<i>Median Difference</i>	<i>Mean Difference</i>	<i>P-Value for <math>H_0: Equal\ Means</math></i>	<i>90% CI</i>	<i>DF</i>
Hours worked per day	0.0697	-0.3612	0.0023	[-0.556, -0.167]	1,698
Productive Seconds Per Output	-9.59	-17.19	$< 10^{-6}$	[-18.34, -16.03]	62,136
Down-Time Seconds Per Output	-1.119	-1.738	$< 10^{-6}$	[-2.304, -1.172]	62,136
Accuracy Rate Per Output	0.0263	0.0216	$< 10^{-6}$	[0.019, 0.025]	62,136

Table 3: Summary statistics: Work Stage.

supply a positive amount of hours every day. The mean number of days worked across all subjects was roughly 5 out of 10. The mean daily shift time was roughly 2 hours, for average earnings of about \$27 per work day. On average, workers would process 54 images on a given shift, implying a per-unit average cost of \$0.498.

The table shows a striking degree of variation across workers. On a given day, many workers could not find time to work, while others occasionally logged shifts upwards of 16 hours. There is considerable heterogeneity on productivity and work quality as well. Mean per-unit production times varied across individuals from 17 seconds to 7.4 minutes. Similarly, mean per-unit paid down-time ranged from 8 seconds to 4 minutes across workers. Slightly less than one quarter of all paid worker

time is down-time, on average. Most workers produced fairly high-quality output but there was substantial heterogeneity in individuals’ mean accuracy, ranging between 30% and 86%.

As with application rates, we see substantial raw differences by recruitment cells. High-wage recruits logged more hours per day, though this change confounds several factors to be disentangled by the structural model. While there is a mechanical increase in their marginal incentives to supply time, underlying labor-supply costs may also be different due to selection. The raw shifts also do not account for the role of within-person, day-to-day variation in time costs. For CSR recruits we see a raw mean reduction in hours worked which is more suggestive of a shift in unobserved costs, since monetary incentives are held fixed. We also see increases in productivity and accuracy among high-wage recruits and CSR recruits, though these numbers do not account for within-person variation or work-stage treatment effects as will be done in our fixed-effects estimator.

## 5.2 Results: General Parameters

Before moving to a discussion of treatments and selection on worker characteristics, we briefly discuss some basic model parameters and measures of model fit displayed in Table 4. For active productivity, we find weak evidence of a learning-by-doing effect. Median production time on the first unit is roughly 3.7 minutes, and the parameter estimate for  $\delta$  is positive and small. This suggests that workers learn to be more productive, with most learning happening in the first 30 or so units of production. However, the 90% confidence interval for  $\delta$  includes zero, which implies that we cannot reject the null hypothesis of no learning effect. The productivity model produces an  $R^2$  of 0.73, which suggests that transitory productivity shocks play a minor, though non-trivial, role in accounting for variation in log production times.<sup>14</sup>

Figure 20 in online supplement C displays estimated learning curves with confidence bands for the 10<sup>th</sup> percentile, median, and 90<sup>th</sup> percentile individuals. After learning has largely ceased, the mean difference between the 10<sup>th</sup> and 90<sup>th</sup> percentile workers is a difference in production times by a factor of over 2.5. Down-time measures are smaller than production times, though transitory shocks play a larger role since

---

<sup>14</sup> $R^2$  for the passive and active productivity models is measured by plugging model estimates into log-transformed versions of equations (1) and (2), and taking the ratio of the variance in the left-hand side to the variance in the predicted component of the right-hand side.

	<i>Parameter</i>	<i>Estimate</i>	<i>Std</i>		
			<i>Error</i>	<i>P-value</i>	<i>90% CI</i>
<b>Learning:</b>	$\delta$	0.1467	0.1347	0.2760	[-0.075, 0.368]
			(H <sub>0</sub> : $\delta=0$ )		
<b>Mean Initial</b>					
<b>Production Time (minutes):</b>	$\tau_1$	3.7016	2.8851	—	[1.022, 13.411]
<b>Active Productivity Model Fit:</b>	$R^2$	0.7314	—	—	—
	(N = 62, 138)				
<b>Baseline Mean</b>					
<b>Down-Time (minutes):</b>	$D_0$	0.3587	0.0425	—	[0.289, 0.429]
<b>Passive Productivity Model Fit:</b>	$R^2$	0.3342	—	—	—
	(N = 62, 138)				
<b>Labor Supply Shock</b>					
<b>Standard Deviation:</b>	$\sigma_l$	0.7658	0.4216	—	[0.5135, 2.0928]

Table 4: Basic Model Parameter Estimates

the  $R^2$  for passive productivity is 0.33. Still, worker heterogeneity remains significant: the 10<sup>th</sup> and 90<sup>th</sup> percentiles of  $\Theta_{di}$  differ by a factor of 4.

For accuracy and labor supply, both models achieve a good fit in the sense that the accuracy model is able to closely predict mean worker-level accuracy rates and the labor supply model estimates are able to replicate the empirical distributions of daily work times quite closely. Figure 21 in online supplement C graphically depicts goodness of fit for accuracy and labor supply. We find that mean within-person standard deviation in accuracy rates is roughly 14.13 percentage points, while the cross-person standard deviation in baseline accuracy rates is 12.91 percentage points.<sup>15</sup> In other words, within the estimated model, permanent worker heterogeneity is quite important, but transitory, within-worker quality shocks also play a substantial role.

For labor supply, this idea is summarized by  $\hat{\sigma}_l$ , which implies a standard deviation of 1.1974 in transitory labor-supply cost shocks. Contrast this to the standard deviation of  $\Theta_l$ , which is 7.0736. Though this may seem a large difference, the two

<sup>15</sup>Within-person standard deviation in accuracy rates is a weighted mean of  $\sigma_{Ai} = \frac{Q_{Ci}\sigma_{Ai}^C + Q_{Ni}\sigma_{Ai}^N}{Q_{Ci} + Q_{Ni}}$ , where  $\sigma_{Ai}^j = \sqrt{\frac{\sum_{q_i=1}^{Q_i} (A_{q_i} - \mu_{Ai}^j)^2 \mathbb{1}(X_{1q_i} = \mathbb{1}(j=C))}{Q_{ji} - 1}}$ , and  $\mu_{Ai}^j = \frac{\sum_{q_i=1}^{Q_i} A_{q_i} \mathbb{1}(X_{1q_i} = \mathbb{1}(j=C))}{Q_{ji}}$ ,  $j = C, N$ . Note that  $\sigma_{Ai}^j$  is excluded whenever  $Q_{ji} < 2$ . Cross-person standard deviation in baseline accuracy rates is the square root of a weighted mean of  $(\Phi(\Theta_{ai}) - \mu_{\hat{A}i})^2$ , where  $\mu_{\hat{A}i}$  is a weighted mean of  $\Phi(\Theta_{ai})$ . Weights on individual  $i$  are calculated as  $\sqrt{Q_i}$ , and normalized to sum to one.

	<i>Parameter</i>	<i>Estimate</i>	<i>P-value</i>	<i>Std Error</i>	<i>90% CI</i>
<b>Active Productivity:</b>	$\mathcal{T}_p$	0.7577	$2.8 \times 10^{-4}$	0.0618	[0, 0.840]
			$(H_0 : \mathcal{T}_p = 1)$		(one-sided)
<b>Passive Productivity:</b>	$\mathcal{T}_d$	0.5752	0.0021	0.1158	[0, 0.737]
			$(H_0 : \mathcal{T}_d = 1)$		(one-sided)
<b>Accuracy/Work Quality:</b>	$\mathcal{T}_a$	-0.0154	0.0017	0.0049	[-0.024, 0.007]
			$(H_0 : \mathcal{T}_a = 0)$		
<b>Labor Supply Costs:</b>	$\mathcal{T}_l$	1.0204	0.5166	0.0314	[0.969, 1.072]
			$(H_0 : \mathcal{T}_l = 1)$		

Table 5: Parameter estimates: Treatment Effects

numbers are much closer when expressed as fractions of their means. Consider a thought experiment where the average worker type  $E[\Theta_l]$  experiences a mean cost shock  $E[U^l]$  on a given day. If we reduce that shock by one standard deviation, her supply costs drop by 58.47%. Alternatively, if we maintain her shock at the mean but reduce her permanent type  $E[\Theta_l]$  by a standard deviation, her supply cost falls by 58.58%. I.e., permanent, cross-worker variation plays roughly the same size role as day-to-day, within-worker variation in aggregate variation of labor supply to the firm. Figure 11 in Appendix A.10 depicts the estimated cost function  $c(h; \hat{\pi}^c)$ , which displays substantial curvature.

### 5.3 Results: Work-Stage Treatment Effects

Table 5 displays estimates for work-stage treatment effects. The interpretation of these parameters is a measure of how an existing worker’s behavior changes (holding unobserved characteristics fixed) when she is prompted about how her work contributes to the firm’s efforts to make the world a better place.

Our first result is that the firm saw substantial productivity gains from the internal CSR advertisement. Active production times dropped by 24.2% under treatment, while non-productive downtime dropped by 42.5%. Both of these shifts are statistically significant. Given the large speed-up in production of output, the expected sign of the treatment effects for work quality may be less obvious. While CSR treatment may inspire workers to exert themselves in producing accurate output, they are producing more quickly and taking less rest time in between. This could lead to a drop in quality through burnout or mental fatigue. In the third row of Table 5 we

see a negative point estimate, but it is negligible, being only enough to reduce mean accuracy for the average worker by about half of a percentage point. This implies that the work-stage CSR treatment substantially increases productivity and that this production speed-up is virtually all valuable to the firm. As for labor supply, we see no significant change in willingness to supply time to the firm, as the confidence interval for  $\mathcal{T}_l$  contains 1.

## 5.4 Results: Unobserved Characteristics and Selection

Here we examine the effect of recruitment-stage variation on the composition of unobserved characteristics within the firm’s labor force. In doing so, we present several comparisons to explore the role of selection on wage and CSR. We first compute the empirical CDFs of worker types separately for the subsamples of CSR recruits and neutral recruits. We then run a bootstrapped testing procedure for stochastic dominance (see Appendices A.8 and A.9.2), repeating the same process for the subsamples of high-wage recruits and low-wage recruits. We also assess combined effects in two ways. First, we compare the baseline sample of neutral recruits to the CSR sample, where the latter are also subject to our estimated work-stage treatment effects. Second, we compare the subsample of workers who were neutrally recruited *and* offered a low wage to the subsample who were offered a high wage *and* CSR recruited.

Tables 9–10 in Appendix A.10 summarize all stochastic dominance tests. For each one our testing procedure, which follows Marmer et al. (2017), compares the null hypothesis of distributional equality to multiple alternative hypotheses: first-order dominance in one direction, first-order dominance in the opposite direction, and non-equality with no dominance either way. Whenever the null hypothesis of equality is rejected, we report the mean percent change in the relevant characteristic both in levels and as a fraction of a standard deviation for the subsamples in the test. We close this section by exploring correlations between productivity, work quality, and the value of a worker’s time.

### 5.4.1 Productivity

Figure 2 displays the distributions of estimated active productivity types  $\Theta_p$ . The step functions are the empirical CDFs used in the stochastic dominance tests, and the continuous curves are smoothed B-spline representations for display purposes. The first feature one notices is a substantial cross-worker heterogeneity. Since the

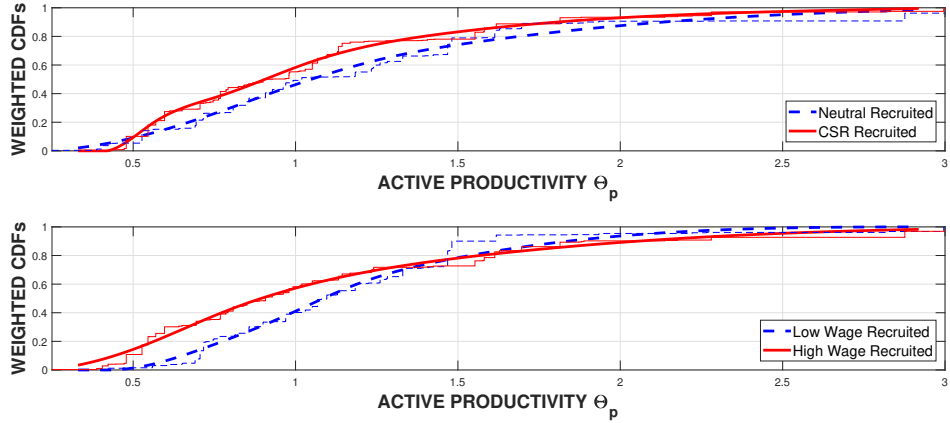


Figure 2: Active Productivity: Selection on Wage and CSR

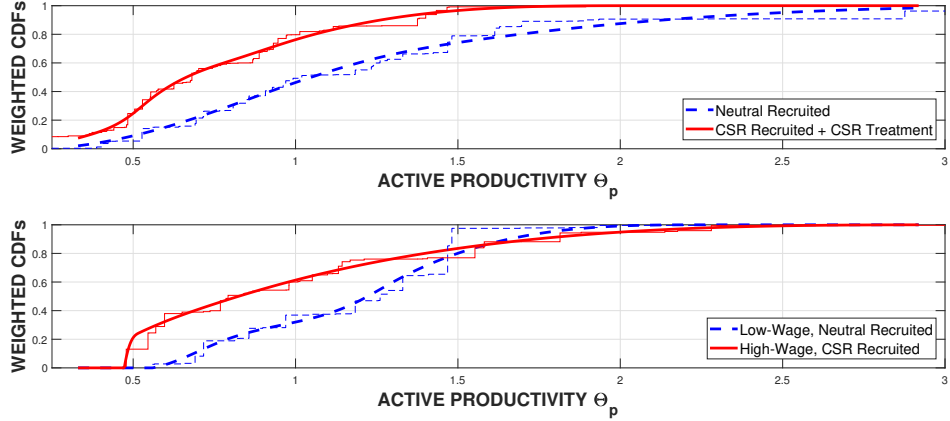


Figure 3: Active Productivity: Combined Effects

unconditional median fixed effect is roughly one, it follows that the 10<sup>th</sup> percentile worker averages production times at roughly half of the median worker, while the 90<sup>th</sup> averages production times at roughly twice those of the median worker.

Figure 2 also shows how selection occurs on both pecuniary and non-pecuniary variation in job offers. In the top panel, the Neutral recruited subsample has a distribution of productivity types that dominate those of the CSR recruited workers. Table 9 contains the results of our stochastic dominance tests and shows that model estimates indeed reject the null hypothesis of distributional equality in favor of advantageous worker selection. We estimate that CSR-recruited workers have productivity types (i.e., mean production times) that are 15.14% lower, on average, which amounts to a reduction of 0.261 standard deviations. As for productivity selection on wage



offers, the results (lower panel of Figure 2) are similar though less decisive: there is a crossing of the two empirical CDFs at about the 80<sup>th</sup> percentile. The null of equality is still strongly rejected, though weakly in favor of the alternative of inequality, rather than a specific dominance ordering. On average, the mean change in productivity under wage selection is positive, though smaller.

The top panel of Figure 3 shows the combined effects of CSR selection and CSR work-stage treatment. Given the large treatment effects discussed in the previous section, the difference between these two distributions is even more striking. The lower panel of the figure shows the combined selection effect of high-wage-CSR recruits versus low-wage-neutral recruits. Interestingly, the joint selection effect is stronger than the sum of the two individual selection margins: together they account for a 0.548 standard deviation improvement. This suggests socially conscious CSR and money are actually complementary recruitment tools.

Figures 9 and 10 depicting estimated CDFs of passive productivity characteristics are relegated to Appendix A.10 since selection effects for  $\Theta_d$  are small and far less conclusive. The estimated CDFs have multiple crossing points and no clear ordering arises. However, cross-worker heterogeneity remains substantial, especially in the upper tail. Most workers' types reside on a relatively narrow band, while workers within the top quintile tend to waste large amounts of off-task time at work.

#### 5.4.2 Accuracy/Work Quality

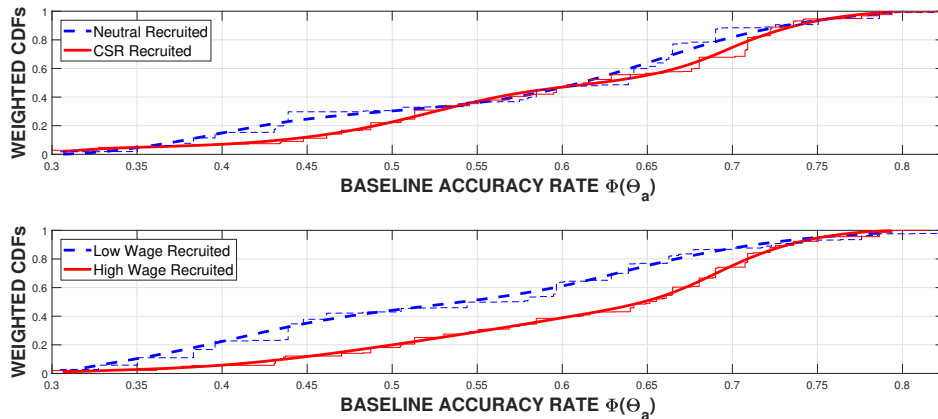


Figure 4: Accuracy Rates: Selection on Wage and CSR

Figure 4 and the left panel of Table 10 display results on individual work quality

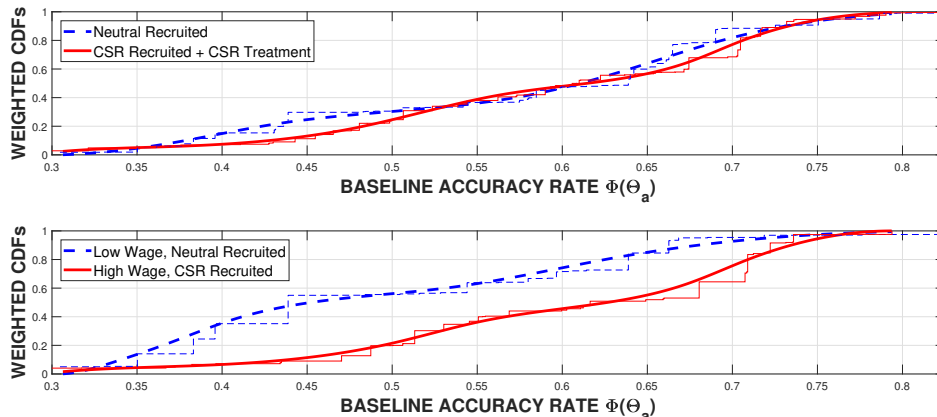


Figure 5: Accuracy Rates: Combined Effects

fixed effects. The figure presents distributions of  $\Phi(\Theta_a)$ , which represent a worker’s baseline average accuracy rate in the absence of a treatment effect. The step functions represent the empirical CDFs used in our stochastic dominance tests, and the continuous functions are smoothed B-spline CDFs for display purposes. Once again, the estimates point to a substantial degree of heterogeneity. Figure 4 indicates that recruiting with either high wage offers or CSR advertisement leads to a first-order dominance shift toward workers who produce superior output (both significant at the 1% level). For CSR, the gains come largely from the second and fourth quintiles of worker quality and result in roughly a 4% (0.165 standard deviations) increase in mean accuracy. For high wage, the change is larger at about 13% (0.518 standard deviations) higher mean accuracy. Though these numbers seem somewhat small, we will see through our model simulations that they can have a big impact on the firm’s bottom line through quality-control costs.

