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

Daniel Hedblom[†], Brent R. Hickman[‡], and John A. List[§]

This version: August 4, 2019
Download latest version HERE

Abstract

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.

JEL Classifications: C14, C93, J22, J30, M52

*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.

[†]University of California, Berkeley School of Information.

[‡]**Corresponding author.** Washington U. in St. Louis, Olin Business School. E-mail: hickmanbr@gmail.com.

[§]University of Chicago Economics Department.

1 Introduction

Historians commonly mark the Industrial Revolution as a major economic turning point: incomes and population began to grow more rapidly, and living standards began to consistently climb for the first time in history. The Industrial Revolution also began a landmark change in social responsibility of business organizations. Whereas many scholars argue that corporate social responsibility (CSR) has its roots in the 1950s, the framework began to emerge in the 19th century. As Wren (2005) points out, with the new factory system in Great Britain came social unrest due to concerns surrounding poverty, slums, and child labor. Businesses soon responded with the industrial welfare movement, which included introduction of many worker perks (see pp. 269-270, Wren (2005)). Whether such actions were profit motivated, socially motivated, or a mixture remains debated in the literature.

Among scientists, the virtues of CSR have been debated at least since Friedman (1970) famously described CSR as a "fundamentally subversive doctrine," and declared that the only social responsibility of a firm is to maximize its profits while avoiding deception or fraud. Nevertheless, today CSR is a common business practice, 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. For instance, proponents have argued that CSR investments actually increase profits. There is, however, no consensus in the empirical literature on whether investments in CSR have a positive effect on the firm's bottom line. While some studies report a positive effect (Waddock and Graves (1997)), 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 with few exceptions, 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, one conjecture is that CSR provides non-pecuniary incentives to (i) advantageously select workers to seek employment at the firm, and once employed CSR can (ii) 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, 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 be more likely to accept a given wage offer within this firm than with 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 natural field experiment allows us to separate such treatment and selection effects of CSR.

We begin the analysis 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, DFT15). The identification argument for time preferences consists of using exogenous wage variation to derive an empirical mapping between observable hours worked and underlying agent types. This mapping allows the researcher to reverse-engineer the worker cost function and the distribution of labor supply costs.

We build upon this framework in several important ways. For instance, while DFT15 require exclusion restrictions that imply no selective worker entry across exogenously varying wage contracts, 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. Importantly, allowing for selection on wage contracts expands the empirical analysis in economically interesting ways. We view this advance as a general methodological contribution, highlighting that field experimental variation can both relax identification assumptions and broaden empirical insights gained on economically relevant questions.

To provide the necessary experimental control and variation to estimate the model, we started our own consultancy firm, *HHL Solutions, LLC* and used it as a vehicle for conducting the field experiment. The stated mission of the firm is to provide data collection services for various clients. Every aspect of the firm's origination and organization is authentic. By starting our own firm, we are not only able to expose subjects to controlled variation in wages and information about the firm's investments in CSR, but we also control the hiring process and the work environment. Since our subjects are recruited from an actual marketplace at wage rates common to adult semi-skilled labor ($11/hr to $15/hr), we avoid sample selection problems that can occur when subjects are recruited from a very specific population (Levitt and List 2007).[1]

Our field experiment consists of randomized treatments administered in two stages: (1) a hiring stage and (2) a real effort work task 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. Application rates in each treatment group are the first outcome variable. This allows 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 serves an additional critical role: it provides an instrument for structural identification of labor supply costs and selection effects.

In the second stage of the experiment, subjects who are hired proceed to conduct 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 clicks, timing, and pay in a non-invasive way. This provides us with detailed information about

[1]According to market research firm Payscale.com, at the time the field experiment was executed our wage offers of $11/hr and $15/hr corresponded to the interquartile range for data entry operators in the US. See `http:www.payscale.comresearchUSJob=Data_Entry_OperatorHourly_Rate` (accessed on April 14, 2017).

the performance of each employee and allows us to construct measures of labor supply, productivity, and output quality. In turn, we are able to identify a worker's productivity and work-quality fixed effects using standard panel-data methods. The website also provides a means of varying task framing—by including prompts concerning the firm's CSR activities—across work days so that we can pin down treatment effects. Importantly, both those who are recruited for a CSR job and those who are recruited neutrally receive the *same* regime of task framing prompts during the work stage. 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, using data from more than 1,100 job seekers during our recruitment stage, our reduced-form analysis reveals 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 33 percent, while advertising the firm as a CSR firm increases application rates by nearly the same amount. What remains unclear, however, is the extent to which we are merely observing increased response rates by the same (on average) set of workers, or whether the reduced-form results signal composition effects that alter the mixture of baseline characteristics across workers within the firm's labor pool.

We therefore use the data generated by our field experiment to estimate a structural model of unobserved worker heterogeneity. This approach allows for an in-depth analysis of a wide array of questions. In particular, 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 phenomena that are common to all workers, 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 welfare 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 both wage offers and CSR advertisement at the recruitment stage. We find evidence of significant worker pool composition shifts on four dimensions of worker heterogeneity—two separate aspects of 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). Specifically, it induces workers to increase their "active productivity"—meaning that a worker requires less time to produce output, conditional on being engaged in a productive task. In addition, it increases their "passive productivity"—meaning that they voluntarily spend a smaller fraction of their work day on non-productive down time. Since treated workers are producing output faster and taking 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 together imply that an internal work-stage 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 strong evidence that selection effects via both CSR and wages can produce economically meaningful improvements in the firm's bottom line. For example, for various levels of quality control on the firm's output, we find that 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 not only produce more units of output per hour, but also, with higher individual work quality fewer redundancy measures are required to ensure a given level of quality control. High-wage recruits on average also supply substantially more output to the firm each day through increased labor supply and productivity. This would reduce 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 of maintaining each worker. These results are of independent interest, and may may have bearing on debates over minimum wage laws and efficiency wages. They relate somewhat to the CSR question in that we find evidence that a profit-motivated firm may wish to voluntarily 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 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 costs and reap productivity benefits, the per-worker profitability of CSR may be higher as well.

The remainder of our paper proceeds as follows. Section 2.1 provides a brief literature survey and describes the field experimental design. Section 3 outlines the theoretical model and discusses its identification, with emphasis on how the experimental variation enables us to generate the requisite set of observables to uniquely pin down structural primitives. The identification argument provides a number of insights regarding how the experiment must be designed to achieve the proper variation in the data. Section 4 breifly discusses a GMM-style estimator based on our identification strategy and a stochastic dominance testing procedure, while relegating most technical details to an appendix. Section 5.1 presents descriptive statistics and preliminary results on worker application rates from the hiring stage. Sections 5.2–5.4 present the results from the structural model, including an exploration of treatment effects and selection effects. Section 6 presents a counterfactual simulation exercise to explore changes to the firm's distribution of costs and outputs from selection on wage offers and/or CSR. Section 7 concludes, and an appendix contains additional technical details, graphs, and tables.

2 Literature Review and Experimental Design

Our study is related to several strands of literature. 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). Through field experiments conducted via mTurk and Elance, Burbano (2016) estimates the reduced-form effects of CSR messages on labor output and salary requirements. Received empirical results suggest that CSR adds value to the firm both through increased output and by lowering employee salary requirements. A particularly interesting set of results we discuss below relates to how the treatment and selection effects in our experiment underpin how CSR and wages positively impact the bottom line. Furthermore, linking a structural model with a field experiment allows us to generate the requisite set of observables needed to identify and estimate the link between unobservable worker characteristics and CSR activities on the part of the firm.[2]

Our paper also contributes to the broader empirical literature on the effects of CSR on firm performance (Greening and Turban (2000)). One recent paper by Bertrand et al. (2018) investigates the degree to which large firms are able to use corporate philanthropy as a tax-exempt form of political lobbying by donating generously to charities connected to elected officials. To complement this work, our study also explores how CSR affects the bottom line, but rather than through lobbying channels, we examine the supply side effects of CSR through its impact on the firm's labor force and its productivity.