In Figure 5 (lower panel) we see once again evidence that CSR and money can be complementary tools for producing advantageous worker selection. The first-order stochastic dominance shift for high-wage-CSR recruits, relative to low-wage-neutral recruits, is highly significant and larger than the sum of the two individual selection effects. The mean accuracy rate improvement is 21% (or 0.785 standard deviations).

### 5.4.3 Worker Labor Supply Costs

Figure 6 and the right panel of Table 10 summarize results on type distributions of  $\Theta_l$ , the shadow value that workers place on their time when making labor-leisure

decisions. Once again, we see a striking level of cross-worker heterogeneity. Relative to the median worker, the 10<sup>th</sup> percentile worker’s cost of supplying time to the firm is 60% lower, and the 90<sup>th</sup> percentile worker’s supply cost is 80% greater than the median. This would account for the long upper tail of observed positive work times where the 90<sup>th</sup> percentile (6.46 hours) is 4 times larger than the median (1.61 hours).

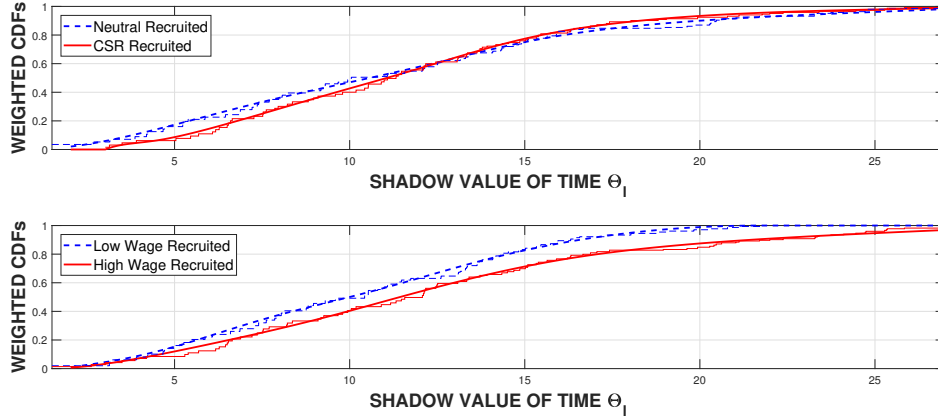


Figure 6: Shadow Value of Time: Selection on Wage and CSR

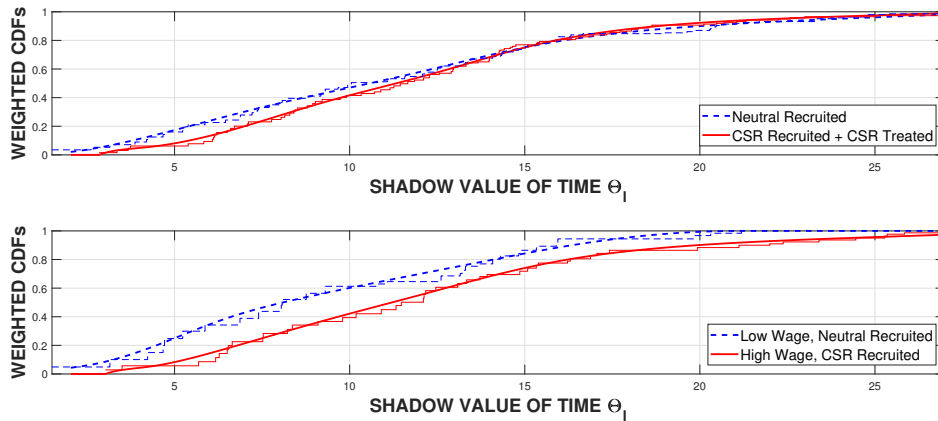


Figure 7: Shadow Value of Time: Combined Effects

Figure 6 depicts selective entry into the worker pool based on wage and CSR. As we noted earlier, selection on CSR recruitment operates similarly as selection on high-wage recruitment: in both cases the null hypothesis of distributional equality is rejected in favor of a shift that is generally consistent with first-order dominance relative to the neutral and low-wage baselines. The main difference is that CSR

selection is concentrated among the firm’s lowest-cost workers (i.e., those who work the most hours), with an overall 3% (0.042 standard deviations) increase in shadow value of time, while wage selection is concentrated among the firm’s highest-cost workers (those who work the least) with an overall increase of 18% (0.334 standard deviations) in the shadow value of their time.

Figure 7 displays compound labor supply effects. Once again, we see evidence of a complementary relationship between CSR status and high wage offer in driving worker selection. The lower pane of the figure shows an impact on workforce composition that is stronger than the sum of the two individual selection effects. The combined result is that high-wage-CSR recruits on average have time that is 29% (0.470 standard deviations) more valuable than the low-wage-neutral baseline.

One might have expected that CSR recruitment would gather a set of workers to the firm who are more eager to supply their time, but here we see that the opposite is true. When a prospective worker is faced with a firm’s offer  $(w, X_0, \mathbf{Z})$ , an increase to some  $w' > w$ , causes an influx of new workers with more valuable time who are willing to accept the offer under  $w'$  but not  $w$ . In similar fashion, holding  $w$  fixed, if CSR recruitment status is increased from  $X_0 = 0$  to  $X_0 = 1$  we once again see entry by an additional set of workers with more valuable time that are willing to accept a fixed wage under  $X_0 = 1$ , but not under  $X_0 = 0$ . In that sense, CSR advertisement seems to play a role similar to money in attracting workers to the firm.

Overall, it may not be surprising that CSR-selected and high-wage-selected workers have more valuable time, given our results that they are also more productive and produce higher quality output. One might expect, therefore, that their time is more valuable due to better outside options in the labor market, given these desirable characteristics. This interpretation is supported by results in Online Appendix C.2.1 where we compute various measures of correlations among our estimates of unobserved worker traits. There we find that variation in  $\hat{\Theta}_p$ ,  $\hat{\Theta}_d$ , and  $\hat{\Theta}_a$  explains 34% of variation in  $\hat{\Theta}_l$ . It is worth emphasizing that residual variation in  $\hat{\Theta}_l$  is partly sampling variability, so 34% is an estimated lower bound on the strength of the relationship between actual time value  $\Theta_l$  and worker productivity and quality  $(\Theta_p, \Theta_d, \Theta_a)$ .

## 6 Bottom-Line Model Simulations

To attach meaning to our estimates, we present model simulations that capture the impacts of advantageous worker selection on a firm’s cost structure. For each worker

in our sample recruited under wage offer  $w$ , we have an estimate of their individual labor-supply cost curve. This allows us to compute counterfactual outcomes under alternative wage raises  $w' > w$  which are assessed after the hiring stage.<sup>16</sup> In order to assess the final impact of selection on wage offer and CSR on the firm’s profitability, we simulated 400 blocks of 10 workdays for each worker in our sample as follows:

1. We assume the worker is inexperienced before their first day, with alternating neutral/CSR task framing on odd/even days. For each worker, we simulate 10 iid draws from the (parametric) distribution of labor-supply shocks  $u^l$  to obtain total work time within each day.
2. We iteratively simulate iid draws from the empirical joint distribution of shocks  $(u^p, u^d)$  to obtain production times, until the simulated work-time for that day is exhausted.<sup>17,18</sup> This determines daily supply of output and per-unit costs.
3. Finally, for each unit of output we simulate from the parametric distribution of accuracy shocks  $u^a$  and record the implied accuracy level for the worker in question, on a single field of the web-form, on each unit of output.

Using this information we measured for subsample  $j$  how selection effects mold distributions of total daily output  $F_{out,j}(q)$ , per-unit costs,  $F_{cost,j}(c)$ , and output quality,  $F_{acc,j}(a)$ , in absence of redundancy—that is, where one unit of output is considered to be one worker-image pair. In many contexts, work quality is not a separate dimension of the production process, but rather, it feeds into production costs through creating a need for costly monitoring and evaluation, or through work redundancy—i.e., having multiple workers redo the same task and aggregating their results to ensure final output quality—as in our data-entry context. Therefore, we executed a second simulation to produce a mapping between single-worker accuracy and the need for different levels of redundancy to ensure a given level of final output quality.

---

<sup>16</sup>We only consider wage raises because labor laws constrain firms from reducing wages after hiring a worker with contract  $w$ . Moreover, an ex-post wage decrease would likely induce attrition.

<sup>17</sup>We assume a worker will not exceed her planned work-time from step 1. At shift end, if she finds that she has a few minutes left, but not enough to accomplish another unit of output, she spends the remaining minutes in downtime. Dropping final minutes instead made little difference.

<sup>18</sup>We found that prolonged learning effects led to unusually low simulated production times for the highest output workers, toward the end of the 10 days. Thus, after the 200<sup>th</sup> unit of output we hold the learning term fixed at  $q^{-\delta} = 200^{-\delta}$  for subsequent output. This learning cutoff produced a close match between the simulated CDF of per-unit production times and its analog in the data.

Specifically, we begin by assuming for simplicity that the objectively correct response for a field-image pair is 3 out of 5 on the Likert scale. For a hypothetical set of  $K$  workers from subsample  $j$ , we simulate  $4 \times 10^6$  sets of  $k = 1, \dots, K$  random draws,  $\{a_1, \dots, a_K\}$ , from the single-worker accuracy distribution  $F_{acc,j}(a)$ . For each set of  $K$ , we then simulate  $K$  uniform random numbers  $\nu_k^{acc}$  and code worker  $k$ 's response as correct if  $\nu_k^{acc} \leq a_k$ . Otherwise, we randomly select a response from among the remaining options according to the empirical distribution of incorrect responses in our raw data.<sup>19</sup> We then recorded a binary variable for whether the simulated modal response was the correct one, and averaged across all  $4 \times 10^6$  observations to get a  $K$ -worker modal response accuracy rate. Since different subsamples of workers (e.g., CSR recruits vs neutral recruits) have different distributions of single-worker accuracy rates, this exercise results in differing levels of redundancy (i.e.,  $K$ ) needed to achieve a given level of accuracy in the modal response. We present results below.

## 6.1 Production Costs and Daily Supply W/o Redundancy

Our model of labor supply focuses on an individual's decision to supply her time to the firm, but the measure most relevant to an employer is not employee time *per se*, but rather, supply of units of output to the firm each day. For various reasons, cross-group comparisons on daily output supply may be ambiguous: some workers are more productive than others, but they also tend to have higher value of time and therefore supply less of it to the firm. Moreover, a worker's wage rate is also correlated with her productivity and labor supply since it induces selection effects.

Suppose a firm wished to solve a cost minimization problem, subject to producing a fixed quantity of outputs. On one hand, the firm could recruit a set of more productive workers with high initial wage offers. On the other hand, it could recruit a set of less productive workers with low initial wage offers, and adopt the strategy of offering ex-post wage-raise incentives to increase output capacity. How would this choice affect per-worker daily supply of output, and per-unit production costs? Moreover, how does the recruitment-stage CSR advertisement affect this cost-minimization calculus compared to a neutral recruitment strategy?

To answer these questions, for each group comparison in our study—i.e., \$11 re-

---

<sup>19</sup>Empirically, when the correct response is 3 out of 5, conditional on responding incorrectly, workers tended to select 1 or 5 roughly one third of the time, and 2 or 4 roughly two thirds of the time. This empirical distribution was relatively stable across different Likert-scale fields.

<i>Recruitment Group:</i>	<b>\$15</b>	<b>\$11</b>	<b>CSR</b>	<b>NTR</b>	<b>CSR15</b>	<b>NTR11</b>
<b>DAILY OUTPUT:</b>	52.71	38.10	44.83	48.15	56.17	45.45
<b>% Differences:</b>	$\% \Delta(\$15, \$11)$	+38.4%	$\% \Delta(CSR, NTR)$	-6.9%	$\% \Delta(CSR15, NTR11)$	+23.6%
<b>PER-PAGE COST:</b>	\$0.555	\$0.436	\$0.473	\$0.553	\$0.500	\$0.446
<b>% Differences:</b>	$\% \Delta(\$15, \$11)$	+27.5%	$\% \Delta(CSR, NTR)$	-14.4%	$\% \Delta(CSR15, NTR11)$	+12.0%
<b>OUTPUT-EQUIV. RAISE:</b>	---	$R^* = \$2.87$	$R^* = \$0.63$	---	---	$R^* = \$1.89$
<b>CF PER-PAGE COST:</b>	---	\$0.538	\$0.495	---	---	\$0.519
<b>CF % Differences:</b>	$\% \Delta(\$15, \$11 + R^*)$	+3.1%	$\% \Delta(CSR + R^*, NTR)$	-10.6%	$\% \Delta(CSR15, NTR11 + R^*)$	-3.7%

Table 6: Cost and Volume Simulations: Baseline and Counterfactual (no redundancy)

recruits versus \$15 recruits; neutral recruits (NTR) versus CSR recruits; and \$11-neutral (NTR11) recruits versus \$15-CSR recruits (CSR15)—we compute the counterfactual *output-equivalent raise*, denoted  $R^*$ , needed for the lower-output group to match mean per-worker daily supply in the other group under baseline wages. We then compare per-unit costs (for a single page processed by one worker) between the two groups at baseline and under the counterfactual wage raise scenario. Table 6 reports relevant means, and Figures 22 – 24 in the online supplemental appendix display simulated CDFs of daily output, per-unit costs, and per-unit accuracy rates.<sup>20</sup> In Table 6, percent changes  $\% \Delta(k, j) \equiv 100 \times (outcome_k - outcome_j) / outcome_j$  are reported, with the convention that group  $j$  (the second argument) is used as the baseline group.

For \$15 recruits relative to \$11 recruits, the former are getting paid more per hour and are on average more productive. These two effects dominate the fact that they have higher opportunity costs of time, so that \$15 recruits average about 53 processed pages per day, or one third more than their \$11 recruit counterparts. However, per-unit costs at baseline are roughly 56 cents, or 27% higher for \$15 recruits. An ex-post wage-raise of  $R^* = \$2.87$  per hour assessed to the \$11 recruits would be enough to equalize mean daily output supply in absence of redundancy. Since these individuals have (on average) lower value of time, this counterfactual raise is strictly less than the \$4 gap between their initial wage offers. However,  $R^*$  would nearly equalize per-page costs to the firm, relative to \$15 recruits, due to cross-group productivity differences.

The comparison between CSR recruits and NTR recruits is somewhat different.

<sup>20</sup>Note that all daily output means reported are unconditional: each worker has positive probability of logging zero units of output on a given day, and these zeros are included as part of the worker’s estimated supply curve under a given wage contract (see Figure 22 in the online supplement).

CSR recruits are more productive, but also have more valuable time. The latter fact dominates so that they produce about 45 units per day on average, or 7% less than their NTR counterparts. An hourly wage increase of  $R^* = \$0.63$  per hour for CSR recruits would be sufficient to close this output gap. The productivity difference results in a more favorable per-page cost among CSR recruits under both the baseline and counterfactual scenarios. The relevant comparisons for the CSR15 group versus the NTR11 group are naturally between the other two in absence of redundancy.

The per-page cost savings among CSR recruits provide an incomplete picture though, since they don't factor in the cost of the firm's CSR operations. Knowing this extra piece of information would allow for a more natural comparison to the \$11 versus \$15 case where the cost of inducing advantageous selection is explicit. The numbers in the middle section of Table 6 imply that a firm could profitably execute a CSR program, based solely on advantageous worker selection, given a budget at or below \$3.58 per worker, per day, under the baseline wage contracts. Therefore, a constant-returns-to-scale firm like ours could profitably sustain an annual CSR budget of \$1,000,000, if it had enough business volume to continuously occupy the labor of 766 workers or more under the status quo contract, which we refer to as *minimum effective scale*.<sup>21</sup> This result provides a possible explanation for why larger firms tend to invest more resources in CSR activities: with a larger employee base over which to spread CSR costs and reap productivity benefits, the profitability of a dollar invested in CSR activity may be higher as well.

## 6.2 Production Costs and Daily Supply with Quality Control

Thus far we have assumed that one unit of output is a single worker-task pair. This idea is appropriate in many contexts (e.g., retail sales), but in other applications (e.g., janitorial work, data entry, various service professions) the quality of an individual worker's output may directly impact cost structure. For example, if a janitorial worker fails to adequately clean a room, the firm may need to dispatch a second employee to redo the job. Likewise, if an exam evaluator does not maintain adequate concentration while scoring an essay, the firm may need to increase the number of her co-workers who re-score the same essay in order to produce a reliable score. In our

---

<sup>21</sup>This simple calculation is as follows:  $minimum\ effective\ scale \equiv \left\lceil \left( \frac{\$1,000,000}{\$3.58 \times 365} \right) \right\rceil = 766$ . We refer to this measure as "minimum effective scale" rather than the more common "minimum efficient scale" since it ignores the role of human resources costs.



data-entry application this is the case: our Likert-scale variables require cognitive effort to process training materials and thoughtfully evaluate each field for each Google Street View image. Following industry practice, we have adopted the modal response across  $K$  independent evaluations as our criterion for determining final output.

Table 7 summarizes the results of our redundancy simulations. We define *mode accuracy rate*, denoted  $A_{mode,K}$ , as the probability that the modal response among  $K$  independent evaluations corresponds to the true value (being 3 in our simulations). We then assumed the firm wishes to maintain accuracy rates at various levels in the set  $A^* \in \{\frac{2}{3}, \frac{4}{5}, \frac{9}{10}\}$  and then computed *required redundancy* as the minimum number  $K^*$  of independent evaluations which result in  $A_{mode,K^*} \geq A^*$ . Given a value of  $K^*$ , we then compute *adjusted daily output* by dividing the output measure from the previous section by  $K^*$ , and *adjusted per-unit cost* by multiplying the per-unit cost from the previous section by  $K^*$ . Finally, we repeat our counterfactual exercise by computing *adjusted output-equivalent raises* and percent differences for subgroup comparisons.

Table 7 shows that the accuracy differences across subgroups are enough to induce non-trivial differences in redundancy requirements. For \$11 recruits, the firm needs between 1 and 3 extra independent evaluations to ensure a fixed mode accuracy, relative to \$15 recruits. For NTR recruits, the requirement is 1 or 2 extra evaluations relative to CSR recruits. This comparison is even more striking for the combined selection effect: among NTR11 recruits, the firm requires between 2 and 8 extra independent evaluations to maintain the same mode accuracy as CSR15 recruits.

These extra redundancy requirements shift the cost structure in favor of high-wage and CSR recruits. Under the status quo contract, per-unit costs across the two wage groups are nearly equalized, while \$15 recruits produce 66%–73% more output per day. Since one employee hired at a \$15 wage offer can accomplish what 1.66 employees (or more) hired at an \$11 wage offer can, if fixed human resources costs (e.g., hiring, payroll, management, etc.) are important then attracting workers by offering to pay them more per hour may actually be *better* for the firm’s bottom line, all things considered. We also note that from the baseline wage contract of \$11/hour, the adjusted output-equivalent counterfactual raises for these recruits actually require pay rates *at or above* the hourly rate for their \$15-recruited counterparts, with the latter producing counterfactual per-unit cost savings of 25%-29%. The broad lesson is that if work quality has a direct influence on the firm’s cost structure it may actually be better to recruit a high-productivity, high-quality workforce with high wage offers.

<i>Recruitment Group:</i>	<i>\$15</i>	<i>\$11</i>	<i>CSR</i>	<i>NTR</i>	<i>CSR15</i>	<i>NTR11</i>	
<i>Mode Accuracy Rate = <math>\frac{2}{3}</math></i>	<b>REDUNDANCY:</b>	$K^* = 5$	$K^* = 6$	$K^* = 5$	$K^* = 6$	$K^* = 5$	$K^* = 7$
	<b>ADJ. DAILY OUTPUT:</b>	10.54	6.35	8.97	8.03	11.23	6.49
	<b>% Differences:</b>	$\% \Delta(\$15, \$11)$	+66.0%	$\% \Delta(CSR, NTR)$	+11.7%	$\% \Delta(CSR15, NTR11)$	+73.0%
	<b>ADJ. PER-UNIT COST:</b>	\$2.776	\$2.614	\$2.366	\$3.318	\$2.5008	\$3.124
	<b>% Differences:</b>	$\% \Delta(\$15, \$11)$	+6.2%	$\% \Delta(CSR, NTR)$	-28.7%	$\% \Delta(CSR15, NTR11)$	-20.0%
	<b>OUTPUT-EQUIV. RAISE:</b>	---	$R^* = \$5.03$	---	$R^* = \$0.93$	---	$R^* = \$6.52$
	<b>CF PER-UNIT COST:</b>	---	\$3.702	---	\$3.521	---	\$4.891
<b>CF % Differences:</b>	$\% \Delta(\$15, \$11 + R^*)$	-25.01%	$\% \Delta(CSR, NTR + R^*)$	-32.8%	$\% \Delta(CSR15, NTR11 + R^*)$	-48.9%	
<i>Mode Accuracy Rate = <math>\frac{4}{5}</math></i>	<b>REDUNDANCY:</b>	$K^* = 8$	$K^* = 10$	$K^* = 8$	$K^* = 9$	$K^* = 9$	$K^* = 13$
	<b>ADJ. DAILY OUTPUT:</b>	6.59	3.81	5.60	5.35	6.24	3.50
	<b>% Differences:</b>	$\% \Delta(\$15, \$11)$	+72.9%	$\% \Delta(CSR, NTR)$	+4.8%	$\% \Delta(CSR15, NTR11)$	+78.5%
	<b>ADJ. PER-UNIT COST:</b>	\$4.442	\$4.356	\$3.785	\$4.977	\$4.500	\$5.801
	<b>% Differences:</b>	$\% \Delta(\$15, \$11)$	+2.0%	$\% \Delta(CSR, NTR)$	-23.9%	$\% \Delta(CSR15, NTR11)$	-22.4%
	<b>OUTPUT-EQUIV. RAISE:</b>	---	$R^* = \$5.33$	---	$R^* = \$0.14$	---	$R^* = \$6.89$
	<b>CF PER-UNIT COST:</b>	---	\$6.277	---	\$5.014	---	\$9.279
<b>CF % Differences:</b>	$\% \Delta(\$15, \$11 + R^*)$	-29.2%	$\% \Delta(CSR, NTR + R^*)$	-24.5%	$\% \Delta(CSR15, NTR11 + R^*)$	-51.5%	
<i>Mode Accuracy Rate = <math>\frac{9}{10}</math></i>	<b>REDUNDANCY:</b>	$K^* = 13$	$K^* = 16$	$K^* = 13$	$K^* = 15$	$K^* = 14$	$K^* = 22$
	<b>ADJ. DAILY OUTPUT:</b>	4.06	2.38	3.45	3.21	4.01	2.07
	<b>% Differences:</b>	$\% \Delta(\$15, \$11)$	+70.3%	$\% \Delta(CSR, NTR)$	+7.4%	$\% \Delta(CSR15, NTR11)$	+94.2%
	<b>ADJ. PER-UNIT COST:</b>	\$7.218	\$6.970	\$6.151	\$8.295	\$6.999	\$9.817
	<b>% Differences:</b>	$\% \Delta(\$15, \$11)$	+3.6%	$\% \Delta(CSR, NTR)$	-25.8%	$\% \Delta(CSR15, NTR11)$	-28.7%
	<b>OUTPUT-EQUIV. RAISE:</b>	---	$R^* = \$5.23$	---	$R^* = \$0.62$	---	$R^* = \$8.47$
	<b>CF PER-UNIT COST:</b>	---	\$9.996	---	\$8.613	---	\$17.044
<b>CF % Differences:</b>	$\% \Delta(\$15, \$11 + R^*)$	-27.8%	$\% \Delta(CSR, NTR + R^*)$	-28.6%	$\% \Delta(CSR15, NTR11 + R^*)$	-58.9%	

Table 7: Redundancy Simulations

As for the individual impact of selection on CSR, the redundancy requirement is enough to reverse the output ordering between the CSR and NTR groups, with the former now being 5%–12% ahead in terms of adjusted daily output. This change then further magnifies the cost difference, with CSR workers’ output now being 24%–29% less expensive under the status quo wage contracts. Redundancy expands the scope for CSR profitability: under the status quo contracts (counterfactual wage raises) a firm can operate a CSR program on a budget of between \$6.67–\$8.54 (\$6.88–\$10.36) per worker, per day. This brings minimum effective scale down considerably (see Table 8), with a conservative estimate being 411 workers or more.

<i>Group Comparison</i> →	<i>CSR vs NTR</i> ( <i>status quo</i> )	<i>CSR vs NTR</i> ( <i>output-equiv</i> <i>raise CF</i> )	<i>CSR15 vs NTR11</i> ( <i>status quo</i> )	<i>CSR15 vs NTR11</i> ( <i>output-equiv</i> <i>raise CF</i> )
<i>Case</i> ↓				
<b>NO REDUNDANCY:</b>	766	973	∞	2,541
<b>MODE ACCURACY</b> = $\frac{2}{3}$ :	321	265	392	103
<b>MODE ACCURACY</b> = $\frac{4}{5}$ :	411	399	338	92
<b>MODE ACCURACY</b> = $\frac{9}{10}$ :	371	323	243	69

Table 8: Minimum Effective Scale (# workers) for Annual CSR Budget of \$1,000,000

## 7 Concluding Remarks

The “business of business is business” has been a mantra associated with Milton Friedman for nearly half a century. Having articulated his view in a 1970 New York Times article that the only responsibility of business was to increase its profits, he described contrarians as “puppets of the intellectual forces that...undermine the basis of a free society.” Today, over 90% of major businesses have specific programs dedicated to corporate social responsibility, and many CEOs and business school deans trumpet their organizations’ commitment to social engagement. While a narrow interpretation of Friedman certainly has proven incorrect—the business world has evolved in a much more social manner than he likely anticipated—but this does not mean that CSR is necessarily at odds with profit seeking.

This study takes a systematic approach to exploring the supply-side effects of CSR. Much previous scholarship has focused on the demand side with mixed results. This has led to hotly-contested debates in corporate law, business ethics, economics, and the social sciences more broadly. Our study combines theory with a natural field experiment to explore how a firm can use CSR to impact the process through which its workers produce output. Our study is unique in that we begin with a well-specified model of CSR and estimate its behavioral parameters using a field experiment.

Our results are generally positive for CSR, in that practical, market-based approaches can be used to affect the supply side of business. Advertising the firm’s CSR endeavors during recruiting increases application rates by almost as much as an increase of hourly wages from \$11 to \$15. In addition, CSR selection works to improve productivity, quality-adjusted supply of output to the firm, and per-unit production costs. A work-stage internal CSR advertisement campaign also increases productivity

of all workers by speeding up production outside of break time, and reducing workers' propensity to take elective breaks between productive tasks.

We conclude that CSR should not be viewed as a necessary distraction from a profit motive, but rather as an important part of profit maximization similar to other non-pecuniary incentives offered to the firm's labor force. Our results highlight that the real challenge for CEOs is not whether CSR should be used, but what are the best social innovations to put in place. We view our work as only a start on how to leverage economic theory and field experiments to lend insights into CSR and non-pecuniary worker compensation generally. Future work will refine our approach and utilize different forms and incentives of CSR to provide a more complete understanding of both the demand and supply sides of the market.

Finally, of independent interest is our finding that, taking quality control and productivity into account, a firm's cost structure may actually improve with higher wage offers during the recruitment process. The resulting set of employees may compensate the firm for their more expensive time by consistently producing more and better output. Of course, this is also not entirely unrelated to the CSR theme: a core finding of this study is that more sustainable and/or socially conscious business practices may not be at odds with profit maximization after all. Moreover, sprinkled throughout our results is the notion that important complementarities exist between CSR and wage offers, both in selection and motivation of workers.

## References

- Ashraf, N., Berry, J., and Shapiro, J. M. (2010). Can higher prices stimulate product use? evidence from a field experiment in zambia. *American Economic Review*, 100:2383–2413.
- Barnett, M. L. and Salomon, R. M. (2012). Does it pay to be really good? addressing the shape of the relationship between social and financial performance. *Strategic Management Journal*, 33(11):1304–1320.
- Barrett, G. F. and Donald, S. G. (2003). Consistent tests for stochastic dominance. *Econometrica*, 71(1):71–104.
- Bertrand, M., Bombardini, M., Fisman, R., and Trebbi, F. (2018). Tax-exempt lobbying: Corporate philanthropy as a tool for political influence. *mimeo, University of British Columbia Economics Dept.*

- Burbano, V. C. (2015). Getting virtual workers to do more by doing good: Field experimental evidence. Technical report, Working paper.
- Burbano, V. C. (2016). Social responsibility messages and worker wage requirements: Field experimental evidence from online labor marketplaces. *Organization Science*, 27:1010–1028.
- Carnahan, S., Kryscynski, D., and Olson, D. (2015). How corporate social responsibility reduces employee turnover: evidence from attorneys before and after 9. Technical report, Working paper.
- Chandler, J., Mueller, P., and Paolacci, G. (2014). Nonnaivite among amazon mechanical turk workers: Consequences and solutions or behavioral researchers. *Behavior research methods*, 46(1):112–130.
- Chen, X. (2007). Large sample sieve estimation of semi-nonparametric models. *Handbook of Econometrics*, 6(76):5549–5632.
- Cotton, C., Hickman, B., List, J., and Price, J. (2019). Productivity versus motivation: Combining field experiments with structural econometrics to study adolescent human capital production. *Working paper, University of Chicago Economics Department*.
- Davidson, R. and Duclos, J.-Y. (2000). Statistical inference for stochastic dominance and for the measurement of poverty and inequality. *Econometrica*, 68(6):1435–1464.
- de Boor, C. (2001). *A Practical Guide to Splines Revised Edition*. New York: Springer-Verlag.
- DellaVigna, S., List, J. A., and Malmendier, U. (2012). Testing for altruism and social pressure in charitable giving. *The Quarterly Journal of Economics*, 127:1–56.
- DellaVigna, S., List, J. A., Malmendier, U., and Rao, G. (2017). Voting to tell others. *The Review of Economic Studies*, 84:143–181.
- DellaVigna, S., List, J. A., Malmendier, U., and Rao, G. (2019). Estimating social preferences and gift exchange at work, nber 22043.
- D’Haultfoeuille, X. and Février, P. (2011). The provision of wage incentives: A structural estimation using contracts variation. *mimeo, CREST*, 2011-29.
- D’Haultfoeuille, X. and Février, P. (2015). Identification of nonseparable triangular models with discrete instruments. *Econometrica*, 83(3):1199–1210.
- Doob, J. L. (1949). Heuristic approach to the kolmogorov-smirnov theorems. *Annals of Mathematical Statistics*, 20:393–403.
- Du, S., Bhattacharya, C., and Sen, S. (2011). Corporate social responsibility

- and competitive advantage: Overcoming the trust barrier. *Management Science*, 57(9):1528–1545.
- Elfenbein, D. W., Fisman, R., and McManus, B. (2012). Charity as a substitute for reputation: Evidence from an online marketplace. *The Review of Economic Studies*, 79(4):1441–1468.
- Flory, J. A., Leibbrandt, A., and List, J. A. (2014). Do competitive workplaces deter female workers? A large-scale natural field experiment on job-entry decisions. *The Review of Economic Studies*, 82(4).
- Friedman, M. (1970). The social responsibility of business is to increase its profits. *The New York Times Magazine*, pages 32–33.
- Gill, D. and Prowse, V. (2012). A structural analysis of disappointment aversion in a real effort competition. *American Economic Review*, 102(1):469–503.
- Godfrey, P. C., Merrill, C. B., and Hansen, J. M. (2009). The relationship between corporate social responsibility and shareholder value: An empirical test of the risk management hypothesis. *Strategic Management Journal*, 30(4):425–445.
- Goodman, J. K., Cryder, C. E., and Cheema, A. (2013). Data collection in a flat world: The strengths and weaknesses of mechanical turk samples. *Journal of Behavioral Decision Making*, 26(3):213–224.
- Greening, D. W. and Turban, D. B. (2000). Corporate social performance as a competitive advantage in attracting a quality workforce. *Business & Society*, 39(3):254–280.
- Harrison, G. W. and List, J. A. (2004). Field experiments. *Journal of Economic Literature*, 42(4):1009–1055.
- Heathcote, A., Brown, S., Wagenmakers, E. J., and Eidels, A. (2010). Distribution-free tests of stochastic dominance for small samples. *Journal of Mathematical Psychology*, 54:454–463.
- Hickman, B. R., Hubbard, T. P., and Paarsch, H. J. (2017). Identification and estimation of a bidding model for electronic auctions. *Quantitative Economics*, 8(2):505–551.
- Huang, J. Z. (2003). Local asymptotics for polynomial spline regression. *Annals of Statistics*, 31(5):1600–1635.
- Kolmogorov, A. (1933). Sulla determinazione empirica di una legge di distribuzione. *Giornale dell’Istituto Italiano degli Attuari*, 4:83–91.
- Leibbrandt, A. and List, J. A. (2014). Do Women Avoid Salary Negotiations? A

- Large-Scale Natural Field Experiment. *Management Science*, 61(9).
- Levitt, S. D. and List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? *The journal of economic perspectives*, 21(2):153–174.
- Linton, O., Song, K., and Whang, Y.-J. (2010). An improved bootstrap test of stochastic dominance. *Journal of Econometrics*, 154:186–202.
- List, J. A. and Momeni, F. (2017). When corporate social responsibility backfires: Theory and evidence from a natural field experiment. *NBER 24169*.
- Ljungqvist, L. and Sargent, T. J. (2004). *Recursive Macroeconomic Theory*. MIT Press.
- Marmer, V., Shneyerov, A., and Kaplan, U. (2017). Identifying collusion in english auctions. *mimeo, University of British Columbia Economics Dept*.
- McDonnell, M.-H. and King, B. (2013). Keeping up appearances reputational threat and impression management after social movement boycotts. *Administrative Science Quarterly*, 58(3):387–419.
- Paolacci, G. and Chandler, J. (2014). Inside the turk: Understanding mechanical turk as a participant pool. *Current Directions in Psychological Science*, 23(3):184–188.
- Roback, J. (1982). Wages, rents, and the quality of life. *The Journal of Political Economy*, pages 1257–1278.
- Rosen, S. (1986). The theory of equalizing differences. *Handbook of labor economics*, 1:641–692.
- Servaes, H. and Tamayo, A. (2013). The impact of corporate social responsibility on firm value: The role of customer awareness. *Management Science*, 59(5):1045–1061.
- Smirnov, N. (1939). Sur les Écarts de la courbe de distribution empirique. *Matematicheskii Sbornik NS*, 6:3–26.
- Torgovitsky, A. (2015). Identification of nonseparable models using instruments with small support. *Econometrica*, 83(3):1185–1197.
- Waddock, S. A. and Graves, S. B. (1997). The corporate social performance-financial performance link. *Strategic management journal*, pages 303–319.
- Wooldridge, J. (2001). *Econometric Analysis of Cross Section and Panel Data*. Cambridge: MIT Press.

# A Appendix: Identification & Estimation Details

Here we present a full discussion of structural identification in order to make clear how the empirical moments in our data pin down various model parameters during the estimation phase. By way of notation, let  $\Phi(\cdot, \mu, \sigma)$  denote the normal distribution with mean  $\mu$  and standard deviation  $\sigma$ , and let  $\phi(\cdot, \mu, \sigma)$  denote its density.

**Assumption 2.**  $\varepsilon_{q_i}^a \sim \Phi(\varepsilon_{q_i}^a, 0, 1) \forall q_i$ , and  $\varepsilon_{it}^l \equiv \log(u_{it}^l) \sim \Phi(\varepsilon_{it}^l, 0, \sigma_l) \forall i, t$ .

Recall that for worker  $i = 1, \dots, I$  we have the following observables: recruiting status,  $X_{0i}$ ; wage offer  $W_i \in \{w_1, w_2\}$ ; total cumulative production  $Q_i$ ; unit-specific production times, accuracy ratings, and treatment values,  $\{\tau_{q_i}, D_{q_i}, A_{q_i}, X_{1q_i}\}_{q_i=1}^{Q_i}$ ; and day-specific total work hours, and treatments,  $\{H_{it}, X_{1it}\}_{t=1}^T$ . In addition, we have the following aggregate observables from the recruiting stage,  $\{\mu_{N,1}, \mu_{N,2}, \mu_{C,1}, \mu_{C,2}\}$ , denoting application rates within each of the 4 potential worker pools recruited under conditions  $(X_0 = 0, W = w_1)$ ,  $(X_0 = 0, W = w_2)$ ,  $(X_0 = 1, W = w_1)$ , and  $(X_0 = 1, W = w_2)$ , respectively. The complete model consists of the following structural components to be recovered from the observables:  $\{G_{pdal}(\cdot, \cdot, \cdot, \cdot), \mathcal{T}_p, \mathcal{T}_d, \mathcal{T}_a, \mathcal{T}_l, \delta, \tau_1, \sigma_l, c(\cdot)\}$ .