We also contribute to the relatively new strand of literature within economics which uses experiments to generate data to identify structural models. Previous work has employed field experiments for structural analyses of, for example, 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. (2017)). 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.

In this manner, the methodology presented in this paper has large applicability in areas beyond the study of CSR. One could readily use an almost identical 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 DFT15 to generate the data necessary to identify the distribution of unobserved worker characteristics while allowing for worker selection effects and treatment 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. While the primary purpose of the hiring stage is to investigate the effect of CSR on the extensive margin—i.e., whether workers select into the job—the focus of the work stage is the intensive margin—i.e., the impact of CSR on the behavior of existing workers. The two stages are linked in the sense that any subject who participates in the work stage was also observed in the hiring 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 allows us to obtain detailed measures of output and labor input during each day of the sample period. Working subjects are assigned to different treatments on a daily basis

[2] 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.

which provides within-subject variation on non-pecuniary incentives.

The experiment was designed with several important conditions in mind. First, exogenous variation is necessary in both stages. This is for identification purposes, as will become clear in the next section when we formalize our model of worker behavior. Second, regardless of whether subjects were recruited neutrally or with CSR prompts, all workers are given the same treatment-control variation after the hiring stage. This aspect of our design is crucial for separately identifying 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 themselves. Fourth, the work task must be uniform across treatments to allow identification of labor supply costs, which include an idiosyncratic component and a common, non-parametric, 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 C 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 on the web. 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. For example, our empirical analysis requires detailed measurements of labor supply, output quantity, and output quality, with a work task that is simple enough to provide a clean data-generating process, but realistic enough to be meaningful as a case study that provides useful output. Creating our own consultancy firm also provided complete control over the size of the workforce, which would not have been possible in a partnership with an existing firm.

At least two further reasons dictated that we start our own firm rather than use an internet crowdsourcing platform. First, the subject pool in a crowdsourcing market is likely to exhibit certain characteristics that would be undesirable for our purposes. For example, subjects recruited via mTurk may have a tendency to pay less attention to experimental materials and may have a different attitude towards money 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)). This would be problematic in our case since the attitude towards pecuniary and non-pecuniary incentives is central to our research question, and our statistical power depends on subjects reading through the materials carefully. Second, the manner in which we assign subjects to treatments in the second stage, and the degree of details in terms of output measures, require a large degree of flexibility in the design of our online task system. This would make it quite difficult to operationalize the experiment on a crowdsourcing market.

2.2 Hiring Stage

Employment advertisements were posted on 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.

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 ad, there would be no way of controlling which ad each subject was exposed to and whether or not a given subject was exposed to one ad.[3] [4] The timing of the first stage of the experiment is summarized in figure 1.

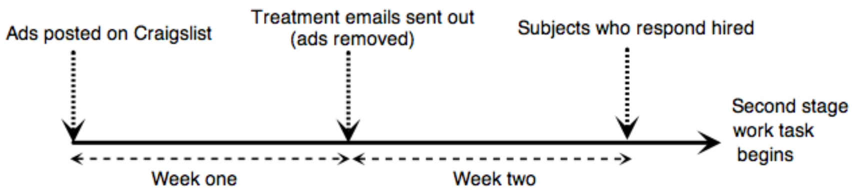

Figure 1: Hiring stage timing

We used two different classes of 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 differing information about the firm's involvement in CSR. The CSR information treatment for subject $i = 1, 2, ..., I$ 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. Treatment cells are summarized in Table 1.

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

Table 1: Hiring stage treatments

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^{th} and 75^{th} percentile wages for data entry operators in the United States.[5] 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

[3] This element of the design is similar to Flory et al. (2014) and Leibbrandt and List (2014) to study gender differences in attitudes towards competition and wage negotiations. 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 how to charge for health care products in developing countries.

[4] The separation between the initial response and the implementation of the treatment also permits us to balance treatment cells based on observables like gender and location. While we are not allowed to ask explicitly about gender, we are able to balance on gender for the purpose of randomization by predicting gender based on name, using probabilities derived from the Social Security Administration database on name popularity.

[5] See http:www.payscale.comresearchUSJob=Data_Entry_OperatorHourly_Rate (accessed on April 14, 2017).

Figure 2: Hiring Stage Nonpecuniary Treatments (for a high-wage subject)

Dear applicant,

Thank you for your interest in the position! We are sending this general first response to interested applicants.

First, a little more information about the job:

The position is focused on data entry tasks for a number of our clients. We provide services for a variety of different firms and organizations.

The hourly wage rate is $15.

In the beginning of each day, you will be provided with a number of files with image-coded data. Your responsibility is to manually enter the data into an online database using our web-based interface. You will be paid by the hour and the system will automatically log the number of hours worked each day. You may work as many or as few hours as you would like during the project period. The first project period lasts for about 2 weeks, but there is a possibility of continued work after the initial period.

Detailed information about the actual task will be provided to the individual(s) we proceed to hire.

If you are interested, please formally apply for the position by sending your resume to info@hhlsolutions.com. (We ask you to do this even if you have already sent us your resume in a previous e-mail.) Please put "Data Entry Position Application" in the subject line.

Best regards,

Diana Cavazos
HHL Solutions
Chicago, Illinois

(a) Neutral Recruitment Letter

Dear applicant,

Thank you for your interest in the position! We are sending this general first response to interested applicants.

First, a little more information about the job:

The position is focused on data entry tasks for a number of our clients. We provide services for a variety of different firms and organizations. Some of them 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 hourly wage rate is $15.

In the beginning of each day, you will be provided with a number of files with image-coded data. Your responsibility is to manually enter the data into an online database using our web-based interface. You will be paid by the hour and the system will automatically log the number of hours worked each day. You may work as many or as few hours as you would like during the project period. The first project period lasts for about 2 weeks, but there is a possibility of continued work after the initial period.

Detailed information about the actual task will be provided to the individual(s) we proceed to hire.

If you are interested, please formally apply for the position by sending your resume to info@hhlsolutions.com. (We ask you to do this even if you have already sent us your resume in a previous e-mail.) Please put "Data Entry Position Application" in the subject line.

Best regards,

Diana Cavazos
HHL Solutions
Chicago, Illinois

(b) CSR Recruitment Letter

in fact true, as discussed below). Figure 2 contains a side-by-side comparison of the hiring stage recruitment letters for a worker who was randomly assigned a high wage offer. As shown in the figure, 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 underpriviledged 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 variation in our pecuniary incentives. For a low-wage subject the fourth block of text below the salutation would instead read "The hourly wage rate is $11."

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 additional information provided in either treatment group, the subject makes the decision whether to apply for the job.

Subjects' subsequent 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 terse 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. Yet, the two treatment groups used in the first stage also provides two distinct groups of individuals entering the second stage of the experiment: those who applied after having been exposed to either treatment. In terms of the model, we obtain a vector $X_0 = (X_{01}, X_{02}, , ..., X_{0I})$ where each element contains the X_{0i} dummy for each agent $i = 1, \ldots, I$. In Section 3, we describe our structural model of unobserved worker characteristics which includes four dimensions of heterogeneity: two separate aspects of worker productivity, work quality, and preferences for leisure which dictate labor supply decisions. The hiring stage variation we have just described is crucial for pinning down how the firm's CSR activities create selection effects that shape the composition of the joint distribution of these characteristics within the firm's applicant pool.

2.3 Work Stage

2.3.1 The Task

In the hiring stage, to ensure that we generated a representative sample of workers across the treatment cells to perform the tasks while staying within our budget, we randomly chose to hire a subset of those who applied for the job. The chosen workers represented the participants in the 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 the project period. Concerning our model, one day corresponds to a period t so that $t = 1, 2, ..., T = 10$ 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 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 provided 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, while avoiding deception of our test-subject workers. Online Appendix C contains various screenshot examples of the main page where employees selected tasks to work on, and the task pages where they actually coded information about the *Google Streetview* images.

2.3.2 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 their recruitment status, within-person variation in treatments consisted of information about the client associated with the task of a particular day, and HHL's dealings with that client. This information was presented to subjects via the information text displayed when logging into the online work system and in information boxes on the main screen after tasks were completed.

The firm served two different types of clients—for-profit clients and non-profit clients—and the employees within each day concentrated on data entry for exactly one of these. On days when the CSR treatment was present, there was a short, color-coded prompt on the screen indicating that today's work was on behalf of 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, short, color-coded prompt appeared in the same location on their screen, indicating that they were working on behalf of a for-profit client, and thanking them for their efforts (see Online Appendix C for screenshots of the work-stage prompts). In order to avoid confusion on the part of the employee, the screen also indicated that their own wage rate is the same, regardless.

Each test subject employed by our firm was given the opportunity to work for a set of 10 contiguous days denoted by $t = 1, \ldots, 10$.[6] The variable X_{1it} served as our instrument for CSR during the work stage. 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

There are three basic outcome variables that we are able to observe during the work stage: (1) worker time inputs, (2) quantity of output, and (3) quality of output. Before outlining our empirical model of worker behavior, we pause here to describe in detail the observables from our experiment that will allow us to achieve econometric identification. The model is partitioned into two parts: labor supply and output quantity/quality, conditional on being at work. It is important to note that 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 stays inactive for longer than two minutes. This prevents subjects from gaming the system by logging in and earning money without actually working (or at least, without paying nominal attention to the computer screen). At the same time, this setup provided latitude for substantial heterogeneity across individuals, as well as within-person behavioral differences in terms of concentration level (reflected in work quality) and "feet dragging" (reflected in productivity).

Daily labor supply, H_{it}, is recorded as the total number of paid 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

[6] All employment spells occurred during September and October of 2016 in 6 waves, with each wave of test subjects having $t = 1, \ldots, 10$ occurring on the same days. Roughly half of all waves began on a Tuesday, and the rest began on a Thursday.

i's output from 1 to Q_i. Each worker processed a given image exactly once, though some were processed over multiple sessions.[7] For each completed unit of worker i's output, we have two measures of time expenditure: τ_{q_i} is the total amount of time that worker i spent on the data entry page for the *Streetview* image corresponding to q_i; and D_{q_i} is the amount of non-productive down-time (e.g., on the home page of the website, rather than on an image processing page) for the q_i^{th} unit of output.[8]

Since workers are paid by the firm for both types of time (as is the case for most 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 day that is actually spent on-task. A possible analog of the passive productivity measure within 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 producing the report.

Measuring work quality, or data entry accuracy rate, denoted A_{q_i}, is a little more involved. For each image there were 19 variables for workers to code on the image processing page.[9] Table 11 contains a brief summary of these. We prepared and distributed to each of our workers a training handout containing explanations and graphic examples of what constitutes an appropriate score—e.g., a "Street Quality" score of 3 as opposed to a 4 or 5—for each Likert-scale variable. In order to determine whether each worker correctly coded each variable we used a 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.[10]

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' tendencies to select the modal response when scoring subjective exam materials (similarly as we do) 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 ac-

[7]It was very rare for a worker to pause in the middle of data entry for one image and begin another before finishing the first. In the few cases when this did occur—a few dozen out of over 60,000—the image that was completed first was counted first in the sequence of worker i's output.

[8]More specifically, D_{q_i} measures the amount of down-time logged by worker i after completion of the $(q_i - 1)^{th}$ unit and before completion of the q_i^{th} unit of output.

[9]In order for an image to be completed, all 19 fields in the data entry form must be filled with some response by the worker. The website provided feedback to the user in cases where he/she left some fields empty. For this reason, it was very rare for images in our data to remain only partially processed by workers (well under 1%), and unfinished responses were dropped from the worker-output-level production data.

[10]In the case of the street name and city variables, correctness was determined by whether the user-input string matched the modal string, including whitespace characters and punctuation, but on a non-case-sensitive basis. In other words, correctness was judged on the basis of spelling but not capitalization.

tual 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.[11]

Table 11 in Online Appendix C displays descriptive statistics on responses and accuracy rates for each individual variable across a total of 62,138 worker-image observations. As the table illustrates, 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, as opposed to a simple crack in the pavement, also required a higher level of attention. Workers also generally appeared more adept at trans-coding street numbers than text strings for street names without making mistakes.

For each worker i and each unit of output q_i, the accuracy rate $A_{q_i} \in \{\frac{0}{7}, \frac{1}{7}, \ldots, \frac{7}{7}\}$ was coded as the mean across the variable-specific correctness outcomes for the final seven variables (below the dashed line in the Table 11). Given that our primary research focus is on worker heterogeneity, we define our accuracy index to focus on more difficult tasks in order to illuminate apsects of job performance that most set workers apart in a vertical sense. Table 11 also displays descriptive statistics on worker-image accuracy observations, with mean and standard deviations of 0.586 and 0.226, respectively.[12]

3 The Model

Our model of workers includes four dimensions of heterogeneity: active productivity, passive productivity, work quality, and leisure preferences. Productivity indexes the average amount of time it takes a worker to complete one unit of output. Work quality in our context is measured as "accuracy," or the degree to which a worker's output conforms to the employer's specifications.

[11]One additional concern was over the possibility that male and female workers may systematically diverge on fields requiring judgment calls, in a way that reflects underlying gender differences of mean opinion, rather than effort or ability. The raw response data suggested this is the case: female workers on average saw lower quality in the cars, buildings, picture, and streets, and on average saw higher quantities of litter and potholes than did their male counterparts. These mean gender differences were between 20% and 39% of the respective standard deviations. Other variables showed no sign of a divergence. In order to filter out systematic gender-based differences in mean judgment, our final accuracy measure defines a given response for one of the above six variables as correct if it matches the modal response within the worker's own gender group for that image-variable pair.

[12]As a robustness check, we also executed our analysis on an alternative form of the accuracy index which uses all variables in the table except v_6, v_7, v_{11}, and v_{12} (the correct values of these had little or no variance across the dataset and were therefore of no use to the client). The qualitative patterns we uncover are very similar, and the empirical magnitudes we estimate come close to what would result after a re-scaling of the above accuracy index, where the number 8 is added to both the numerator and denominator: the mean accuracy rate under this alternative measure is 78.6%, with a standard deviation that is about half as large, 0.116.

Leisure preference is quantified as the worker's monetized utility cost of supplying an hour of time to the firm. We describe in detail the sources of variation in observables which lead to identification of the structural primitives, while striving to impose as few *a priori* assumptions as possible. Our identification argument combines a principal-agent framework with panel-data methods and provides a guide for our experimental design. The purpose of the field experiment is to generate the requisite set of observables needed to identify and estimate the link between unobservable worker characteristics and CSR activities on the part of the firm.

3.1 The Baseline Model

Workers are recruited to produce outputs that the firm then sells on an open market each period $t = 1, 2, \ldots$ For simplicity, the firm produces only a single good, but not all units of output, denoted 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 is to some extent at the discretion of the worker. The firm offers a fixed hourly wage contract, w, for the worker to supply her available time to the firm in each period t.[13] Holding work quality fixed, maintaining peak productivity also requires exertion on the part of the worker, and is 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 of worker activities in real time 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,

$$\left(\Theta_{pi}, \Theta_{di}, \Theta_{ai}, \Theta_{li}\right),$$

which indexes their baseline productivity, $(\Theta_{pi}, \Theta_{di})$, "work quality" (or accuracy rate in our data entry context), Θ_{ai}, and shadow value of leisure time, Θ_{li}. The firm is characterized by a vector of commonly known and fixed characteristics, $\boldsymbol{Z}$—*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.[14] Let $(\mathcal{T}_p, \mathcal{T}_d, \mathcal{T}_a, \mathcal{T}_l)$ denote the treatment effects on productivity, accuracy, and labor supply, respectively, of laboring under condition $X_{1t} = 1$.[15] The idea here is that 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, thus motivating more and/or better outputs.