## A.1 IDENTIFICATION: Productivity

Given that our passive productivity model is essentially identical to our active productivity model with the restriction that  $\delta = 0$  in the former, it will suffice to discuss identification under the latter, for which we require a standard exogeneity condition on the log-error terms. Letting  $\varepsilon_{q_i}^p \equiv \ln(u_{q_i}^p)$  and  $\varepsilon_{q_i}^d \equiv \ln(u_{q_i}^d)$ , we require

**Assumption 3.**  $E[\varepsilon_{q_i}^p \ln(\Theta_{pi})] = 0$ ,  $E[\varepsilon_{q_i}^p] = 0$ ,  $E[\varepsilon_{q_i}^p \ln(q_i)] = 0$ , and  $E[\varepsilon_{q_i}^p X_{1q_i}] = 0$ ,  $E[\varepsilon_{q_i}^d \ln(\Theta_{pi})] = 0$ , and  $E[\varepsilon_{q_i}^d] = 0$   $E[\varepsilon_{q_i}^d X_{1q_i}] = 0$ .

In other words, the unit-specific shocks to production times are uncorrelated with unobserved worker types, treatments, or cumulative output. This condition follows from our randomization scheme described above.

Now, taking the natural logarithm of equation (1) gives us

$$\ln(\tau_{q_i}) = \ln(\tau_1) - \delta \ln(q_i) \mathbb{1}(q_i > 1) + \ln(\Theta_{pi}) + \ln(\mathcal{T}_p) X_{1q_i} + \varepsilon_{q_i}^p. \quad q_i = 1, 2, \dots, Q_i \quad (8)$$

Provided  $Q_i \geq 2$  for all  $i$ , and  $Q_i > 2$  for at least one  $i$ , Assumption 3 and equation (8) establish that the active productivity model is a standard linear regression with fixed effects, whose parameters are identified from the within-person panel structure of the data. Identification obtains in the passive productivity model for similar reasoning.



## A.2 IDENTIFICATION: Accuracy

For the accuracy model, we require the following:

**Assumption 4.**  $E[\varepsilon_{it}^l \Theta_{ai}] = 0$ ,  $E[\varepsilon_{q_i}^a] = 0$ ,  $E[\varepsilon_{q_i}^a q_i] = 0$ ,  $E[\varepsilon_{q_i}^a X_{1q_i}] = 0$ .

With the mean and scale normalization of the error term  $\varepsilon_{q_i}^a$  from Assumption 2, identification of  $\Theta_{ai}$  and  $\mathcal{T}_a$  follows a standard panel-data Probit argument:  $A_{q_i}$  is a Bernoulli random variable<sup>22</sup> whose conditional probability mass function is

$$\Pr[A_{q_i} = j | \Theta_{ai}, X_{1q_i}] = \Phi(\Theta_{ai} + X_{1q_i} \mathcal{T}_a)^j [1 - \Phi(\Theta_{ai} + X_{1q_i} \mathcal{T}_a)]^{1-j}, \quad j \in \{0, 1\}. \quad (9)$$

Therefore, the within-person panel structure of the observables are sufficient to identify treatment effects and worker fixed effects. Of course, during estimation a bias correction will be necessary due to the non-linearity in the normal CDF  $\Phi$ , but this can be accomplished without difficulty (see Appendix A.4 for further discussion).

## A.3 IDENTIFICATION: Labor Supply

Our strategy for labor supply builds upon and extends recent identification results proven by D’Haultfoeulle and Février (2015, Theorem 2) and Torgovitsky (2015, Theorem 2) (henceforth DFT). We begin by imposing some key identifying assumptions on the log of the error term in the labor cost equation:

**Assumption 5.**  $\varepsilon_{it}^l \perp\!\!\!\perp \ln(\Theta_{li})$ ,  $E[\varepsilon_{it}^l] = 0$ ,  $E[\varepsilon_{it}^l X_{1it}] = 0$ .

The independence assumption is related to a sample selection problem that we will address in Section A.3.2 below. Three further assumptions are necessary:

---

<sup>22</sup>In reality, our data contain a finer (but still discrete) accuracy measure as explained in Section 2.4. One could alternatively specify a more elaborate model where for each coded variable, say  $v_1, \dots, v_7$ , there is a variable-specific, iid, normal shock  $\varepsilon_{v_k}^a$ , so that  $A_{v_k, q_i}^* = \Theta_{ai} + X_{1q_i} \mathcal{T}_a + \varepsilon_{v_k}^a + \varepsilon_{q_i}^a$ ,  $A_{v_k, q_i} = \mathbb{1}(A_{v_k, q_i}^* > 0)$ , and  $A_{q_i} = \sum_{k=1}^7 A_{v_k, q_i} / 7$ . One could potentially identify the relative variances of the  $\varepsilon_{v_k}^a$ ’s (variable-specific shocks) after imposing a scale normalization—say,  $Var(\varepsilon_{q_i}^a) = 1$  for example—which is always a necessary identifying assumption with limited dependent variables models. Yet, knowing the relative variances of the variable-specific shocks is not central to the main theme of this paper, so we opt for a simpler approach. The identification argument and estimator are based on averaging accuracy measures within worker and across multiple units of output to estimate treatment effects and worker fixed effects. Since these averages belong to an asymptotically dense set, the method applies to any scenario where the values of  $A_{q_i}$  are limited to a finite grid and shocks are uncorrelated across different  $q_i$ ’s. Therefore, for the purpose of exposition we concentrate on the simpler case where  $A_{q_i}$  is binary.

**Assumption 6.** There exist known subsets  $S_1$  contained in the set of all workers offered wage  $w_1$  and  $S_2$  contained in the set of all workers offered wage  $w_2$  such that  $G_{l|S_1}(\Theta_l|S_1) = G_{l|S_2}(\Theta_l|S_2) = G_l^*(\Theta_l)$

**Assumption 7.** Letting  $c_j(\cdot)$  denote the cost function which applies to workers from wage group  $w_j$ ,  $j = 1, 2$ , we assume  $c_1(h) = c_2(h) = c(h)$ ,  $\forall h \in \mathbb{R}_+$ .

**Assumption 8.**  $G_l(\Theta_l)$  has full support, with  $g_l(\Theta_l) \geq a > 0$ ,  $\forall \Theta \in (\underline{\Theta}_l, \bar{\Theta}_l) \subset \mathbb{R}_{++}$ .

In words, Assumptions 6 and 7 provide crucial exclusion restrictions on which DFT's seminal identification strategy is based. In words, the econometrician must be able to observe two type-equivalent subsets of workers across contract groups (i.e., no selective entry conditional on  $(S_1, S_2)$ ) and the cost functions must be the same. Under these conditions, the variation in wage offers can essentially serve as an instrument to disentangle the common component of costs  $c(\cdot)$  from the idiosyncratic component  $\Theta_{li}$ . In what follows it will be useful to let  $G_{H|S_1}(h|S_1)$  and  $G_{H|S_2}(h|S_2)$  denote the distributions of *non-zero work times* logged, conditional on events  $S_1$  and  $S_2$ , respectively. Our recruitment-stage procedure induces self-selection on wage offer which might seem to threaten Assumption 6, but in the following sections we explain how our research design is crafted to solve this problem. Assumption 8 is a technical condition to avoid problems of partial identification, which will become clear below.

### A.3.1 Case I: Noiseless Labor Supply

To develop intuition for how different sources of variation in the data pin down model parameters, we first simplify the discussion along several dimensions. For now we will assume no selective entry so that  $S_1$  and  $S_2$  each constitute the full sets of workers offered wages  $w_1$  and  $w_2$ , respectively. Under this scenario, we have  $G_{l|S_1}(\Theta_l|S_1) = G_{l|S_2}(\Theta_l|S_2) = G_l^*(\Theta_l) = G_l(\Theta_l)$ . Also, let  $\underline{\Theta} = 0$  with  $G_l(0) = 0$ —so that work time distributions have unbounded support but individuals choose finite optimal  $h$ —and for convenience we assume  $G_{H|w_1}(0) = 0$ , whereas  $G_{H|w_2}(h) = 0$  at some  $h > 0$ . Finally, we assume for now degenerate labor-supply shocks where  $E[u_{it}^l] = 1$  and  $Var[u_{it}^l] = 0$  with  $X_{1it} = 0$  in every period  $t$ . Before proceeding, it will also be convenient to define  $h_j(\Theta)$  as the optimal choice profile consistent with first-order condition (6) (under degenerate supply shocks), given contract  $w_j$ ,  $j = 1, 2$ , and denote its inverse by  $\Theta_j(h)$ .

The mathematics behind the fully general frameworks of DFT are fairly complicated, but in the current setting with noiseless labor supply and no selective entry, a simple geometric argument, proposed by [D’Haultfoeuille and Février \(2011, henceforth, DF11\)](#), illustrates the intuition behind why the the model and the observables together uniquely pin down the structural primitives. First, note that Assumption 6 combined with the fact that  $h_j(\cdot)$  is strictly decreasing (by strict convexity of  $c(\cdot)$ ) implies that a fixed quantile of labor supply in either wage group (e.g., median hours worked under  $w_1$  or  $w_2$ ) corresponds to the unobserved type at that same quantile in the type distribution (i.e., median  $\Theta_l$ ):

$$\begin{aligned} G_l(\Theta_l) &= 1 - G_{H|S_1}(h_1(\Theta_l)|S_1) = 1 - G_{H|S_2}(h_2(\Theta_l)|S_2) \\ \Rightarrow h_1(\Theta_l) &= G_{H|S_1}^{-1} \left( G_{H|S_2}[h_2(\Theta_l)|S_2] \Big| S_1 \right). \end{aligned} \tag{10}$$

DF11 referred to this relationship as the *horizontal transform* operator: given knowledge of the labor supply distributions and some  $(\Theta_l, h_2(\Theta_l))$  pair under wage contract  $w_2$ , one can infer the counterfactual labor supply choice for the same worker type under contract  $w_1$ . Assumption 7 produces an inverse to this operation which DF11 refer to as the *vertical transform*. If we recall that  $\Theta_j(h) = \frac{w_j}{c'(h)}$  from the FOCs, then dividing these equations across the two contracts we obtain the following relationship:

$$\Theta_1(h) = \frac{w_2}{w_1} \Theta_2(h). \tag{11}$$

Given knowledge of some  $(h, \Theta_2(h))$  pair under contract  $w_2$ , one can infer the unobservable type  $\Theta_1(h)$  that would optimally choose the same  $h$  under contract  $w_1$ .

Finally, since  $c'(0) = 1$  we can infer that  $\Theta_1(0) = w_1$  lay on the inverse choice profile under contract  $w_1$ .<sup>23</sup> From this known  $(\Theta_l, h) = (w_1, 0)$  pair that lay on the supply curve under contract 1, we can perform a sequence of horizontal and vertical transform operations to infer other points that lay on both inverse choice profiles. This process is graphically depicted in Figure 8. Since the limiting behavior of the two CDFs is identical, it follows that both inverse choice mappings are non-parametrically identified since there is one and only one  $\Theta_l$  value that could rationalize each observed  $(h, w_j)$  pair under the model. This is because there is a (potentially infinite) sequence

---

<sup>23</sup>From this the reader can see that the scale normalization in Assumption 1 gives rise to an interpretation of  $c(\cdot)$  as being the labor supply cost profile for the baseline type  $\Theta_l = \frac{\Theta_1(0)}{w_1}$ , and an interpretation of idiosyncratic types as scaling costs up or down relative to the baseline cost profile.

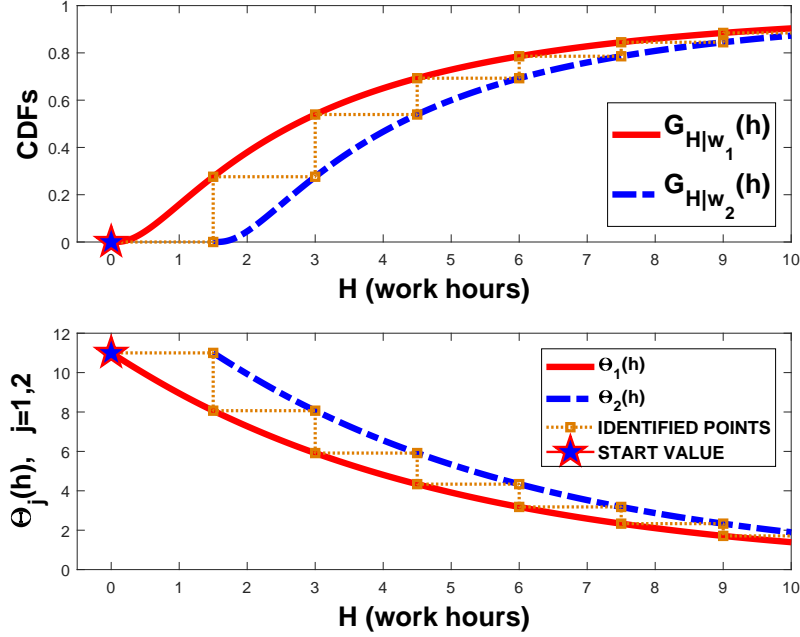


Figure 8: Geometry of Identification

of transform operations that can be constructed to connect any known point  $(h, \Theta_j(h))$  on some choice function, to any other point  $(h', \Theta_{j'}(h'))$  on another choice function. [Torgovitsky \(2015, Theorem 2\)](#) and [D'Haultfoeulle and Février \(2015, Theorem 2\)](#) formalize and generalize this geometric intuition.

Once either of the functions  $\Theta_j(h)$  are known, the econometrician can also use the FOCs to recover the cost function from the differential equation  $c'(h) = \frac{w_j}{\Theta_j(h)}$ ,  $c(0) = 0$ . Perhaps a more intuitive way to view the basic identification argument is that there is a unique form of the common cost function  $c(\cdot)$  (or equivalently, of  $G_l(\Theta_l)$ ) that could rationalize each observed  $(G_{H|S_1}(h|S_1), G_{H|S_2}(h|S_2))$  pair under the model. Although quite elegant, at first glance this identification logic may seem non-constructive as it does not immediately seem to suggest an estimation strategy. However, our estimator in [Section A.7](#) below implicitly builds in the vertical transform concept—by flexibly parameterizing  $c(\cdot)$ —and the horizontal transform concept—by using counterfactual predictions of worker behavior to match the empirical work-time CDFs.

### A.3.2 CASE II: Noisy Labor Supply

Now we return to the full model where within-person variation in work times across different days arises from transitory shocks to each worker's labor-supply costs. For

the time being we will assume that  $S_1$  and  $S_2$  are known (though they may now be strict subsets) and we defer discussion on how to derive them to the next section. The first complication that arises is one of sample selection: on a day when a worker receives a cost shock large enough to trigger a corner solution  $H_{it} = 0$ , it is not possible to make inference about the precise magnitude of her shock that day. Assumption 2  $\varepsilon_{it}^l = \log(u_{it}^l) \sim \phi(\varepsilon; 0, \sigma_l)$  helps to correct for this selection problem.

Note that each worker receives two iid samples of labor supply shocks: there are  $T/2$  shocks that occur on days when  $X_{1it} = 1$  and  $T/2$  additional shocks occur on days when  $X_{1it} = 0$ . To simplify notation, let  $T = 10$ . We denote the order statistics of log-shocks on CSR-treated days and Neutral days by  $\{\varepsilon_{C_i}^l(1:5) < \varepsilon_{C_i}^l(2:5) < \varepsilon_{C_i}^l(3:5) < \varepsilon_{C_i}^l(4:5) < \varepsilon_{C_i}^l(5:5)\}$ , and  $\{\varepsilon_{N_i}^l(1:5) < \varepsilon_{N_i}^l(2:5) < \varepsilon_{N_i}^l(3:5) < \varepsilon_{N_i}^l(4:5) < \varepsilon_{N_i}^l(5:5)\}$ , respectively. Since work times are monotone in shocks, we denote the corresponding order statistics of work times for CSR and Neutral workdays  $\{H_{C_i}(1:5) \geq \dots \geq H_{C_i}(5:5)\}$  and  $\{H_{N_i}(1:5) \geq \dots \geq H_{N_i}(5:5)\}$ , respectively. Note in this case that the indexing is reversed because higher cost shocks lead to lower optimal work-time choices.

We first show that if  $\sigma_l$  and  $c(\cdot)$  are known, then all other labor supply parameters can be treated as known. Recall that Assumption 2 implies the log-shock order statistics have the following densities:

$$\begin{aligned}
\varepsilon_{C_i}^l(1:5), \varepsilon_{N_i}^l(1:5) &\sim \phi_{(1:5)}(x; \sigma_l) = 5[1 - \Phi(x; \sigma_l)]^4 \phi(x; \sigma_l), \\
\varepsilon_{C_i}^l(2:5), \varepsilon_{N_i}^l(2:5) &\sim \phi_{(2:5)}(x; \sigma_l) = 20\Phi(x; \sigma_l)[1 - \Phi(x; \sigma_l)]^3 \phi(x; \sigma_l), \\
\varepsilon_{C_i}^l(3:5), \varepsilon_{N_i}^l(3:5) &\sim \phi_{(3:5)}(x; \sigma_l) = 30\Phi(x; \sigma_l)^2[1 - \Phi(x; \sigma_l)]^2 \phi(x; \sigma_l), \\
\varepsilon_{C_i}^l(4:5), \varepsilon_{N_i}^l(4:5) &\sim \phi_{(4:5)}(x; \sigma_l) = 20\Phi(x; \sigma_l)^3[1 - \Phi(x; \sigma_l)] \phi(x; \sigma_l), \text{ and} \\
\varepsilon_{C_i}^l(5:5), \varepsilon_{N_i}^l(5:5) &\sim \phi_{(5:5)}(x; \sigma_l) = 5\Phi(x; \sigma_l)^4 \phi(x; \sigma_l).
\end{aligned} \tag{12}$$

From these we can define  $\mathbb{E}_{(j:5)}(\sigma_l) \equiv E[\varepsilon_{ki}^l(j:5); \sigma_l] = \int_{-\infty}^{\infty} x \phi_{(j:5)}(x; \sigma_l) dx$  as the expectation of the  $j^{\text{th}}$  log-shock order statistic, for  $j = 1, \dots, 5$ . Moreover, equation (6) allows us to construct a model-generated analog to the supply shock for any of worker  $i$ 's positive observed work times, being  $E_{C_i}(j:5) \equiv \ln\left(\frac{W_i}{c^{[H_{C_i}(j:5)]}}\right) - \ln(\Theta_{li}) - \ln(\mathcal{T}_l)$ , if  $H_{C_i}(j:5) > 0$ , and  $E_{N_i}(j:5) \equiv \ln\left(\frac{W_i}{c^{[H_{N_i}(j:5)]}}\right) - \ln(\Theta_{li})$ , if  $H_{N_i}(j:5) > 0$ . For convenience, we simply define  $E_{ki}(j:5) = 0$  whenever  $H_{ki}(j:5) = 0$ , for  $k = C, N$ .