Upon receiving a job offer $(w, X_0, \boldsymbol{Z})$ 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

[13] Our model assumes that workers have no set schedule, and are allowed to choose their own hours at their discretion. This is to permit 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 entirely unnatural in the context of hourly wage 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 hours.

[14] One can alternatively think of the time-varying X_{1t} as the firm engaging in an advertising campaign in period t to remind or educate its workers of its CSR-oriented contributions to the greater good.

[15] Our framework and research design could also be applied to non-pecuniary investments in workplace quality, more generally. Although our applied research question focuses specifically on CSR, there is nothing in the economic or econometric theory which specifically requires that interpretation. If investment in a non-pecuniary workplace characteristic is one time and permanent thereafter, then $X_0 \in \{0, 1\}$, and $X_{1t} = X_0 \ \forall t$.

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 characteristicsG_{pdal} depends on the 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, given by

$$\tau\left(q_i; \Theta_{pi}, X_{1q_i}\right) = \Theta_{pi} \times \mathcal{T}_p^{X_{1q_i}} \times \tau_1 q_i^{-\delta} \times u^p_{q_i}, \quad q_i = 1, \ldots, Q_i. \tag{1}$$

Here, τ_1 is the baseline average time required for the first unit of output, δ represents the learning effect or elasticity of mean production time with regard to cumulative output, Q_i is total cumulative output by i, and $u^p_{q_i}$ is an exogenous, unit-specific shock with support on $\mathbb{R}_{++}$. In a slight shift of notation, X_{1q_i} is the CSR treatment that existed during the period when i produced her q_i^{th} unit of output.

Here, Θ_{pi} represents a permanent, idiosyncratic scaling of production time, with smaller values representing higher productivity workers. As Θ_{pi} or $\mathcal{T}_p$ decrease, the worker becomes more productive, holding other factors fixed. Because workers can select themselves into the firm's applicant pool, a firm engaging in CSR may attract a set of applicants whose unobserved characteristics are fundamentally different, *before* they even 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 activities as an investment in their bottom line, and therefore, they expect the baseline productivity distribution $G_p(\Theta_{pi}|X_0 = 0)$ to dominate $G_p(\Theta_{pi}|X_0 = 1)$, leading to advantageous selection in the worker pool. The sign and magnitude of this shift is a central focus of our empirical application.

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, and therefore the learning-by-doing component is omitted. Specifically, we model downtime on the q_i^{th} unit of output as

$$D\left(q_i; \Theta_{di}, X_{1q_i}\right) = \Theta_{di} \times \mathcal{T}_d^{X_{1q_i}} \times D_0 \times u^d_{q_i}. \tag{2}$$

Here, D_0 is the baseline average down-time and $u^d_{q_i}$ is an exogenous, unit-specific shock with support on $\mathbb{R}_{++}$. As before, Θ_{di} represents a permanent, idiosyncratic scaling of one's propensity to log unproductive down-time, with smaller values representing higher productivity workers. As Θ_{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

Similar to productivity, work quality is measured at the worker-unit level. With an eye toward our empirical application involving data entry workers, we use the terms "work quality" and "accuracy" interchangeably, and we will use the letter "A" to denote this dimension of worker heterogeneity. To fix ideas, we define a continuous, latent accuracy index $A^*_{q_i} \in \mathbb{R}$ as

$$A^*_{q_i} = \Theta_{ai} + X_{1q_i}\mathcal{T}_a + \varepsilon^a_{q_i}, \tag{3}$$

where $\varepsilon^a_{q_i}$ is an exogenous, iid, unit-specific shock that follows a symmetric distribution $\Phi(\cdot)$. The latent accuracy index encapsulates both permanent worker characteristics, treatment effects, and transitory output quality shocks. Conceptually, quality or accuracy of one's work on a particular data entry task becomes perfect as $A^*_{q_i} \to \infty$ and perfectly flawed as $A^*_{q_i} \to -\infty$.

However, a limited dependent variable problem is present, and we cannot observe the latent continuous accuracy index A^*_{qi} directly. Instead, we observe a binary measure $A_{qi} \in \{0,1\}$ where[16]

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

If we assume Φ is the standard normal CDF, we get the familiar probit formulation for accuracy,

$$\begin{aligned} \Pr[A_{qi} = 1|\Theta_{ai}] = \Pr[A^*_{qi} > 1|\Theta_{ai}] &= \Pr[\varepsilon^a_{q_i} > -\Theta_{ai} - X_{1q_i}\mathcal{T}_a] \\ &= \Phi(\Theta_{ai} + X_{1q_i}\mathcal{T}_a). \end{aligned} \tag{4}$$

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. Specifically, 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 $t = 1, 2, \ldots$ 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

$$C_{it}(H_{it}; \Theta_{li}, X_{1it}) = \Theta_{li} \times \mathcal{T}_l^{X_{1it}} \times c(H_{it}) \times u^l_{it}, \tag{5}$$

where u^l_{it} is an exogenous, iid shock with support on $\mathbb{R}_{++}$, and $c(\cdot)$ is a baseline cost function common to all workers in all periods. Worker type Θ_{li} and treatment $\mathcal{T}_l$ govern the shadow value of her 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. We assume standard regularity conditions which ensure a well-behaved decision problem:

[16] In reality, our data contain a finer (but still discrete) accuracy measure as explained in Section 2.4. One could alternatively specify a more complete model where for each coded variable, say $v_1, \ldots, v_7$, there is a variable-specific, iid, normal shock $\varepsilon^a_{v_k}$, so that $A^*_{v_k,q_i} = \Theta_{ai} + X_{1q_i}\mathcal{T}_a + \varepsilon^a_{v_k} + \varepsilon^a_{q_i}$, $A_{v_k,q_i} = \mathbb{1}\left(A^*_{v_k,q_i} > 0\right)$, and $A_{q_i} = \sum_{k=1}^{7} A_{v_k,q_i}/7$. One could potentially identify the relative variances of the $\varepsilon^a_{v_k}$'s (variable-specific shocks) after imposing a scale normalization—say, $Var\left(\varepsilon^a_{q_i}\right) = 1$ 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_{qi} are limited to a finite grid and shocks are uncorrelated across different q_i's.

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 Θ_{li}.[17] The cost shock term u^l_{it} is meant to represent day-to-day variation in a worker's time constraint. For example, one day her car might unexpectedly break down, necessitating a trip to a repair shop over an hour that could have otherwise been devoted to work. Another day, a friend might cancel a lunch appointment, leaving her with an extra, unanticipated hour of leisure to allocate.

On any day where worker i supplies positive time to the firm, she will choose H_{it} to optimize 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:

$$w = \Theta_{li} \mathcal{T}_l^{X_{1it}} c'(H_{it}) u^l_{it}. \tag{6}$$

Since costs are strictly convex, equation (6) implies that labor supply is monotone decreasing in Θ_{li}, $\mathcal{T}_l$, and u^l_{it} (for u^l_{it} 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. Intuitively, on days where cost shocks are very large, workers will find it optimal to simply not work. Accordingly, the worker's initial, extensive margin decision obeys the following rule:

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

Note that we have implicitly assumed the worker has observed only the daily shock u^l_{it} when she makes her extensive-margin decision. Since she must show up to work in order to observe treatment status X_{1it} on a given day, 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 it's 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. For example, if the CSR firm is successful at attracting a more productive set of workers who produce higher quality work, then their outside employment options may also be more valuable, leading to a mean increase in Θ_{li}. Alternatively, if the CSR-selected worker pool derives utility from contributing to the firm's socially beneficial mission, then their costs of supplying time to the firm may fall.

3.6 Structural Identification

Having outlined the basic components of our model of worker heterogeneity, we now turn to a discussion of structural identification. For reasons that will become clear below, certain functional form restrictions must be imposed on some model errors as identifying assumptions in order to cope with problems of limited dependent variable observations. Let $\Phi(\cdot, \mu, \sigma)$ denote the normal distribution with mean μ and standard deviation σ, and let $\phi(\cdot, \mu, \sigma)$ be its density.

Assumption 2. $\varepsilon^a_{q_i} \sim \Phi(\varepsilon^a_{q_i}, 0, 1) \; \forall q_i$, and $\varepsilon^l_{it} \equiv \log(u^l_{it}) \sim \Phi(\varepsilon^l_{it}, 0, \sigma_l) \; \forall i, t$.

[17] Note that any cost model $C(h) = \Theta_{li} c(h)$ is equivalent to any other cost model $\widetilde{C}(h) = \widetilde{\Theta}_{li} \widetilde{c}(h)$, where $\widetilde{\Theta}_{li} = a\Theta_{li}$ and $\widetilde{c}(h) = \frac{c(h)}{a}$ for come constant $a > 0$. Since the units of Θ_l have no inherent meaning, the model can only be identified up to a scale normalization. By constraining the boundary derivative to equal one, we are simply fixing the units of total costs by re-scaling $\Theta_{li} = \widetilde{\Theta}_{li} \times \widetilde{c}'(0)$ and $c(h) = \widetilde{c}(h)/\widetilde{c}'(0)$, with the result being that $c'(0) = 1$.

Let $G_{H|w_1}(h)$ and $G_{H|w_2}(h)$ denote the work time distributions conditional on wage contracts w_1 and w_2, respectively. For worker $i = 1, \ldots, I$ 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, $\{H_{it}, X_{1it}\}_{t=1}^{T}$. The complete model consists of the following structural components to be recovered from the observables: $\{G_{pdal}(\cdot,\cdot,\cdot,\cdot), \mathcal{T}_p, \mathcal{T}_d, \mathcal{T}_a, \mathcal{T}_l, \delta, \tau_1, \sigma_l, c(\cdot)\}$.

3.6.1 IDENTIFICATION: Productivity

Given that our passive productivity model is essentially identical to our active productivity model with the restriction that $\delta = 0$, it will suffice to discuss identification under the latter. In order to establish identification of the productivity model, we require a standard exogeneity condition on the log error terms. Letting $\varepsilon_{qi}^{p} \equiv \ln(u_{qi}^{p})$, we assume

Assumption 3. $E[\varepsilon_{qi}^{p}\ln(\Theta_{pi})] = 0$, $E[\varepsilon_{q_i}^{p}] = 0$, $E[\varepsilon_{q_i}^{p}\ln(q_i)] = 0$, and $E[\varepsilon_{q_i}^{p}X_{1q_i}] = 0$, $E[\varepsilon_{qi}^{d}\ln(\Theta_{pi})] = 0$, 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 both sides of equation (1) gives us

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

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. It is straightforward to see that identification obtains in the passive productivity case for the same reasoning when the second term on the right-hand side above is absent.

3.6.2 IDENTIFICATION: Accuracy

When considering the accuracy component of the model, it is necessary to impose the following:

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

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

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

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 function Φ, but this is something that can be accomplished without difficulty (see Appendix B for further discussion).

3.6.3 IDENTIFICATION: Labor Supply

Our identification strategy for labor supply builds upon the novel work of D'Haultfoeuille and Février (2015) and Torgovitsky (2015). We begin by imposing some key identifying assumptions on the log of the error term in the labor cost equation:

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

The reason we require independence of shocks and types is because of a sample selection problem that we will address in Section 3.6.5 below. Three further assumptions are necessary:

Assumption 6. $G_L(\Theta_l|w_1) = G_L(\Theta_l|w_2) = G_L(\Theta_l)$

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 $g_L(\Theta_l) \geq a > 0$, $\forall\, \Theta \in \left(\underline{\Theta}_l, \overline{\Theta}_l\right) \subset \mathbb{R}_{++}$.

In words, Assumptions 6 and 7 provide crucial exclusion restrictions that there must be no selective entry into the different contract groups in terms of labor supply costs. 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 Θ_{li}. The reader may have noticed that our recruitment-stage procedure induces self-selection on wage offer, but in the following section we explain how our research design is crafted to ensure that the above assumptions are met on a known subset of the data. For simplicity however, as we begin our discussion on identification we simply take them as given. Assumption 8 is a technical condition to avoid problems of partial identification, which will become clear below.[18]

3.6.4 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. Specifically, assume $G_{H|w_1}(0) = 0$, $\underline{\Theta} = 0$ (so that the support of work times is the positive real line for both labor contracts), and consider the baseline case of degenerate labor supply shocks where $E[U^l] = 1$ and $Var[U^l] = 0$ with $X_{1it} = 0$ in every period t. Before proceeding, it will be convenient to define $h_j(\Theta)$ as the optimal choice profile consistent with equation (6) (under a degenerate supply shock distribution), given wage contract w_j, $j = 1, 2$, and we also denote its inverse by $\Theta_j(h)$.

The mathematics behind the fully general identification results in DFT15 are fairly complicated, but in the basic setting with noiseless labor supply, a simple geometric argument, proposed by D'Haultfoeuille and Février (2011, henceforth, DF), 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 monotone $h_j(\cdot)$ leads to the implication 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 Θ_l):

$$\begin{aligned} G_L(\Theta_l) &= 1 - G_{H|w_1}(h_1(\Theta_l)|w_1) = 1 - G_{H|w_2}(h_2(\Theta_l)|w_2) \\ \Rightarrow h_1(\Theta_l) &= G^{-1}_{H|w_1}\left(G_{H|w_2}\left[h_2(\Theta_l)\right]\right). \end{aligned} \tag{9}$$

[18] Assumption 8 is testable, however: if the type distribution has full support and workers may choose any work time $h \in \mathbb{R}_+$, it follows that the work time distributions must also have full support.

DF 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. This is because monotonicity and the exclusion restriction imply that corresponding quantiles of work hours across different contract groups map into the same unobservable type Θ_l. Assumption 7 produces an inverse to this operation which DF refer to as the *vertical transform*. If we recall that $\Theta_j(h) = \frac{w_j}{c'(h)}$ from the first-order conditions, then by dividing the inverse choice profiles across the two contracts, we obtain the following relationship:

$$\Theta_1(h) = \frac{w_2}{w_1}\Theta_2(h). \tag{10}$$

Once again, given knowledge of some $(h, \Theta_2(h))$ pair under wage contract w_2, one can infer the unobservable type Θ_1 that would optimally choose the same labor supply 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.[19] 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 3. 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 Θ_l value that could rationalize each observed (h, w_j) pair under the model. This is because there is a (potentially infinite) sequence 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. Once either of the functions $\Theta_j(h)$ are known, the econometrician can also use the first-order conditions to recover the cost function from the differential equation $c'(h) = \frac{w_j}{\Theta_j(h)}$, $c(0) = 0$.[20]

3.6.5 CASE II: Noisy Labor Supply

Now we return to the full model where within-person variation in work times across different days arises from iid shocks to labor supply costs. The complication that arises in this case 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 on that day. In order to correct for this source of sample selection we require a parametric assumption (Assumption 2) on the distribution of labor supply shocks: $\varepsilon^l_{it} \sim \phi(\varepsilon; 0, \sigma_l)$.