Moreover, equation (7) implies that whether or not  $H_{C_i}(j:5) > 0$  or  $H_{N_i}(j:5) > 0$  is true depends solely on whether  $\ln\left(\frac{W_i}{c^{(0)}}\right) - \ln(\Theta_{li}) < \varepsilon_{ki}^l(j:5)$  for  $k = C, N$ , respectively. Since  $\varepsilon_{C_i}^l(j:5)$  and  $\varepsilon_{N_i}^l(j:5)$  are independent of  $\Theta_{li}$  (Assumption 5), this implies

purely random data loss for the order statistics of log-shocks across individuals. In other words, although the within-person samples of supply shocks are systematically selected, the samples of order statistics for supply shocks *across people* are random. Therefore, the treatment effects  $\mathcal{T}_l$  and individual fixed effects  $\Theta_{li}$  are uniquely pinned down by the following set of up to  $2 \times 5$  moment conditions for each  $i$ :

$$\begin{aligned} \mathbb{E}_{(j:5)}(\sigma_l) - E \left[ \ln \left( \frac{W_i}{c'[H_{Ci}(j:5)]} \right) - \ln(\Theta_{li}) - \ln(\mathcal{T}_l) \right] &= 0, \quad \forall j = 1, \dots, 5 \text{ s.t. } H_{Ci}(j:5) > 0, \\ \mathbb{E}_{(j:5)}(\sigma_l) - E \left[ \ln \left( \frac{W_i}{c'[H_{Ni}(j:5)]} \right) - \ln(\Theta_{li}) \right] &= 0, \quad \forall j = 1, \dots, 5 \text{ s.t. } H_{Ni}(j:5) > 0. \end{aligned} \quad (13)$$

Thus,  $\Theta_l$  and  $\mathcal{T}_l$  are known if  $\sigma_l$  and  $c(\cdot)$  are known, since fixing the latter two implies a unique solution to the system of moment conditions (13).

In order to close the identification argument, we show that  $\sigma_l$  and  $c(\cdot)$  uniquely determine a set of work-hour CDFs. To see why this is true, first recall that for a continuous random variable  $U$  with CDF  $G_U(U)$ , if we construct a second random variable  $H = f(U)$  using a monotone decreasing function  $f$ , then the CDF of  $H$  can be derived from the CDF of  $U$  by  $G_H(h) = 1 - G_U(f^{-1}(h))$ . This fact is useful since equation (6) indicates that optimal hours supplied  $H_{it}$  are monotone decreasing in cost shocks  $u_{it}^l$  since  $c(\cdot)$  is strictly convex. On any day when labor supply is positive, we can express the cost shock on that day as  $u_{it}^l = \frac{W_i}{\Theta_{li}\mathcal{T}_l^{X_{1it}}c'(H_{it})}$ , or alternatively, we can express labor supply as a function of the shock by inverting this equation to get  $H_{it} = (c')^{-1} \left[ \frac{W_i}{u_{it}^l \Theta_{li} \mathcal{T}_l^{X_{1it}}} \right]$ . We also know that worker  $i$  will choose positive labor supply only on days when her cost shock does not exceed the cutoff  $\bar{u}_i^l = \frac{W_i}{c'(0)\Theta_{li}}$ . Moreover, the conditional shock CDF, given  $u_{it}^l \leq \bar{u}_i^l$ , is  $\frac{\ln\Phi(U_{it}^l, \sigma_l)}{\ln\Phi(\bar{u}_i^l, \sigma_l)}$ , where  $\ln\Phi(\cdot, \sigma_l)$  is the CDF of the log-normal distribution.

These facts characterize the conditional CDF of positive work times implied by a  $(\sigma_l, c)$  pair. We denote these model-implied CDFs by  $\ddot{G}_H(h; \sigma_l, c)$  to differentiate them from their known empirical analogs  $G_H(h)$ , while noting that under the true  $(\sigma_l, c)$  they are the same. On CSR and Neutral days, for each individual  $i$  laboring under wage  $W_i = w_1$  the conditional CDFs of positive work time are given by

$$\begin{aligned} H_i(U_{it}^l)|_{W_i=w_1, X_1=1} &\sim \ddot{G}_{H_i|S_1, X_1}(h_i|S_1, X_1=1; \sigma_l, c) = 1 - \frac{\ln\Phi\left(\frac{w_1}{\Theta_{li}\mathcal{T}_l c'[h_i]}; \sigma_l\right)}{\ln\Phi(\bar{u}_i^l, \sigma_l)}, \quad \text{and} \\ H_i(U_{it}^l)|_{W_i=w_1, X_1=0} &\sim \ddot{G}_{H_i|S_1, X_1}(h_i|S_1, X_1=0; \sigma_l, c) = 1 - \frac{\ln\Phi\left(\frac{w_1}{\Theta_{li} c'[h_i]}; \sigma_l\right)}{\ln\Phi(\bar{u}_i^l, \sigma_l)}, \end{aligned} \quad (14)$$

respectively. The right-hand sides above truncate the support of the distribution at the cutoff  $\bar{u} = \frac{W_i}{c'(0)\Theta_{li}}$  where the corner solution of  $H_{it} = 0$  obtains. Similarly, we can also characterize  $i$ 's counterfactual conditional work-time distributions, had she been offered wage  $W_i = w_2$  instead, as

$$\begin{aligned} H_i(U_{it}^l)|_{W_i=w_2, X_1=1} &\sim \ddot{G}_{H_i|S_2, X_1}(h_i|S_2, X_1 = 1; \sigma_l, c) = 1 - \frac{\ln\Phi\left(\frac{w_2}{\Theta_{li}c'[h_i]}; \sigma_l\right)}{\ln\Phi(\bar{u}_i^l; \sigma_l)}, \quad \text{and} \\ H_i(U_{it}^l)|_{W_i=w_2, X_1=0} &\sim \ddot{G}_{H_i|S_2, X_1}(h_i|S_2, X_1 = 0; \sigma_l, c) = 1 - \frac{\ln\Phi\left(\frac{w_2}{\Theta_{li}c'[h_i]}; \sigma_l\right)}{\ln\Phi(\bar{u}_i^l; \sigma_l)}, \end{aligned} \quad (15)$$

for CSR and neutral workdays, respectively. Therefore,  $i$ 's actual and counterfactual conditional work-time CDFs can both be expressed as functions of  $\sigma_l$  and  $c(\cdot)$ :

$$\begin{aligned} \ddot{G}_{H_i|S_1}(h_i|S_1; \sigma_l, c) &= \frac{\ddot{G}_{H_i|S_1, X_1}(h_i|S_1, X_1 = 1; \sigma_l, c) + \ddot{G}_{H_i|S_1, X_1}(h_i|S_1, X_1 = 0; \sigma_l, c)}{2} \quad \text{and} \\ \ddot{G}_{H_i|S_2}(h_i|S_2; \sigma_l, c) &= \frac{\ddot{G}_{H_i|S_2, X_1}(h_i|S_2, X_1 = 1; \sigma_l, c) + \ddot{G}_{H_i|S_2, X_1}(h_i|S_2, X_1 = 0; \sigma_l, c)}{2}. \end{aligned} \quad (16)$$

Finally, the aggregate conditional work-time CDFs under the model are similarly expressible as functions of  $\sigma_l$  and  $c(\cdot)$ :

$$\ddot{G}_{H|S_j}(h|S_j; \sigma_l, c) = \frac{\sum_{i=1}^I \ddot{G}_{H_i|S_j}(h|S_j; \sigma_l, c)}{I}; \quad j = 1, 2. \quad (17)$$

Recalling that the empirical distributions of positive work times  $G_{H|S_j}$ ,  $j = 1, 2$  are known, [Torgovitsky \(2015, Theorem 2\)](#) and [D'Haultfoeulle and Février \(2015, Theorem 2\)](#) imply that there is a unique  $(\sigma_l, c(\cdot))$  pair which make equation (17) consistent with them while satisfying moment conditions (13).

### A.3.3 Controlling for Sample Selection on Wage Offer

Recall from Section A.3 that we need an exclusion restriction, that known subsets  $S_1$  and  $S_2$  exist of wage groups  $w_1$  and  $w_2$ , respectively, such that  $G_{l|S_1}(\Theta_{li}|S_1) = G_{l|S_2}(\Theta_{li}|S_2) = G_l^*(\Theta_{li})$ , in order to identify the labor-supply model. However, since workers are allowed to respond to their randomly assigned offer  $W_i \in \{w_1, w_2\}$  by applying for the position or not, it is likely that worker selection on wage offer is present in the data-generating process. Table 2 confirms this: holding recruitment status fixed, a shift from wage offer  $w_1 = \$11$  to  $w_2 = \$15$  increased the number of applications by roughly 32%. Therefore, if we define  $S_j = \{H_{it}|H_{it} > 0 \ \& \ W_i =$

$w_j$ ,  $j = 1, 2$  as the set of strictly positive work times logged under wage  $w_j$ , then the distribution of cost types  $\Theta_l$  that gave rise to  $\mathcal{S}_1$  will not be the same that gave rise to  $\mathcal{S}_2$ . This is because the latter set is influenced by an additional margin of workers who were unwilling to supply labor to the firm at a price of  $w_1$  but who were willing at a price of  $w_2$ . However, since we randomly assign wage offers, and since application rates by recruitment cells are known, a correction can be made to restore the exclusion restriction.

Consider a broader model of labor search where our wage offer represents one draw from an offer distribution from which each worker may take several draws over time by waiting longer, rather than accepting a given offer at one point in time. Suppose each worker uses a cutoff rule, say  $\zeta(\Theta_{li})$  that is strictly increasing in the worker's type  $\Theta_{li}$ , such that the worker accepts a job with wage offer  $w$  if and only if  $w \geq \zeta(\Theta_{li})$ .<sup>24</sup> A cutoff rule implies simplistic composition effects: shifts in  $w$  induce different truncation in the tails of the distribution of workers who accept.

More concretely, monotonicity in the cutoff rule implies that if  $\bar{\Theta}_1(X_0)$  is the maximal cost type who applies under wage  $w_1$  then, holding all else fixed (including CSR/neutral recruitment status), the maximum cost type under  $w_2 > w_1$  satisfies  $\bar{\Theta}_2(X_0) > \bar{\Theta}_1(X_0)$ . Moreover, the additional applicants entering the pool under  $w_2$  all have cost types above  $\bar{\Theta}_1(X_0)$ , with the conditional distribution of applicants, given  $\Theta \leq \bar{\Theta}_1(X_0)$  being the same as it was under  $w_1$ . Within the worker choice model, conditional on being hired, mean within-person labor supply,  $E[H_{it}]$  is also monotone decreasing in  $\Theta_{li}$ . Combining these two ideas implies a simple sample trimming rule.

We first define  $\mathcal{S}_1 = \{H_{it} | H_{it} > 0 \ \& \ W_i = w_1\}$  as the set of all positive work times by hired workers assigned to the low-wage group. Next, let  $\mu_{kj}$ ,  $k = C, N$ ,  $j = 1, 2$ , denote the fraction of all potential applicants who accepted wage offer  $(\mathbf{Z}, X_0 = \mathbb{1}(k = C), w_j)$ , and let  $G_{\bar{H}^{kj}}(h)$  denote the CDF of within-worker mean work times across all  $T$  periods. Note that these are both known objects since they can be

---

<sup>24</sup>Such phenomena are common in labor search models (see [Ljungqvist and Sargent \(2004\)](#)).



derived from observables. With that, we can define

$$\begin{aligned}
\tilde{\mathcal{S}}_{N2} &\equiv \left\{ H_{it} > 0 : i \in \left\{ n : W_n = w_2, X_{0n} = 0, E[H_{nt}|X_{0n} = 0, W_n = w_2] > G_{\bar{H}^{N2}}^{-1}(e_N) \right\} \right\}, \\
&\quad \text{s.t. } e_N \equiv \frac{\mu_{N2} - \mu_{N1}}{\mu_{N1}}, \quad \text{and} \\
\tilde{\mathcal{S}}_{C2} &\equiv \left\{ H_{it} > 0 : i \in \left\{ n : W_n = \$15, X_{0n} = 1, E[H_{nt}|X_{0n} = 1, W_n = w_2] > G_{\bar{H}^{C2}}^{-1}(e_C) \right\} \right\}, \\
&\quad \text{s.t. } e_C \equiv \frac{\mu_{C2} - \mu_{C1}}{\mu_{C1}}.
\end{aligned} \tag{18}$$

Note that these two sets exclude the additional influx of marginal, high-cost applicants who enter the worker pool when the wage changes from  $w_1$  to  $w_2$ . This is because, for the neutrally recruited group,  $e_N = \frac{\mu_{N2} - \mu_{N1}}{\mu_{N1}}$  represents the additional fraction of high-cost types who entered the worker pool under the high wage offer ( $w_2 = \$15$ ), relative to the lower wage offer ( $w_1 = \$11$ ). Thus, we find the cutoff at the  $e_N^{th}$  percentile of the distribution of  $\bar{H}^{N2}$  and trim high-wage, neutral recruits with mean labor supply below that cutoff. We then do a similar operation for high-wage CSR recruits.

With that completed, if we define  $\mathcal{S}_2 = \tilde{\mathcal{S}}_{C2} \cup \tilde{\mathcal{S}}_{N2}$ , then  $\mathcal{S}_1$  and  $\mathcal{S}_2$  qualify as subsets of workers which satisfy the exclusion restriction in Assumption 6. Moreover, with that the common cost function  $c$  and shock variance  $\sigma_l$  can be identified using the arguments from Section A.3.2. Further, once the common cost function  $c$  is known, idiosyncratic types  $\Theta_l$  for *all* test subjects can be identified, including those who were excluded by the trimming rule, through the system of moment conditions in (13).

## A.4 GMM Estimation

Here we present the full technical details of our structural model estimator. In doing so we lay out an extended methodology which allows for estimation of treatment effects that are heterogeneous by worker observables. This is to demonstrate the level of generality our framework can accommodate for researchers studying non-pecuniary job traits in broad settings where heterogeneous treatments are more central than in our empirical application. To fix ideas, suppose workers differ by a binary observable characteristic, such as gender.<sup>25</sup> We chose gender as an illustrative example partly because it was one of few demographic variables that were easily verifiable, and partly

<sup>25</sup>The interested reader is directed to Supplemental Online Appendix C where we present model treatment effects allowing for heterogeneity by gender. There our overall results are similar, but we

because several existing studies on valuations for non-pecuniary job characteristics focus on schedule flexibility, for which it is reasonable to assume that preferences may vary by gender (e.g., single mothers may place a relatively high value on flexibility). However, there is nothing particularly crucial about gender in a methodological sense: a similar approach could be applied to any categorical worker observables.

Let  $F_i$  denote a dummy variable equalling 1 if worker  $i$  is female, and 0 otherwise. For active productivity, passive productivity, and leisure preference—the components of the model where treatment effects enter multiplicatively—we re-define gender-dependent treatment effects as  $\mathcal{T}_{ji} \equiv \beta_j \times \beta_{jf}^{F_i}$ ,  $j = p, d, l$ . Under this setup, for each  $j = p, d, l$ ,  $\beta_j$  represents the baseline treatment effect on males for whom  $F_i = 0$ . For females, the treatment effect is  $\beta_j \times \beta_{jf}$ , where the latter term in the product is a female treatment differential. For accuracy, we analogously re-define gender-dependent treatment effects as  $\mathcal{T}_{ai} = \beta_a + F_i\beta_{af}$ . Similarly as above,  $\beta_a$  represents the baseline treatment effect on males, while for females it is  $\beta_a + \beta_{af}$ , with the latter term being interpreted as a female treatment differential.

## A.5 Productivity Estimator

If we consider two subsequent units of output  $q_i \geq 2$  and  $q_i - 1$ , we can use equation (8) to form the log difference in time required to produce each of them:

$$\ln\left(\frac{\tau_{q_i}}{\tau_{q_i-1}}\right) = -\delta \ln\left(\frac{q_i}{q_i-1}\right) + (\ln(\beta_p) + \ln(\beta_{pf})F_i)(X_{1q_i} - X_{1(q_i-1)}) + (\varepsilon_{q_i}^p - \varepsilon_{q_i-1}^p). \quad (19)$$

Recall that treatment status is held constant within each period  $t$ , so for the majority of outputs produced by worker  $i$  we will have  $X_{1q_i} = X_{1(q_i-1)}$  and the second term on the right-hand side will drop out. With that in mind, equation (19) leads to a straightforward estimator for  $\delta$ :

$$\hat{\delta} = -\frac{\sum_{i=1}^I \sum_{q_i=1}^{Q_i} \left[ \ln\left(\frac{\tau_{q_i}}{\tau_{q_i-1}}\right) / \ln\left(\frac{q_i}{q_i-1}\right) \right] \mathbb{1}(X_{1q_i} = X_{1(q_i-1)} \ \& \ q_i \geq 3)}{\sum_{i=1}^I \sum_{q_i=1}^{Q_i} \mathbb{1}(X_{1q_i} = X_{1(q_i-1)} \ \& \ q_i \geq 3)}. \quad (20)$$

find that the productivity gains (in both the active and passive senses) from work-stage treatments are higher for male workers, who are on average less productive than their female counterparts at baseline. One possible explanation for this is that male workers operate further from their production possibility frontier at baseline. Productivity treatment effects are positive and statistically significant for both genders, but under treatment the gender productivity gap is no longer statistically significant.