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, we assume that the number of days over which a worker is observed is $T = 10$. We denote the order statistics of log-shocks on CSR-treated days and Neutral days by $\{\varepsilon^l_{Ci}(1{:}5) < \varepsilon^l_{Ci}(2{:}5) < \varepsilon^l_{Ci}(3{:}5) < \varepsilon^l_{Ci}(4{:}5) < \varepsilon^l_{Ci}(5{:}5)\}$, and $\{\varepsilon^l_{Ni}(1{:}5) < \varepsilon^l_{Ni}(2{:}5) < \varepsilon^l_{Ni}(3{:}5) < \varepsilon^l_{Ni}(4{:}5) < \varepsilon^l_{Ni}(5{:}5)\}$, respectively. Since work times are monotone in shocks, we denote the corresponding order statistics of work times for CSR and Neutral workdays $\{H_{Ci}(1{:}5) \geq \ldots \geq H_{Ci}(5{:}5)\}$ and $\{H_{Ni}(1{:}5) \geq \ldots \geq H_{Ni}(5{:}5)\}$, respectively. Note in this case that the indexing is reversed because higher shocks to labor supply costs lead to lower optimal work time choices.

[19] 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.

[20] An alternative way to view the basic identification argument is that there is one and only one value of $G_l(\Theta_l)$ that could rationalize each observed $\left(G_{H|w_1}(h), G_{H|w_2}(h)\right)$ pair under the model.

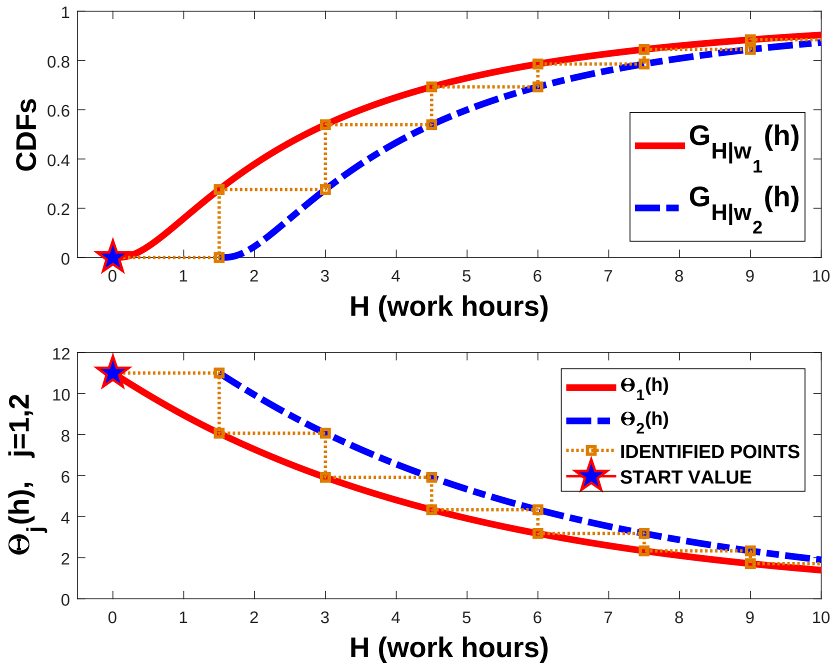

Figure 3: Geometry of Identification

We first show that if σ_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:

$$
\begin{aligned}
&\varepsilon^l_{Ci}(1{:}5), \varepsilon^l_{Ni}(1{:}5) \sim \phi_{(1:5)}(x;\sigma_l) = 5[1-\Phi(x;\sigma_l)]^4\phi(x;\sigma_l),\\
&\varepsilon^l_{Ci}(2{:}5), \varepsilon^l_{Ni}(2{:}5) \sim \phi_{(2:5)}(x;\sigma_l) = 20\Phi(x;\sigma_l)[1-\Phi(x;\sigma_l)]^3\phi(x;\sigma_l),\\
&\varepsilon^l_{Ci}(3{:}5), \varepsilon^l_{Ni}(3{:}5) \sim \phi_{(3:5)}(x;\sigma_l) = 30\Phi(x;\sigma_l)^2[1-\Phi(x;\sigma_l)]^2\phi(x;\sigma_l),\\
&\varepsilon^l_{Ci}(4{:}5), \varepsilon^l_{Ni}(4{:}5) \sim \phi_{(4:5)}(x;\sigma_l) = 20\Phi(x;\sigma_l)^3[1-\Phi(x;\sigma_l)]\phi(x;\sigma_l), \text{ and}\\
&\varepsilon^l_{Ci}(5{:}5), \varepsilon^l_{Ni}(5{:}5) \sim \phi_{(5:5)}(x;\sigma_l) = 5\Phi(x;\sigma_l)^4\phi(x;\sigma_l).
\end{aligned}
\tag{11}
$$

From these we can define $\mathbb{E}_{(j:5)}(\sigma_l) \equiv E[\varepsilon^l_{ki}(j{:}5);\sigma_l] = \int_{-\infty}^{\infty} x\phi_{(j:5)}(x;\sigma_l)dx;\ k = C, N;\ i = 1, \ldots, I;$ as the expectation of the j^{th} log-shock order statistic, for $j = 1, \ldots, 5$. 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 $E_{Ci}(j{:}5) \equiv \ln\left(\frac{W_i}{c'[H_{Ci}(j:5)]}\right) - \ln(\Theta_{li}) - \ln(\mathcal{T}_l)$, if $H_{Ci}(j{:}5) > 0$ on CSR days, and on Neutral days, $E_{Ni}(j{:}5) \equiv \ln\left(\frac{W_i}{c'[H_{Ni}(j:5)]}\right) - \ln(\Theta_{li})$, if $H_{Ni}(j{:}5) > 0$. For convenience, we simply define $E_{ki}(j{:}5) = 0$ whenever $H_{ki}(j{:}5) = 0$, for $k = C, N$.

Alternatively, equation (7) implies that whether or not $H_{Ci}(j{:}5) > 0$ or $H_{Ni}(j{:}5) > 0$ is true depends solely on whether $\ln\left(\frac{W_i}{c'(0)}\right) - \ln(\Theta_{li}) < \varepsilon^l_{ki}(j{:}5)$ for $k = C, N$, respectively. Since $\varepsilon^l_{Ci}(j{:}5)$ and $\varepsilon^l_{Ni}(j{:}5)$ are independent of Θ_{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

Θ_{li} are uniquely pinned down by the following set of up to 2×5 moment conditions for each i:

$$\begin{aligned}
\mathbb{E}_{(j:5)}(\sigma_l) &= E\left[\ln\left(\frac{W_i}{c'[H_{Ci}(j:5)]}\right) - \ln(\Theta_{li}) - \ln(\mathcal{T}_l)\right],\ \forall j = 1,\ldots,5\ \ s.t.\ \ H_{Ci}(j:5) > 0,\\
\mathbb{E}_{(j:5)}(\sigma_l) &= E\left[\ln\left(\frac{W_i}{c'[H_{Ni}(j:5)]}\right) - \ln(\Theta_{li})\right],\ \forall j = 1,\ldots,5\ \ s.t.\ \ H_{Ni}(j:5) > 0.
\end{aligned} \tag{12}$$

This shows why all other labor supply parameters are known if σ_l and $c(\cdot)$ are known, since fixing these objects implies a unique solution to the system of moment conditions in (12).

In order to close the identification argument, we note that σ_l and $c(\cdot)$ uniquely determine a set of work hour distributions as well. 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 $G_H(h) = 1 - G_U\left(f^{-1}(h)\right)$. This proves quite useful for the current purpose, since equation (6) indicates that hours supplied H_{it} are monotone decreasing in labor-supply cost shocks u^l_{it} since the common cost function is strictly convex. On any day when labor supply is positive, we can express the cost shock on that day as $u^l_{it} = \frac{W_i}{\Theta_{li}\mathcal{T}_l^{X_{1it}}c'(H_{it})}$, or alternatively, we can express labor supply as a function of the shock by inverting this equation to get $H_{it} = (c')^{-1}\left[\frac{W_i}{u^l_{it}\Theta_{li}\mathcal{T}_l^{X_{1it}}}\right]$. But, we also know that worker i will choose positive labor supply only on days when her cost shock does not exceed the cutoff $\overline{u}^l_i = \frac{w_1}{c'(0)\Theta_{li}}$. Moreover, the conditional shock CDF, given $U^l_{it} \le \overline{u}^l_i$, is $\frac{\ln\Phi(U^l_{it},\sigma_l)}{\ln\Phi(\overline{u}^l_i,\sigma_l)}$, where $\ln\Phi(\cdot,\sigma_l)$ is the CDF of the log-normal distribution.

We can use this to characterize the conditional distribution of positive work times implied by a (σ_l, c) pair for each individual i laboring under wage w_1. On CSR and Neutral days, they are

$$\begin{aligned}
H_i(U^l_{it})|_{X_1=1} &\sim G_{H_i|w_1,X_1=1}(h_i|w_1, X_1 = 1; \sigma_l, c) = 1 - \frac{\ln\Phi\left(\frac{w_1}{\Theta_{li}\mathcal{T}_l c'[h_i]};\sigma_l\right)}{\ln\Phi(\overline{u}^l_i,\sigma_l)},\ \text{ and}\\
H_i(U^l_{it})|_{X_1=0} &\sim G_{H_i|w_1,X_1=0}(h_i|w_1, X_1 = 0; \sigma_l, c) = 1 - \frac{\ln\Phi\left(\frac{w_1}{\Theta_{li}c'[h_i]};\sigma_l\right)}{\ln\Phi(\overline{u}^l_i,\sigma_l)},
\end{aligned} \tag{13}$$

respectively. The right-hand sides above truncate the support of the distribution at the cutoff $\overline{u} = \frac{w_1}{c'(0)\Theta_{li}}$ 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_2 instead, as

$$\begin{aligned}
H_i(U^l_{it})|_{X_1=1} &\sim G_{H_i|w_2,X_1=1}(h_i|w_2, X_1 = 1; \sigma_l, c) = 1 - \frac{\ln\Phi\left(\frac{w_2}{\Theta_{li}\mathcal{T}_l c'[h_i]};\sigma_l\right)}{\ln\Phi\left(\overline{u}^l_i;\sigma_l\right)},\ \text{ and}\\
H_i(U^l_{it})|_{X_1=0} &\sim G_{H_i|w_2,X_1=0}(h_i|w_2, X_1 = 0; \sigma_l, c) = 1 - \frac{\ln\Phi\left(\frac{w_2}{\Theta_{li}c'[h_i]};\sigma_l\right)}{\ln\Phi\left(\overline{u}^l_i;\sigma_l\right)},
\end{aligned} \tag{14}$$

for CSR and neutral workdays, respectively. Once again, the right-hand sides of the expressions above truncate the distributions at the appropriate cutoff where $H_{it} = 0$ is optimal. Therefore, i's model-generated conditional work time CDF and counterfactual work time CDF follow

$$\begin{aligned}
H_i(U^l_{it}) &\sim G_{H_i|w_1}(h_i;\sigma_l,c) = \frac{G_{H_i|w_1,X_1=1}(h_I|w_1,X_1=1;\sigma_l,c) + G_{H_i|w_1,X_1=0}(h_i|w_1,X_1=0;\sigma_l,c)}{2}\ \text{ and}\\
H_i(U^l_{it}) &\sim G_{H_i|w_2}(h_i;\sigma_l,c) = \frac{G_{H_i|w_2,X_1=1}(h_i|w_2,X_1=1;\sigma_l,c) + G_{H_i|w_2,X_1=0}(h_i|w_2,X_1=0;\sigma_l,c)}{2}.
\end{aligned} \tag{15}$$

Finally, to derive the overall conditional work time distributions implied by the model, we simply average across all observed individuals, i:

$$G_{H|w_1}(h;\sigma_l,c) = \frac{\sum_{i=1}^{I} G_{H_i|w_1}(h;\sigma_l,c)}{I} \quad \text{and} \quad G_{H|w_2}(h;\sigma_l,c) = \frac{\sum_{i=1}^{I} G_{H_i|w_2}(h;\sigma_l,c)}{I}. \tag{16}$$

Note that by characterizing both actual work time distributions and counterfactual work time distributions for every individual, equation (16) implicitly has the vertical and horizontal transform operations built-in through equations (14) and (15). The role that the parametric assumption has in our formulation is to pin down the appropriate truncation points in those same expressions, so as to correct for the sample selection problem.

Finally, in the model with non-degenerate shocks, all within-person variation in labor supply decisions across days with the same treatment status is driven entirely by variation in the shocks. Therefore, within-person variation in labor supply is a sufficient statistic for pinning down σ_l. Thus, by combining the baseline identification results proven by DFT15 with our sample selection correction it follows that the model components $\{\Theta_{li}\}_{i=1}^{I}$, $\mathcal{T}_l$, σ_l, and c are identified from the empirical distributions of positive work times $G_{H|w_1}$ and $G_{H|w_2}$, given our assumption of log-normal shocks.

3.6.6 Controlling for Sample Selection on Wage Offer

Recall from Section 3.6.3 that an exclusion restriction, $G_L(\Theta_{li}|w_1) = G_L(\Theta_{li}|w_2) = G_L(\Theta_{li})$, is necessary to identify the labor supply model. However, since workers are allowed to respond to their randomly assigned $W_i \in \{\$11, \$15\}$ by applying for the position (or not), it is likely that worker selection on wage offer is also present in the data-generating process. Indeed, Table 2 confirms this is so: holding recruitment status fixed, a shift from wage offer $w_1 = \$11$ to $w_2 = \$15$ increased the number of applications by roughly 30%. Therefore, if we define $\mathcal{S}_j = \{H_{it} | H_{it} > 0 \;\; \& \;\; W_i = w_j\}, \;\; j = 1, 2$ as the set of strictly positive work times logged under wage w_j, then it follows that the distribution of cost types Θ_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 supplied labor at a price of w_2. However, since we randomly assign wages, and since we have data on differential application rates by wage groups during the hiring stage, a correction can be made to restore validity of the identifying 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 Θ_{li}, such that the worker accepts a job with wage offer w if and only if $w \geq \zeta(\Theta_{li})$.[21] 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 $\overline{\Theta}_1(X_0)$ 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 $\overline{\Theta}_2(X_0) > \overline{\Theta}_1(X_0)$. Moreover, the additional applicants entering the pool under w_2 all have cost types above $\overline{\Theta}_1(X_0)$, with the conditional distribution of applicants, given $\Theta \leq \overline{\Theta}_1(X_0)$ being the same as it was under w_1. Within the worker choice model, conditional on being hired, mean within-person labor supply, $\mathrm{E}[H_{it}]$ is also monotone decreasing in Θ_{li}. Combining these two ideas implies a sample trimming rule to

[21] Such phenomena commonly arise from standard labor search models; e.g., see Ljungqvist and Sargent (2004).

derive a subset of high-wage workers for whom the exclusion restriction will be satisfied when comparisons to the set of low-wage workers are made.

To better understand why, 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 $G_{\overline{H}^{kj}}(h)$ denote the CDF of (within-worker) mean work times. Note that these are both known objects since they can be derived directly from the observables. With that, we can define for each recruitment group $k = C, N$ the trimmed positive work time samples

$$\begin{aligned} \widetilde{\mathcal{S}}_{N2} &\equiv \left\{ H_{it} : H_{it} > 0,\ i \in \left\{ {\scriptstyle n: W_n = w_2, X_{0n} = 0, E[H_{nt} | X_{0n} = 0, W_n = w_2] > G^{-1}_{\overline{H}^{N2}}(e_N)} \right\} \right\},\ \ e_N \equiv \frac{\mu_{N2} - \mu_{N1}}{\mu_{N1}}, \quad \text{and} \\ \widetilde{\mathcal{S}}_{C2} &\equiv \left\{ H_{it} : H_{it} > 0,\ i \in \left\{ {\scriptstyle n: W_n = \$15, X_{0n} = 1, E[H_{nt} | X_{0n} = 1, W_n = w_2] > G^{-1}_{\overline{H}^{C2}}(e_C)} \right\} \right\},\ \ e_C \equiv \frac{\mu_{C2} - \mu_{C1}}{\mu_{C1}}. \end{aligned} \tag{17}$$

Note that these two sets exclude the additional influx of 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 $\overline{H}^{N2}$ and trim high-wage, neutrally recruited workers with mean labor supply below that cutoff. We then do the same thing within the distribution of CSR-recruited workers.