For treatment effects, our approach is also intuitive, though notationally intense: with our estimate  $\hat{\delta}$  in hand, if we take an average of the quantity  $\ln(\tau_{q_i}) + \hat{\delta} \ln(q_i)$  within CSR-treated days and neutral days separately for each  $i$ , then equation (8) and Assumption 3 imply that the difference between the control mean and the treatment mean will reflect the treatment effect. To implement this idea, we first define  $Q_{Ci} \equiv \sum_{q_i=1}^{Q_i} \mathbf{1}(X_{1q_i} = 1)$  and  $Q_{Ni} \equiv \sum_{q_i=1}^{Q_i} \mathbf{1}(X_{1q_i} = 0)$  as the number of units of output produced on CSR-treated days and Neutral days, respectively. We also define mean shocks to production time as

$$\bar{\varepsilon}_{Ci}^p \equiv \begin{cases} \frac{\sum_{q_i=1}^{Q_i} \varepsilon_{q_i}^p \mathbf{1}(X_{1q_i}=1)}{Q_{Ci}} & \text{if } Q_{Ci} > 0 \\ 0 & \text{otherwise,} \end{cases} \quad \text{and} \quad \bar{\varepsilon}_{Ni}^p \equiv \begin{cases} \frac{\sum_{q_i=1}^{Q_i} \varepsilon_{q_i}^p \mathbf{1}(X_{1q_i}=0)}{Q_{Ni}} & \text{if } Q_{Ni} > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Whenever  $Q_{Ni} > 0$  and  $Q_{Ci} > 0$  for the same individual  $i$ , we have  $\bar{Y}_{Ci}^p - \bar{Y}_{Ni}^p = \ln(\beta_p) + \ln(\beta_{pf})F_i + (\bar{\varepsilon}_{Ci}^p - \bar{\varepsilon}_{Ni}^p)$ , where

$$\bar{Y}_{ji}^p \equiv \begin{cases} \frac{\sum_{q_i=1}^{Q_i} [\ln(\tau_{q_i}) + \hat{\delta} \ln(q_i)] \mathbf{1}[X_{1q_i} = \mathbf{1}(j=C)]}{Q_{ji}} & \text{if } Q_{ji} > 0 \\ 0 & \text{otherwise,} \end{cases} \quad j = C, N$$

represents mean log production times (net of learning) for  $i$  under treatment condition  $j = C, N$ . This sets up a straightforward least squares estimator for treatment effects:

$$\begin{bmatrix} \hat{\beta}_p \\ \hat{\beta}_{pf} \end{bmatrix} = \exp \left[ \underset{[\ln(\beta_p), \ln(\beta_{pf})]^\top}{\text{arg min}} \left\{ \sum_{i=1}^I \omega_i^p (\bar{Y}_{Ci}^p - \bar{Y}_{Ni}^p - \ln(\beta_p) - \ln(\beta_{pf})F_i)^2 \right\} \right], \quad (21)$$

where we use the standard inverse variance sampling weights given by<sup>26</sup>

$$\omega_i^p \equiv \begin{cases} \frac{1}{\widehat{\text{var}}(\bar{Y}_{Ci}^p - \bar{Y}_{Ni}^p)} & \text{if } Q_{Ci} > 0 \ \& \ Q_{Ni} > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Next we estimate the baseline initial production time  $\tau_1$  and fixed effects  $\{\Theta_{pi}\}_{i=1}^I$ . If we let  $Z_{qi}^p \equiv \ln(\tau_{q_i}) + \hat{\delta} \ln(q_i) \mathbf{1}(q_i \geq 2) - X_{1q_i} [\ln(\hat{\beta}_p) + \ln(\hat{\beta}_{pf})F_i]$ , then since  $E[\varepsilon_{q_i}^p] = 0$  we have :

$$\hat{\tau}_1 = \exp \left[ \frac{\sum_{i=1}^I \sum_{q_i=1}^{Q_i} Z_{qi}^p}{\sum_{i=1}^I Q_i} \right] \quad (22)$$

and

$$\hat{\Theta}_{pi} = \exp \left[ \frac{\sum_{q_i=1}^{Q_i} (Z_{qi}^p - \ln(\hat{\tau}_1))}{Q_i} \right], \quad i = 1, \dots, I. \quad (23)$$

---

<sup>26</sup>We use the usual pooled variance formula to calculate  $\widehat{\text{Var}}(\bar{Y}_{Ci}^p - \bar{Y}_{Ni}^p)$  in the sampling weights.

## A.6 Accuracy Estimator

As in the case of productivity, we will construct a differencing estimator for accuracy. In this case, however, we need to make some adjustments to cope with non-linearity in parameters introduced by the CDF  $\Phi(\cdot)$ . We begin with a difference equation for the treatment effects  $\beta_a$  and  $\beta_{af}$ . We then back out individual fixed effects  $\Theta_{ai}$ , given known values for treatments.

Note that for all of  $i$ 's units produced under the same task framing, the treatment effect is constant. Let  $i$ 's within-treatment empirical accuracy rate be defined as

$$\bar{Y}_{ji}^a \equiv \begin{cases} \frac{\sum_{q_i=1}^{Q_i} \mathbb{1}[A_{q_i}=1 \ \& \ X_{1q_i}=\mathbb{1}(j=C)]}{Q_{ji}} & \text{if } \sum_{q_i=1}^{Q_i} \mathbb{1}[X_{1q_i} = \mathbb{1}(j = C)] > 0 \\ 0 & \text{otherwise.} \end{cases} \quad j = C, N,$$

Whenever both  $Q_{Ni} > 0$  and  $Q_{Ci} > 0$  for the same individual  $i$  we have  $\Phi^{-1}(\Pr[A_{q_i} = 1|X_{1q_i} = 1]) = \Theta_{ai} + \beta_a + \beta_{af}F_i$  and  $\Phi^{-1}(\Pr[A_{q_i} = 1|X_{1q_i} = 0]) = \Theta_{ai}$ . If we substitute the empirical analogs of the left-hand sides of these equations and difference them we obtain

$$\mathbb{E} [\Phi^{-1}(\bar{Y}_{Ci}^a) + BC_{Ci} - \Phi^{-1}(\bar{Y}_{Ni}^a) - BC_{Ni}] = \beta_a + \beta_{af}F_i,$$

where the expectation is taken across individuals  $i$  and the  $BC_{ji}$ ,  $j = C, N$  are known bias correction terms (discussed below) to adjust for finite-sample variability of the empirical accuracy scores  $\bar{Y}_{ji}^a$ . Thus, accuracy treatment effects are estimable via a simple least squares routine:

$$[\hat{\beta}_a, \hat{\beta}_{af}]^\top = \arg \min_{[\beta_a, \beta_{af}]^\top} \left\{ \sum_{i=1}^I \omega_i^a (\Phi^{-1}[\bar{Y}_{Ci}^a] + BC_{Ci} - \Phi^{-1}[\bar{Y}_{Ni}^a] - BC_{Ni} - \beta_a - \beta_{af}F_i)^2 \times \mathbb{1}[(Q_{Ci} > 0) \ \& \ (Q_{Ni} > 0)] \right\}, \quad (24)$$

where the  $\omega_i^a$ 's are inverse variance sampling weights given by<sup>27</sup>

$$\omega_i^a \equiv \begin{cases} \frac{1}{\text{var}[\Phi^{-1}(\bar{Y}_{Ci}^a) - \Phi^{-1}(\bar{Y}_{Ni}^a)]} & \text{if } Q_{Ci} > 0 \ \& \ Q_{Ni} > 0 \\ 0 & \text{otherwise.} \end{cases}$$

---

<sup>27</sup>Here we use the standard pooled variance formula to calculate  $\text{Var}[\Phi^{-1}(\widehat{\bar{Y}_{Ci}^a}) - \Phi^{-1}(\bar{Y}_{Ni}^a)]$  in the sampling weights. Moreover, we calculate the variance of  $\Phi^{-1}(\bar{Y}_{ji}^a)$  using the delta method, the inverse function theorem, and standard properties of the sample mean for Bernoulli random samples:  $\text{Var}[\Phi^{-1}(\bar{Y}_{ji}^a)] \approx \left(\frac{1}{\phi(\bar{Y}_{ji}^a)^2}\right) \left(\frac{\bar{Y}_{ji}^a(1-\bar{Y}_{ji}^a)}{7Q_{ji}}\right)$ ,  $j = C, N$ .

As for fixed effects and selections, let  $\bar{\Omega}_{rji}^{ak}$  denote the empirical mean accuracy score for a person of gender  $k = f, m$ , hiring treatment group  $r = C, N$ , on days where task framing treatment  $j = C, N$  is in effect. Specifically, this is defined as

$$\bar{\Omega}_{rji}^{ak} \equiv \begin{cases} \frac{\sum_{q_i=1}^{Q_i} \mathbb{1} \left[ (A_{q_i}=1) \& (F_i=\mathbb{1}[k=f]) \& (X_{0i}=\mathbb{1}[r=C]) \& (X_{1q_i}=\mathbb{1}[j=C]) \right]}{Q_{rji}^k} & \text{if } Q_{rji}^k > 0 \\ 0 & \text{otherwise} \end{cases} \quad k=f,m; r=C,N; j=C,N,$$

where  $Q_{rji}^k \equiv \sum_{q_i=1}^{Q_i} \mathbb{1} \left[ (F_i = \mathbb{1}[k = f]) \& (X_{0i} = \mathbb{1}[r = C]) \& (X_{1q_i} = \mathbb{1}[j = C]) \right]$  is the total number of units produced by person  $i$ , given his/her gender  $k$  and hiring treatment  $r$ , under task framing treatment  $j$ . By similar logic as above, it follows that we can estimate individual fixed effects  $\Theta_{ai}$  for individual  $i$  with gender  $k$  and hiring treatment  $r$  using a simple weighted average for each  $i = 1, \dots, I$ :

$$\hat{\Theta}_{ai} = \left[ \frac{Q_{rCi}^k}{Q_{rCi}^k + Q_{rNi}^k} \left( \Phi^{-1} \left[ \bar{\Omega}_{rCi}^{ak} \right] + BC_{rCi}^k - \hat{\beta}_a - \hat{\beta}_{af} \mathbb{1}[k = f] \right) + \frac{Q_{rNi}^k}{Q_{rCi}^k + Q_{rNi}^k} \left( \Phi^{-1} \left[ \bar{\Omega}_{rNi}^{ak} \right] + BC_{rNi}^k \right) \right]. \quad (25)$$

### A.6.1 Bias Correction

A problem that remains for the above estimator is the fact that the empirical accuracy rates (the  $\bar{Y}$ 's and  $\bar{\Omega}$ 's) above are subject to finite-sample variability, and  $E[\Phi^{-1}(Z)] \neq \Phi^{-1}(E[Z])$  for any random variable  $Z$ , due to the non-linearity in  $\Phi$ . However, one can use the known functional form of  $\Phi$  and the statistical properties of sample means to correct for the bias. The empirical accuracy rates are sample means from Bernoulli random variables, so the central limit theorem provides us with a sampling distribution theory. Specifically, if  $\{a_n\}_{n=1}^N$  is a random sample of Bernoulli observations with mean  $\rho$ , then the sample mean  $\bar{A} = \sum_{n=1}^N a_n/N$  is (approximately) normal with mean  $\rho$  and standard deviation  $\rho(1 - \rho)/\sqrt{N}$ . Using this information, one can compute the expectation

$$E[\Phi^{-1}(\bar{A})] = \int_0^1 \Phi^{-1}(t; 0, 1) \phi \left( t; \rho, \sqrt{\frac{\rho(1 - \rho)}{N}} \right) dt.$$

This expression allows for the researcher to implement a finite-sample bias correction through substitutions such as  $\Phi^{-1}[\bar{A}] + BC \approx E[\Phi^{-1}(\bar{A})]$  in equations (24) and (25) above.

## A.7 A Two-Stage Labor Supply Estimator

As a dimension-reduction measure, we begin by adopting flexible but finite-dimensional parametric forms for several objects in the model. We specify the functional form of the cost function  $c(\cdot)$  and the conditional work time distributions  $G_{H|S_1}$ ,  $G_{H|S_2}$  as B-spline functions. B-splines are convenient tools for parametric curve fitting that combine the most attractive properties of both orthogonal polynomials and piecewise splines. Like polynomials they are computationally convenient, being linear combinations of globally defined basis functions which can be constructed to fit any continuous curve to arbitrary precision. Similar to piecewise splines, they are much better behaved than global polynomials, being locally low-dimensional and facilitating shape restrictions easily through parsimonious linear constraints on their parameters (see [Hickman et al. \(2017\)](#) a brief primer on B-splines and their application to empirical microeconomics).

In order to define a set of B-spline basis functions, one must first specify a fixed partition of the relevant (compact) empirical domain  $[\underline{h}, \bar{h}] = [\min_{i,t}(h_{it}), \max_{i,t}(h_{it})]$ . The set of points which establish cutoffs for the subintervals in the partition is referred to as a *knot vector*. For the cost function, we choose a knot vector to partition the support of work times into  $K_c$  subintervals, denoted  $\mathbf{k}_c = \{k_{c1} = \underline{h} < k_{c2} < \dots < k_{c,K_c+1} = \bar{h}\}$  and for the conditional work time CDFs we choose a partition of  $K_h$  subintervals, denoted  $\mathbf{k}_h = \{k_{h1} = \underline{h} < k_{h2} < \dots < k_{h,K_h+1} = \bar{h}\}$ . These in combination with the Cox-de Boor recursion formula (see [de Boor \(2001\)](#)) uniquely pin down a set of globally-defined basis functions  $\mathcal{B}_{jk} : [\underline{h}, \bar{h}] \rightarrow \mathbb{R}$ ,  $k = 1, \dots, K_j + 3$ ,  $j = c, h$ , which are twice continuously differentiable and are mathematically equivalent to cubic piecewise splines with continuity and differentiability conditions imposed at the knots.<sup>28</sup> The parameterized B-spline functions themselves exist as linear combinations of their respective bases:  $\widehat{G}_{H|S_1}(h; \boldsymbol{\pi}_1^h) \equiv \sum_{k=1}^{K_h+3} \pi_{1k}^h \mathcal{B}_{1k}^h(h)$ ;  $\widehat{G}_{H|S_2}(h; \boldsymbol{\pi}_2^h) \equiv \sum_{k=1}^{K_h+3} \pi_{2k}^h \mathcal{B}_{2k}^h(h)$ ;  $\widehat{c}(h; \boldsymbol{\pi}^c) \equiv \sum_{k=1}^{K_c+3} \pi_k^c \mathcal{B}_k^c(h)$ .

In order for the first two to constitute valid CDFs we need to impose a number of requisite shape restrictions such as monotonicity, boundary conditions, and stochastic dominance, and for the last one to be a valid cost function we need to

---

<sup>28</sup>Each cubic basis function is nonzero on at most 4 subintervals of the partition, though some basis functions are nonzero on fewer than four. Conversely, for each subinterval in the partition there are exactly 4 out of  $K_j + 3$  basis functions that are nonzero and therefore wield a direct influence over functional fit on that subinterval.

impose monotonicity, convexity, and a boundary condition. One of the virtues of B-splines as a parametric form is the ease with which such information can be readily incorporated into a remarkably flexible functional form. By construction, a B-spline function is monotone if and only if its parameters are ordered monotonically. Thus, the monotonicity conditions are equivalent to imposing the following linear constraints on the parameters:  $\pi_{jk}^h \leq \pi_{j(k+1)}^h$ ,  $\forall k = 1, 2, \dots, K_h + 2$ ,  $j = 1, 2$  and  $\pi_k^c \leq \pi_{(k+1)}^c$ ,  $\forall k = 1, 2, \dots, K_c + 2$ . Similarly, for B-spline functions with a common knot vector (and therefore common basis functions as well), ordering of the functions themselves is equivalent to a simple ordering of their parameters; that is, stochastic dominance is equivalent to  $\pi_{2k}^h \leq \pi_{1k}^h$ ,  $\forall k = 1, 2, \dots, K_h + 3$ . Boundary conditions are also a simple matter, involving equality constraints on the first and last parameter values:  $\pi_{11}^h = \pi_{21}^h = 0$ ,  $\pi_{1, K_h+3}^h = \pi_{2, K_h+3}^h = 1$ , and  $\pi_1^c = 0$ . The remaining boundary derivative and convexity conditions are also imposed on the functional form via an additional set of parsimonious linear constraints.

With that, our estimator follows a two-stage process which closely tracks the logic of the identification argument. In the first stage, we smooth the empirical conditional work time CDFs with our B-spline forms. In the second stage, we choose the parameters of the cost function and shock variance to minimize the distance between our empirical CDFs with their model-generated analogs, and match the moments of the shock order statistics.

### A.7.1 Stage I Estimation

we begin Stage I by collecting all positive work time observations for contract  $w_1$  and  $w_2$  into subsamples  $\mathcal{S}_1 = \{H_{it} : H_{it} > 0, W_i = w_1\}$  and, shifting notation from the previous subsection somewhat, re-define  $\mathcal{S}_2 = \widehat{\mathcal{S}}_{N2} \cup \widehat{\mathcal{S}}_{C2}$ , where

$$\begin{aligned} \widehat{\mathcal{S}}_{N2} &\equiv \left\{ H_{it} : H_{it} > 0, i \in \left\{ n : W_n = w_2, X_{0n} = 0, \overline{H}_n^{N2} > \widehat{G}_{\overline{H}^{N2}}^{-1} \left( \frac{\hat{\mu}_{N2} - \hat{\mu}_{N1}}{\hat{\mu}_{N1}} \right) \right\} \right\} \\ \widehat{\mathcal{S}}_{C2} &\equiv \left\{ H_{it} : H_{it} > 0, i \in \left\{ n : W_n = w_2, X_{0n} = 1, \overline{H}_n^{C2} > \widehat{G}_{\overline{H}^{C2}}^{-1} \left( \frac{\hat{\mu}_{C2} - \hat{\mu}_{C1}}{\hat{\mu}_{C1}} \right) \right\} \right\}, \end{aligned} \quad (26)$$

are the empirical selection-corrected samples of positive work times under wage  $w_2$ . In the above expression, for subjects assigned to recruitment group  $j = C, N$ , and wage offer group  $k = 1, 2$  (henceforth, “group  $jk$ ”,  $\hat{\mu}_{jk}$  is the empirical fraction of all potential applicants who chose to formally apply for the position,  $\overline{H}_i^{jk} \equiv \sum_{t=1}^T H_{it}/T$  is the within-person sample mean of work times for hired workers from group  $jk$  ( for

completeness,  $H_i^{jk} \equiv 0$  if  $i$  was not assigned to group  $jk$ ), and

$$\hat{G}_{\bar{H}^{jk}}(x) = \sum_{i=1}^I \frac{\mathbb{1}(\bar{H}_i^{jk} \leq x)}{\sum_{i=1}^I \mathbb{1}(X_{0i} = \mathbb{1}(j = C) \& W_i = w_k)}$$

is the empirical CDF of mean work times for group  $jk$ . For notational simplicity, let  $S_1$  and  $S_2$  denote the sample sizes of the sets  $\mathcal{S}_1 = \{h_{11}, h_{12}, \dots, h_{1S_1}\}$  and  $\mathcal{S}_2 = \{h_{21}, h_{22}, \dots, h_{2S_2}\}$ , respectively.

Intuitively, from the perspective of stage one estimation, we view each positive work time as if it were an independent observation of a single workday for a distinct individual with type  $\tilde{\Theta}_{js} = \Theta_{li} \mathcal{T}_l^{X_{lit}} u_{it}^l$  for some  $i, t$  and contract  $j$ . We then choose the B-spline parameters to provide a constrained, least squares, best fit to the empirical CDFs of  $\mathcal{S}_1$  and  $\mathcal{S}_2$ ,  $\hat{G}_{S_j}(x) = \sum_{s=1}^{S_j} \mathbb{1}(h_{js} \leq x) / S_j$ .<sup>29</sup> This idea implies the following estimator for the B-spline CDF parameters:

$$\begin{aligned} \begin{bmatrix} \hat{\pi}_1^h \\ \hat{\pi}_2^h \end{bmatrix} &\equiv \arg \min \left\{ \sum_{h_{1s} \in \mathcal{S}_1} \left[ \hat{G}_{H|\mathcal{S}_1}(h_{1s}; \pi_1^h) - \hat{G}_{S_1}(h_{1s}) \right]^2 + \sum_{h_{2s} \in \mathcal{S}_2} \left[ \hat{G}_{H|\mathcal{S}_2}(h_{2s}; \pi_2^h) - \hat{G}_{S_2}(h_{2s}) \right]^2 \right\} \\ &\text{subject to :} \\ &\pi_{jk}^h \leq \pi_{j(k+1)}^h, \quad \forall k = 1, 2, \dots, K_h + 2, \quad j = 1, 2 \quad (\text{monotonicity}) \\ &\pi_{2k}^h \leq \pi_{1k}^h, \quad \forall k = 1, 2, \dots, K_h + 3 \quad (\text{first-order dominance}) \\ &\pi_{11}^h = \pi_{21}^h = 0, \quad (\text{initial conditions}) \quad \text{and} \\ &\pi_{1, K_h+3}^h = \pi_{2, K_h+3}^h = 1 \quad (\text{terminal conditions}). \end{aligned} \tag{27}$$

## A.7.2 Stage II Estimation

With the work time CDF parameters in hand, the second stage estimator uses this information to recover the cost function parameters  $\pi^c$ , the labor-supply shock variance  $\sigma_l$ , the treatment parameters  $\beta_l = [\beta_l, \beta_{fl}]^\top$ , and the worker-specific supply cost types  $\Theta_l = [\Theta_{l1}, \dots, \Theta_{lU}]^\top$ . To do so, we construct what we refer to as a ‘‘counterfactual estimator’’ which derives directly from equations (12) – (17) above.

To begin, let  $\tilde{\Theta}_l = [\Theta_{l1}, \dots, \Theta_{lU-q}]^\top$  denote the subset of test subjects whose work-time data survived the trimming rule in equation (26), and assume without loss of

<sup>29</sup>In stage 2 of estimation, it will be necessary to evaluate the CDF of positive work times at arbitrary domain points, which is why we smooth them using the B-splines.



generality that the final  $n$  test subjects in the full list did not. We denote the model-generated analogs to the empirical work time CDFs by  $\ddot{G}_{H|\mathcal{S}_j}(h; \sigma_l, \boldsymbol{\pi}^c, \boldsymbol{\beta}_l, \Theta_l)$ ,  $j = 1, 2$ , and we construct them through the following process:

For each individual  $i = 1, \dots, I - n$  we can characterize the distribution of her positive work times under wage  $j = 1, 2$  implied by a  $(\sigma_l, \boldsymbol{\pi}^c, \boldsymbol{\beta}_l, \Theta_l)$  quadruple on CSR days and Neutral days as

$$\begin{aligned} \ddot{G}_{H_i|\mathcal{S}_j, X_1}(h_i|\mathcal{S}_j, X_1 = 1; \sigma_l, \boldsymbol{\pi}^c, \boldsymbol{\beta}_l, \Theta_{li}) &= 1 - \frac{\ln\Phi\left(\frac{w_j}{\Theta_{li}\mathcal{T}_l\mathcal{C}'(h_i; \boldsymbol{\pi}^c)}; \sigma_l\right)}{\ln\Phi\left(\frac{w_1}{\Theta_{li}}; \sigma_l\right)}, \quad \text{and} \\ \ddot{G}_{H_i|\mathcal{S}_j, X_1}(h_i|\mathcal{S}_j, X_1 = 0; \sigma_l, \boldsymbol{\pi}^c, \boldsymbol{\beta}_l, \Theta_{li}) &= 1 - \frac{\ln\Phi\left(\frac{w_j}{\Theta_{li}\mathcal{C}'(h_i; \boldsymbol{\pi}^c)}; \sigma_l\right)}{\ln\Phi\left(\frac{w_1}{\Theta_{li}}; \sigma_l\right)}, \end{aligned} \quad (28)$$

respectively. As before,  $\ln\Phi(\cdot; \sigma_l)$  denotes the CDF of the lognormal distribution of labor-supply shocks (with location parameter  $\mu = 0$ ), and note that the right-hand sides of the equations above characterize the distribution of shocks after truncating the support at the cutoff  $\frac{w_j}{\Theta_{li}}$  above which the corner solution of  $H_{it} = 0$  obtains. Note also that the above expressions imply four equations for each worker  $i$ : two which were in effect under the actual wage  $W_i = w_k$  she was offered, and two counterfactual equations which would apply if she had been offered  $w_j \neq w_k$  instead. However, since the difference in her work times depends only on the prevailing wage, both her actual work distribution and the counterfactual distribution can be characterized.

Because of the fact that half of her work days are under treatment condition  $X_{1it} = 1$  and the other half are under  $X_{1it} = 0$  (by design), we can further characterize  $i$ 's model-generated work time CDFs conditioning only on wage  $j = 1, 2$  by

$$\begin{aligned} \ddot{G}_{H_i|\mathcal{S}_j}(h_i|\mathcal{S}_j; \sigma_l, \boldsymbol{\pi}^c, \boldsymbol{\beta}_l, \Theta_{li}) &= \frac{\ddot{G}_{H_i|\mathcal{S}_j, X_1}(h_i|\mathcal{S}_j, X_1 = 1; \sigma_l, \boldsymbol{\pi}^c, \boldsymbol{\beta}_l, \Theta_{li})}{2} \\ &\quad + \frac{\ddot{G}_{H_i|\mathcal{S}_j, X_1}(h_i|\mathcal{S}_j, X_1 = 0; \sigma_l, \boldsymbol{\pi}^c, \boldsymbol{\beta}_l, \Theta_{li})}{2}, \quad j = 1, 2. \end{aligned} \quad (29)$$

Once again, note here that we get two equations for each  $i$ ; one to cover her actual wage treatment, and another to cover the counterfactual scenario where she was offered the other wage instead. Finally, for a given guess of the parameter values  $(\sigma_l, \boldsymbol{\pi}^c)$  we can compute from these equations for each  $H_{it} > 0$  the aggregate model-

generated work time distributions:

$$\begin{aligned}\ddot{G}_{H|\mathcal{S}_1}(h|\mathcal{S}_1; \sigma_l, \boldsymbol{\pi}^c, \boldsymbol{\beta}_l, \boldsymbol{\Theta}_l) &= \frac{\sum_{i=1}^{I-q} \ddot{G}_{H_i|\mathcal{S}_1}(h|\mathcal{S}_1; \sigma_l, \boldsymbol{\pi}^c, \boldsymbol{\beta}_l, \boldsymbol{\Theta}_{li})}{I} \\ \ddot{G}_{H|\mathcal{S}_2}(h|\mathcal{S}_2; \sigma_l, \boldsymbol{\pi}^c, \boldsymbol{\beta}_l, \boldsymbol{\Theta}_l) &= \frac{\sum_{i=1}^{I-q} \ddot{G}_{H_i|\mathcal{S}_2}(h|\mathcal{S}_2; \sigma_l, \boldsymbol{\pi}^c, \boldsymbol{\beta}_l, \boldsymbol{\Theta}_{li})}{I}.\end{aligned}\tag{30}$$

With that definition, the last thing we must do is define a grid of  $L$  domain points,  $\{h_{c1}, h_{c2}, \dots, h_{cL}\}$ , spanning the work time support, at which convexity of the cost function will be enforced. This sets up the second-stage empirical objective function where we select the relevant parameter values to satisfy the empirical analogs of the moment conditions in equation (13) and to optimize fit between the empirical and model-generated work-time CDFs:

$$\begin{aligned}\begin{bmatrix} \hat{\sigma}_l \\ \hat{\boldsymbol{\pi}}^c \\ \hat{\boldsymbol{\beta}}_l \\ \hat{\boldsymbol{\Theta}}_l \end{bmatrix} &\equiv \arg \min_{\substack{(\sigma_l, \boldsymbol{\pi}^c, \boldsymbol{\beta}_l, \boldsymbol{\Theta}_l) \in \\ \mathbb{R}_{++}^{3+L+K_c+3}}} \left\{ \sum_{h_s \in \mathcal{S}_1 \cup \mathcal{S}_2} \left[ \hat{G}_{H|\mathcal{S}_1}(h_s|\mathcal{S}_1; \hat{\boldsymbol{\pi}}_1^h) - \ddot{G}_{H|\mathcal{S}_1}(h_s|\mathcal{S}_1; \sigma_l, \boldsymbol{\pi}^c, \boldsymbol{\beta}_l, \boldsymbol{\Theta}_l) \right]^2 \right. \\ &+ \sum_{h_s \in \mathcal{S}_1 \cup \mathcal{S}_2} \left[ \hat{G}_{H|\mathcal{S}_2}(h_s|\mathcal{S}_2; \hat{\boldsymbol{\pi}}_2^h) - \ddot{G}_{H|\mathcal{S}_2}(h_s|\mathcal{S}_2; \sigma_l, \boldsymbol{\pi}^c, \boldsymbol{\beta}_l, \boldsymbol{\Theta}_l) \right]^2 \\ &+ \sum_{i=1}^I \sum_{j=1}^5 \left( \mathbb{E}_{(j:5)}(\sigma_l) - \ln \left( \frac{W_i}{c'(H_{C_i}(j:5); \boldsymbol{\pi}^c)} \right) + \ln(\boldsymbol{\Theta}_{li}) + \ln(\mathcal{T}_l) \right)^2 \mathbf{1}(H_{C_i}(j:5) > 0) \\ &\left. + \sum_{i=1}^I \sum_{j=1}^5 \left( \mathbb{E}_{(j:5)}(\sigma_l) - \ln \left( \frac{W_i}{c'(H_{N_i}(j:5); \boldsymbol{\pi}^c)} \right) + \ln(\boldsymbol{\Theta}_{li}) \right)^2 \mathbf{1}(H_{N_i}(j:5) > 0) \right\} \\ &\text{subject to :} \\ &\pi_k^c \leq \pi_{(k+1)}^c, \quad \forall k = 1, 2, \dots, K_c + 2, \quad (\text{monotonicity}) \\ &\hat{c}''(h_j; \boldsymbol{\pi}^c) > 0, \quad \forall j = 1, 2, \dots, J \quad (\text{convexity}) \\ &\hat{c}'(0; \boldsymbol{\pi}^c) = 1 \quad (\text{boundary condition}) \\ &\boldsymbol{\Theta}_{li} > 0, \quad i = 1, \dots, I; \quad \beta_l > 0; \quad \beta_{lf} > 0; \quad \sigma_l > 0 \quad (\text{positivity}).\end{aligned}\tag{31}$$

In the objective function, note that only workers who survived the trimming rule (26) enter into the CDF fitting criterion (first two lines), whereas all workers' data contribute to the moment conditions of shock order statistics (third and fourth lines).

Note that implementing the estimator in equation (31) is computationally difficult, as it requires a solver to search through a  $(3 + L + K_c + 3)$ -dimensional space for the

optimal parameter values. However, a mathematically equivalent approach would involve sequentially choosing values for the shock variance  $\sigma_l$  and cost function  $\pi^c$ , and then solving for the treatment parameters  $\beta_l$  and fixed effects  $\Theta_l$  using equations (13) in an inner loop. This sequential approach involves a small increase in computing time per iteration, but comes at the benefit of greater numerical stability, fewer iterations, and less tendency to become stuck at local optima.

### A.7.3 Practical Concerns

In order to complete the definition of the labor supply estimator, some discussion of knot choice is in order. There are three B-spline functions to be estimated: the work hour distributions  $G_{H|S_1}(h; \pi_1^h)$ ,  $G_{H|S_2}(h; \pi_2^h)$ , and the cost function,  $c(h; \pi^c)$ . The former two depend on B-spline basis functions defined by a common knot vector  $\mathbf{k}_h$  and the latter depends on another knot vector  $\mathbf{k}_c$ . Choice of number and locations of knots is crucial for the statistical performance of the estimator.

We chose to locate the knots uniformly in quantile rank space, so as to evenly spread the influence of the data across all parameters to be estimated. That is, if the number of sub-intervals of the domain spanned by the knot vector is  $K$ , then we spaced the knots at the  $100 \times \frac{j}{K+1}$ <sup>th</sup> empirical quantiles, for each  $j = 0, 1, 2, \dots, K + 1$ . For the distributions of work hours, there are long upper tails, so we modify this basic rule by inserting an additional knot at the midpoint between the upper two knots chosen by the above rule.

Once this convention is established, the only remaining choice is how many knots. For the work time CDFs, we chose  $K_h = 5$ —i.e., knots at the empirical quartiles with an extra knot at the midpoint of the upper quartile—and for the cost function we chose  $K_c = 7$  (with locations chosen similarly). This resulted in six free parameters each in the vectors  $\pi_1^h$  and  $\pi_2^h$  (eight parameters total, with two boundary conditions), and nine free parameters in the vector  $\pi^c$  (ten parameters total, with one boundary condition). We settled on these numbers for our empirical implementation because adding further knots did not make any appreciable difference in the model estimates.

## A.8 Type Distributions

To obtain estimates of the distributions of worker characteristics  $G_p, G_d, G_a, G_l$  the main difficulty is that the data  $\{\widehat{\Theta}_{pi}, \widehat{\Theta}_{di}, \widehat{\Theta}_{ai}, \widehat{\Theta}_{li}\}_{i=1}^I$  are estimates derived from an unbalanced panel, so some fixed-effect estimates are more informative than others. Thus, we construct weighted empirical CDFs  $\widehat{G}_j(x) = \sum_{i=1}^I \eta_i^j \mathbf{1}(\widehat{\Theta}_{ji} \leq x)$ ,  $j = p, d, a, l$

using sample weights

$$\eta_i^p = \eta_i^d = \eta_i^a \equiv \frac{Q_{Ci} + Q_{Ni}}{\sum_{j=1}^I Q_{Cj} + Q_{Nj}}, \text{ and } \eta_i^l \equiv \frac{\sum_{t=1}^T \mathbf{1}(H_{it} > 0)}{\sum_{j=1}^I \sum_{t=1}^T \mathbf{1}(H_{jt} > 0)}. \quad (32)$$

Before moving on, one caveat of the type distribution estimators is worth mentioning. Because they are based on a sample of stochastic observations—that is, the available data for estimating the type distributions are *estimates* of worker characteristics, rather than the actual values of the workers’ characteristics—we effectively have a measurement error problem that will induce attenuation bias in some estimates above. The sample weights used above help to mitigate the problem by relying most on the more precisely estimated datapoints, but it cannot solve the problem entirely. In particular, the estimated CDFs will tend toward over-estimating variance, and mean differences (across subsamples), covariances, and correlations will be biased toward zero in finite samples. Thus, the point estimates we derive on mean differences and correlations from the above method can be thought of as conservative lower bounds.

## A.9 Asymptotic Theory and Inference

Our estimator falls within the class of GMM estimators, since each parameter is chosen to match empirical moments from the data. For the productivity and accuracy models, standard econometric theory establishes asymptotic normality of un-balanced panel data estimators (see [Wooldridge 2001](#)). For the labor-supply model, two alternative views are possible. One can consider the B-spline CDFs and cost function to be fixed parametric forms that will not be altered—i.e., the knot vectors and basis functions to be held fixed—as the sample size grows. Under this view, standard asymptotic theory once again applies. Alternatively, one could think of them as *sieve estimators*, meaning that as the sample size grows, the researcher will gradually add additional knots until in the limit the knot vectors will become a dense set on the relevant domains and the B-spline functions will therefore become arbitrarily flexible. This view complicates matters, but a recent econometric literature has established some results on point-wise asymptotic normality of sieve functionals (e.g., [Chen \(2007\)](#) and [Huang \(2003\)](#)), though this requires regularity conditions that are difficult to verify.

The main caveat to recognize, given the unbalanced panel structure of the data, is which dimension governs the asymptotics of a given parameter estimate. Parameter

estimates that are common to all subjects—treatment effects,  $\delta$ ,  $\tau_1$ , and  $c(\cdot)$ —will be consistent at a rate of either  $\sqrt{\sum_{i=1}^I Q_i}$  or  $\sqrt{T \times \bar{T}}$ , being the length times the width of the panel. Thus, the researcher may obtain higher precision in the corresponding estimates simply by increasing the number of test subjects  $I$ , while holding fixed other aspects of the research design. On the other hand, worker fixed-effect estimates—e.g.,  $\Theta_{pi}$ ,  $\Theta_{ai}$ , and  $\Theta_{li}$ —are consistent at a rate of either  $\sqrt{Q_i}$  or  $\sqrt{T}$ .

### A.9.1 Bootstrap Inference

Due to the computational complexity of our estimator, we chose to compute standard errors and confidence bounds by the bootstrap method. We executed a double bootstrap routine (due to the panel structure of our data) that mimicked our experimental sampling scheme and allowed for the possibility of accuracy and productivity shocks being correlated. Specifically, to construct each bootstrap sample, we first partitioned test subjects into recruitment-stage treatment bins. We then re-sampled subject identities from each bin, with replacement, to obtain a sub-sample of workers the same size as the original bin. Then, for each sampled individual, we re-sampled five times, with replacement, from that individual’s work-time sample on CSR days; and again five times, with replacement, from that individual’s work time distribution on neutral days to construct a panel of labor supply data. Then, for each sampled  $i$  we re-sampled, with replacement,  $Q_i$  times from  $i$ ’s empirical distribution of  $(\varepsilon_{qi}^p, \varepsilon_{qi}^d, A_{qi})$ -triples to construct a sample of output data.

We executed this re-sampling scheme for  $S = 100,000$  iterations, and for the  $s^{\text{th}}$  bootstrap sample all model parameters  $\left\{ G_{pdal,s}^*(\cdot, \cdot, \cdot), \mathcal{T}_s^*, \delta_s^*, \tau_{1s}^*, D_{0s}^*, \sigma_{ls}^*, c_s^*(\cdot) \right\}_{s=1}^S$  were estimated. We then computed standard errors and bias-corrected confidence intervals and p-values in the usual way using our bootstrapped estimates.

### A.9.2 Inference on Stochastic Dominance of Worker Characteristics

There are several papers that propose tests for stochastic dominance relationships between two distributions. The earliest work in this vein is [Kolmogorov 1933](#) and [Smirnov 1939](#), but more recent work has developed various alternatives or improvements (e.g., [Davidson and Duclos 2000](#), [Barrett and Donald 2003](#), and [Linton et al. 2010](#); see [Heathcote et al. 2010](#) for a survey). For some brief intuition on how these methods work, suppose there are two random variables  $X \sim F_x$  and  $Y \sim F_y$  defined on a common support  $[\underline{z}, \bar{z}]$ , and two corresponding random samples,  $\{x_n\}_{n=1}^{N_x}$  and  $\{y_n\}_{n=1}^{N_y}$ . When testing for first-order stochastic dominance one must account for the

fact that if the null hypothesis of distributional equality  $F_x = F_y$  is rejected, then there are three alternative possibilities: (i)  $F_x < F_y$ , (ii)  $F_x > F_y$ , and (iii)  $\exists \{z, z'\} \in (z, \bar{z})$  such that  $F_x(z) < F_y(z)$  and  $F_x(z') > F_y(z')$ . Thus, the KS test for dominance proceeds by estimating the standard empirical CDFs  $\widehat{F}_x(x)$  and  $\widehat{F}_y(y)$ , and then by simultaneously constructing two symmetrically defined test statistics based on the vertical differences  $T^{(i)} = \sqrt{\frac{N_x N_y}{N_x + N_y}} \sup_{z \in [z, \bar{z}]} [\widehat{F}_x(z) - \widehat{F}_y(z)]$  and  $T^{(ii)} = \sqrt{\frac{N_x N_y}{N_x + N_y}} \sup_{z \in [z, \bar{z}]} [\widehat{F}_y(z) - \widehat{F}_x(z)]$ . These are then evaluated at their limiting distribution (due to [Doob 1949](#)) to get p-values  $[p^{(i)}, p^{(ii)}] = \exp(-2(T^{(i)}, T^{(ii)})^2)$ . Finally, for some fixed significance level  $\alpha \in (0, 1)$ , if  $\min(p^{(i)}, p^{(ii)}) > \alpha$  then we fail to reject equality; if  $\max(p^{(i)}, p^{(ii)}) < \alpha$  we reject equality in favor of alternative (iii); and if  $p^{(j)} < \alpha$  is true for exactly one  $j \in \{i, ii\}$ , then we reject equality in favor of alternative (j). Different versions of the stochastic dominance test generally follow a similar procedure but vary by how the test statistics and p-values are computed.

A drawback to these methods in our context is that they all assume non-stochastic observations; i.e., that the samples  $\{x_n\}_{n=1}^{N_x}$  and  $\{y_n\}_{n=1}^{N_y}$  are direct observations from  $F_x$  and  $F_y$ , respectively. In our case the data are not direct observations from the distributions of worker characteristics  $(\Theta_p, \Theta_d, \Theta_a, \Theta_l) \sim G_{pdal}(\cdot, \cdot, \cdot, \cdot)$ , but rather, they are *estimates* of worker characteristics:  $\{\widehat{\Theta}_{pi}, \widehat{\Theta}_{di}, \widehat{\Theta}_{ai}, \widehat{\Theta}_{li}\}_{i=1}^I$ . Thus, the testing procedures above produce test statistics and p-values that would under-reject the null hypothesis of equality. Intuitively, this is because the finite-sample variability in worker type estimates creates a measurement error problem that attenuates differences between the estimated distributions. With that in mind, we use a bootstrapped adaptation of the KS test for stochastic dominance that closely follows [Marmer et al. \(2017\)](#).

For the CSR vs neutral comparison, we define a grid of points in quantile rank space by  $\mathbf{r} = \{r_1, r_2, \dots, r_K\} = \{0.010, 0.011, \dots, 0.990\}$  and  $\boldsymbol{\theta}_j = \{\theta_{j1}, \dots, \theta_{jK}\} = \widehat{G}_j^{-1}(\mathbf{r})$ ,  $j = p, a, l$ . We then evaluate the bootstrapped empirical CDFs  $\widehat{G}_{j_s}^*(\theta_{jk}|X_0=0)$  and  $\widehat{G}_{j_s}^*(\theta_{jk}|X_0=1)$   $j = p, d, a, l$ ,  $k = 1, \dots, K$  (conditional on recruitment status) at each of the points in these grids. Similarly as in previous methods, for each  $j = p, d, a, l$  we specify a null hypothesis of equality  $H_0 : G_j(\cdot|X_0=0) = G_j(\cdot|X_0=1)$  with three alternatives  $H_1 : G_j(\cdot|X_0=0) > G_j(\cdot|X_0=1)$ ;  $H_2 : G_j(\cdot|X_0=0) < G_j(\cdot|X_0=1)$ ; and  $H_3 : G_j(\cdot|X_0=0) \ll G_j(\cdot|X_0=1)$ . Finally, we compute a p-value for  $H_1$  by  $p_{1j}^* = 1 - \max_{k=1, \dots, K} \frac{\sum_{s=1}^S \mathbb{1}[\widehat{G}_{j_s}^*(\theta_{jk}|X_0=0) - \widehat{G}_{j_s}^*(\theta_{jk}|X_0=1) > 0]}{S}$ , and symmetrically for  $H_2$  by

$p_{2j}^* = 1 - \max_{k=1, \dots, K} \frac{\sum_{s=1}^S \mathbb{1}[\widehat{G}_{js}^*(\theta_{jk}|X_0=1) - \widehat{G}_{js}^*(\theta_{jk}|X_0=0) > 0]}{S}$ .<sup>30</sup> In words, the p-value for a given alternative is one minus the maximal point-specific frequency with which the null hypothesis is violated in favor of that alternative. For a fixed significance level  $\alpha \in (0, 0.5)$  we draw conclusions from the test by the following process:

1. If  $\min(p_{1j}^*, p_{2j}^*) > \alpha$  we fail to reject  $H_0$
2. Else, if  $p_{1j}^* < \alpha$ ,  $p_{2j}^* > \alpha$  we reject  $H_0$  in favor of  $H_1$
3. Else, if  $p_{1j}^* > \alpha$ ,  $p_{2j}^* < \alpha$  we reject  $H_0$  in favor of  $H_2$
4. Else, if  $\max(p_{1j}^*, p_{2j}^*) < \alpha$  we reject  $H_0$  in favor of  $H_3$ .

## A.10 Additional Tables and Figures

Figures 9 and 10 display results for estimated CDFs of passive productivity fixed effects  $\Theta_d$ . Figure 11 displays the common cost function and marginal cost function, scaled by the median of time preference types  $\Theta_l$ . Tables 9 and 10 report the results of stochastic dominance tests. In cases where the null of equality is rejected, they also report estimated magnitudes of the mean change in levels and as a fraction of a standard deviation relative to the populations involved in the test. The interested reader is directed to Supplemental Online Appendix C for a more complete picture, where Tables 14 – 21 display point-specific p-values over a fine grid of quantiles.

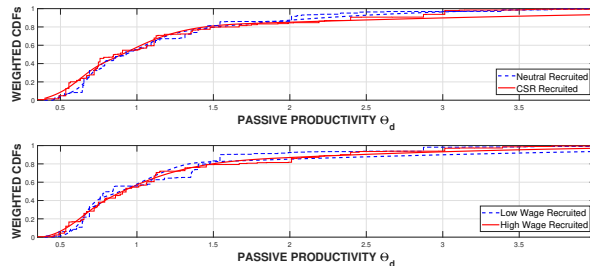


Figure 9: Passive Productivity: Selection on Wage and CSR

<sup>30</sup>We also bias correct our bootstrap testing procedure for each  $\theta$  by re-centering the bootstrapped sample of differences  $\left\{ \left[ \widehat{G}_{js}^*(\theta|X_0=1) - \widehat{G}_{js}^*(\theta|X_0=0) \right] \right\}_{s=1}^S$  at the point estimate difference  $\left[ \widehat{G}_j(\theta|X_0=1) - \widehat{G}_j(\theta|X_0=0) \right]$  in the usual way for each  $\theta \in \text{supp}(\Theta_j)$ .

TEST	HYPOTHESIS	P-VALUE	RESULT	TEST	HYPOTHESIS	P-VALUE	RESULT
$\Theta_p$ Selection on CSR FOSD	$H_0 : G_{p CSR} = G_{p Ntr}$ $H_1 : G_{p CSR} > G_{p Ntr}$ $H_2 : G_{p CSR} < G_{p Ntr}$ $H_3 : G_{p CSR} \neq G_{p Ntr}$ <b>Mean Prod. Time Change:</b> -15.1% (26.1% St. Dev.)	— $< 1 \times 10^{-5}$ 0.1561 —	<i>reject</i> <b>adopt</b> <i>reject</i> <i>reject</i>	$\Theta_d$ Selection on CSR FOSD	$H_0 : G_{d CSR} = G_{d Ntr}$ $H_1 : G_{d CSR} > G_{d Ntr}$ $H_2 : G_{d CSR} < G_{d Ntr}$ $H_3 : G_{d CSR} \neq G_{d Ntr}$ <b>Mean Prod. Time Change:</b> +4.0% (6.4% St. Dev.)	— $< 1 \times 10^{-5}$ 0.0577 —	<i>reject</i> <i>reject</i> <i>reject</i> <b>adopt</b>
$\Theta_p$ Selection on WAGE FOSD	$H_0 : G_{p \$15} = G_{p \$11}$ $H_1 : G_{p \$15} > G_{p \$11}$ $H_2 : G_{p \$15} < G_{p \$11}$ $H_3 : G_{p \$15} \neq G_{p \$11}$ <b>Mean Prod. Time Change:</b> -3.6% (8.7% St. Dev.)	— $< 1 \times 10^{-5}$ 0.1003 —	<i>reject</i> <i>reject</i> <i>reject</i> <b>adopt</b>	$\Theta_d$ Selection on WAGE FOSD	$H_0 : G_{d \$15} = G_{d \$11}$ $H_1 : G_{d \$15} > G_{d \$11}$ $H_2 : G_{d \$15} < G_{d \$11}$ $H_3 : G_{d \$15} \neq G_{d \$11}$ <b>Mean Prod. Time Change:</b> +5.7% (8.5% St. Dev.)	— $< 1 \times 10^{-5}$ 0.0068 —	<i>reject</i> <i>reject</i> <i>reject</i> <b>adopt</b>
<b>COMBINED EFFECTS</b>							
$\Theta_p$ CSR Select. + CSR Trtmt FOSD	$H_0 : G_{\Theta_p, \mathcal{T}_p CSR} = G_{p Ntr}$ $H_1 : G_{\Theta_p, \mathcal{T}_p CSR} > G_{p Ntr}$ $H_2 : G_{\Theta_p, \mathcal{T}_p CSR} < G_{p Ntr}$ $H_3 : G_{\Theta_p, \mathcal{T}_p CSR} \neq G_{p Ntr}$ <b>Mean Prod. Time Change:</b> -38.5% (66.4% St. Dev.)	— $< 1 \times 10^{-5}$ 0.7927 —	<i>reject</i> <b>adopt</b> <i>reject</i> <i>reject</i>	$\Theta_d$ CSR Select. + CSR Trtmt FOSD	$H_0 : G_{\Theta_d, \mathcal{T}_d CSR} = G_{d Ntr}$ $H_1 : G_{\Theta_d, \mathcal{T}_d CSR} < G_{d Ntr}$ $H_2 : G_{\Theta_d, \mathcal{T}_d CSR} > G_{d Ntr}$ $H_3 : G_{\Theta_d, \mathcal{T}_d CSR} \neq G_{d Ntr}$ <b>Mean Prod. Time Change:</b> -48.1% (76.0% St. Dev.)	— $< 1 \times 10^{-5}$ 0.9376 —	<i>reject</i> <b>adopt</b> <i>reject</i> <i>reject</i>
$\Theta_p$ CSR Select. + Wage Select. FOSD	$H_0 : G_{p CSR, \$15} = G_{p Ntr, \$11}$ $H_1 : G_{p CSR, \$15} > G_{p Ntr, \$11}$ $H_2 : G_{p CSR, \$15} < G_{p Ntr, \$11}$ $H_3 : G_{p CSR, \$15} \neq G_{p Ntr, \$11}$ <b>Mean Prod. Time Change:</b> -15.3% (54.8% St. Dev.)	— $< 1 \times 10^{-5}$ 0.0376 —	<i>reject</i> <i>reject</i> <i>reject</i> <b>adopt</b>	$\Theta_d$ CSR Select. + Wage Select. FOSD	$H_0 : G_{d CSR, \$15} = G_{d Ntr, \$11}$ $H_1 : G_{d CSR, \$15} < G_{d Ntr, \$11}$ $H_2 : G_{d CSR, \$15} > G_{d Ntr, \$11}$ $H_3 : G_{d CSR, \$15} \neq G_{d Ntr, \$11}$ <b>Mean Prod. Time Change:</b> +13.3% (21.3% St. Dev.)	— $< 1 \times 10^{-5}$ 0.0068 —	<i>reject</i> <i>reject</i> <i>reject</i> <b>adopt</b>

Table 9: Stochastic Dominance Tests: Passive and Active Productivity



TEST	HYPOTHESIS	P-VALUE	RESULT	TEST	HYPOTHESIS	P-VALUE	RESULT
$\Theta_a$ Selection on CSR FOSD	$H_0 : G_a CSR = G_a Ntr$ $H_1 : G_a CSR > G_a Ntr$ $H_2 : G_a CSR < G_a Ntr$ $H_3 : G_a CSR \neq G_a Ntr$ <b>Mean Accuracy Change: +3.8% (16.5% St. Dev.)</b>	— 0.0087 0.1870 —	<i>reject</i> <b>adopt</b> <i>reject</i> <i>reject</i>	$\Theta_l$ Selection on CSR FOSD	$H_0 : G_l CSR = G_l Ntr$ $H_1 : G_l CSR > G_l Ntr$ $H_2 : G_l CSR < G_l Ntr$ $H_3 : G_l CSR \neq G_l Ntr$ <b>Mean Time Value Change: +2.7% (4.2% St. Dev.)</b>	— $< 1 \times 10^{-5}$ 0.0365 —	<i>reject</i> <i>reject</i> <i>reject</i> <b>adopt</b>
$\Theta_a$ Selection on WAGE FOSD	$H_0 : G_a \$15 = G_a \$11$ $H_1 : G_a \$15 > G_a \$11$ $H_2 : G_a \$15 < G_a \$11$ $H_3 : G_a \$15 \neq G_a \$11$ <b>Mean Accuracy Change: +13.2% (51.8% St. Dev.)</b>	— 0.0035 0.2790 —	<i>reject</i> <b>adopt</b> <i>reject</i> <i>reject</i>	$\Theta_l$ Selection on WAGE FOSD	$H_0 : G_l \$15 = G_l \$11$ $H_1 : G_l \$15 > G_l \$11$ $H_2 : G_l \$15 < G_l \$11$ $H_3 : G_l \$15 \neq G_l \$11$ <b>Mean Time Value Change: +18.1% (33.8% St. Dev.)</b>	— $< 1 \times 10^{-5}$ 0.3458 —	<i>reject</i> <b>adopt</b> <i>reject</i> <i>reject</i>
<b>COMBINED EFFECTS</b>							
$\Theta_a$ CSR Select. + CSR Trtmt FOSD	$H_0 : G_{\Theta_a+T_a} CSR = G_a Ntr$ $H_1 : G_{\Theta_a+T_a} CSR > G_a Ntr$ $H_2 : G_{\Theta_a+T_a} CSR < G_a Ntr$ $H_3 : G_{\Theta_a+T_a} CSR \neq G_a Ntr$ <b>Mean Accuracy Change: +2.8% (12.2% St. Dev.)</b>	— 0.0087 0.1481 —	<i>reject</i> <b>adopt</b> <i>reject</i> <i>reject</i>	$\Theta_l$ CSR Select. + CSR Trtmt FOSD	$H_0 : G_{\Theta_l} CSR = G_l Ntr$ $H_1 : G_{\Theta_l} CSR < G_l Ntr$ $H_2 : G_{\Theta_l} CSR > G_l Ntr$ $H_3 : G_{\Theta_l} CSR \neq G_l Ntr$ <b>Mean Time Value Change: +4.6% (7.2% St. Dev.)</b>	— $< 1 \times 10^{-5}$ 0.0677 —	<i>reject</i> <i>reject</i> <i>reject</i> <b>adopt</b>
$\Theta_a$ CSR Select. + Wage Select. FOSD	$H_0 : G_a CSR,\$15 = G_a Ntr,\$11$ $H_1 : G_a CSR,\$15 > G_a Ntr,\$11$ $H_2 : G_a CSR,\$15 < G_a Ntr,\$11$ $H_3 : G_a CSR,\$15 \neq G_a Ntr,\$11$ <b>Mean Accuracy Change: +20.6% (78.5% St. Dev.)</b>	— $< 1 \times 10^{-5}$ 0.3911 —	<i>reject</i> <b>adopt</b> <i>reject</i> <i>reject</i>	$\Theta_l$ CSR Select. + Wage Select. FOSD	$H_0 : G_l CSR,\$15 = G_l Ntr,\$11$ $H_1 : G_l CSR,\$15 < G_l Ntr,\$11$ $H_2 : G_l CSR,\$15 > G_l Ntr,\$11$ $H_3 : G_l CSR,\$15 \neq G_l Ntr,\$11$ <b>Mean Time Value Change: +29.0% (47.0% St. Dev.)</b>	— $< 1 \times 10^{-5}$ 0.5111 —	<i>reject</i> <b>adopt</b> <i>reject</i> <i>reject</i>

Table 10: Stochastic Dominance Tests: Accuracy and Time Value

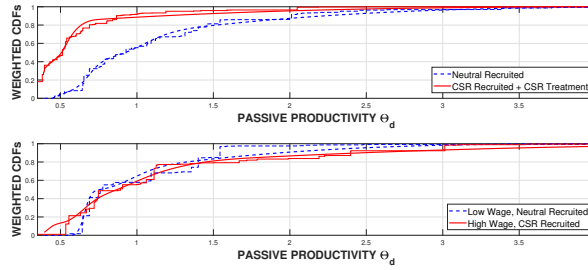


Figure 10: Passive Productivity: Combined Effects

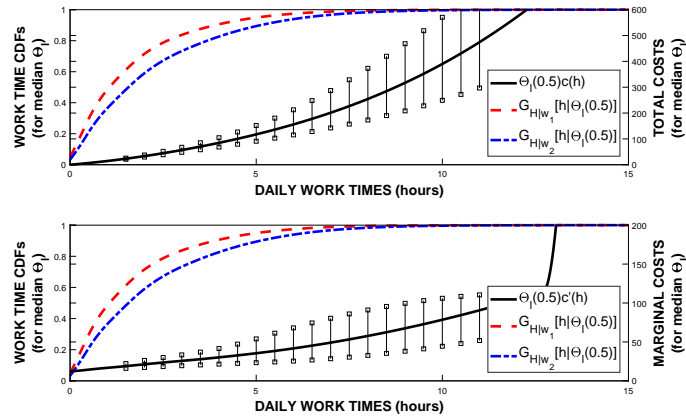


Figure 11: Labor Supply Costs and Marginal Costs (w/90% confidence bounds)