With that completed, if we define $\mathcal{S}_2 = \widetilde{\mathcal{S}}_{C2} \cup \widetilde{\mathcal{S}}_{N2}$, then the resulting type distributions that underlay $\mathcal{S}_1$ and $\mathcal{S}_2$ will now satisfy the exclusion restriction, so that the common cost function c and shock variance σ_l can be identified using the arguments from Section 3.6.5. Further, once the common cost function c is known, idiosyncratic types Θ_l for *all* test subjects can be identified, including those who were excluded by the trimming rule, through the system of moment conditions in (12). In this way, we are able to both identify the underlying type distribution and study the effects of worker selection on wage offer as well.

4 Estimation

Having established how the moments in the data pin down the structural components of the model, it is not difficult to imagine how an estimator could be implemented by moment matching. Since the ideas behind our estimator follow from the identification argument, but the math needed to formally define the estimator is quite involved, we discuss estimation on an intuitive level here, while relegating the technical details to Appendix B. Recall that the various model components to estimate include basic common parameters, $(\delta, \tau_1, D_0, \sigma_l, c(\cdot))$; treatment effects, $\boldsymbol{\mathcal{T}} = (\mathcal{T}_p, \mathcal{T}_d, \mathcal{T}_a, \mathcal{T}_l)$; and the distribution of worker fixed effects, $G_{pdal}(\Theta_p, \Theta_d, \Theta_a, \Theta_l)$.

The productivity and work quality models are estimated via a simple differencing estimator. In order to obtain estimates of treatment effects, we difference mean production times and accuracy rates across treatment states ($X_1 = 0$ vs $X_1 = 1$), and then average these differences across individuals. We average initial production times across workers to obtain τ_1 (and D_0). To estimate the learning effect δ 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 in hand, estimating individual fixed effects is essentially computing a within-worker mean of production times or accuracy rates, conditional on treatment and/or learning. One important caveat here is that for the accuracy model there is a necessary bias correction (described in Appendix B) due to the interaction of non-linearity in the function $\Phi(\cdot)$ and finite-sampling variation in observed accuracy rates A_{q_i}.

The labor supply estimator is somewhat more involved, but still akin to a basic moment matching scheme. We begin by adopting flexible parametric forms to represent the work time distributions, $G_{H|w_1}$ and $G_{H|w_2}$, and the common cost function, $c(h)$. In our implementation we use B-splines and find that they provide an excellent mixture of flexibility and computational convenience (see Appendix B for further discussion). In a first step we estimate the aggregate empirical work time CDFs $\widehat{G}_{H|w_j}(h),\ j = 1,2$ by matching the B-spline forms (subject to shape restrictions and boundary conditions) to the empirical quantiles after implementing the selection correction in equation (17). In a second step, we use equations (11)–(17) to construct the model-generated (selection-corrected) analogs of both $\widehat{G}_{H|w_1}(h)$ and $\widehat{G}_{H|w_2}(h)$. Recall that these depend on the various model parameters including treatment effect, $\mathcal{T}_l$, the common cost function, $c(h)$, the supply-cost shock variance term σ_l, and individual fixed effects $\mathbf{\Theta}_l = \{\Theta_{l1}, \Theta_{l2} \ldots, \Theta_{lI}\}$. Therefore, model parameter values are chosen to optimize a least-squares fit criterion between the empirical aggregate work-time CDFs and their model-generated analogs.

Finally, with individual worker fixed effects $(\widehat{\Theta}_{pi}, \widehat{\Theta}_{di}, \widehat{\Theta}_{ai}, \widehat{\Theta}_{li})$ estimated, we can straightforwardly estimate the marginal distributions of these characteristics and their pairwise correlations using between-worker and within-worker variation. The final caveat to estimation is that for various model components (e.g., treatment effects $\boldsymbol{\mathcal{T}}$ and type distributions G_{pdal}), we obtain more precise information (in a statistical sense) for inference from workers who work more hours and produce more output. Essentially, this is because our data constitute an unbalanced panel. Therefore, whenever appropriate, we construct between-person means using inverse variance weighting; that is, by weighting according to total hours worked or total units produced. The interested reader is directed to Appendix B for additional detail.

4.0.1 Bootstrap Inference

Due to the computational complexity of our estimator, we chose to compute standard errors and confidence bounds by the bootstrap method, rather than computing the usual GMM standard errors (see Appendix B for a brief discussion of the applicable GMM sampling distribution theory). 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 $(\varepsilon^p_{q_i}, A_{q_i})$-pairs to construct a sample of output data.

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

4.0.2 Inference on Stochastic Dominance of Worker Characteristics

There are several papers in the statistics and econometrics literature that focus on 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 to the statistical properties of the original Kolmogorov-Smirnov (KS) method. Examples include Davidson and Duclos 2000, Barrett and Donald 2003, and Linton et al. 2010 (see Heathcote et al. 2010 for a survey and comparison of the methods in this literature).

To develop some intuition for how these methods work, we suppose there are two random variables $X \sim F_x$ and $Y \sim F_y$ defined on a common support $[\underline{z}, \overline{z}]$, 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 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) $\exists\ z \in (\underline{z}, \overline{z})$ such that $F_x(z) < F_y(z)$ and $\exists\ z' \in (\underline{z}, \overline{z})$ such that $F_x(z') > F_y(z')$. Therefore, 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 $T^{(i)} = \sqrt{\frac{N_x N_y}{N_x+N_y}} \sup_{z\in[\underline{z},\overline{z}]}[\widehat{F}_x(z) - \widehat{F}_y(z)]$ and $T^{(ii)} = \sqrt{\frac{N_x N_y}{N_x+N_y}} \sup_{z\in[\underline{z},\overline{z}]}[\widehat{F}_y(z) - \widehat{F}_x(z)]$. Finally, these are then evaluated at their limiting distribution (due to Doob 1949) to get p-values $[p^{(i)}, p^{(ii)}] = \exp\left(-2(T^{(i)}, T^{(ii)})^2\right)$. Finally, for some fixed significance level $\alpha \in (0, 1)$, if $\min(p^{(i)}, p^{(ii)}) > \alpha$ then we fail to reject the hypothesis of 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). Later variants of the stochastic dominance test (e.g., Davidson and Duclos 2000, Barrett and Donald 2003, and Linton et al. 2010) generally follow a similar proceedure but vary by how the test statistics and p-values are computed.

An important drawback to these existing 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, an important complication is that our data are not direct observations from the distributions of worker characteristics $(\Theta_p, \Theta_a, \Theta_l) \sim G_{pal}(\cdot,\cdot,\cdot|X_0 = 1)/G_{pal}(\cdot,\cdot,\cdot|X_0 = 0)$, but rather, they are *estimates* of worker characteristics: $\{\widehat{\Theta}_{pi}, \widehat{\Theta}_{ai}, \widehat{\Theta}_{li}\}_{i=1}^{I}$. Thus, the testing procedures mentioned above produce test statistics and p-values that would tend to under-reject the null hypothesis of equality. Intuitively, this is because the finite-sample variability in worker type estimates can be thought of as a measurement error problem that attenuates differences between the estimated distributions; e.g., $\widehat{G}_p(\Theta_p|X_0 = 0)$ versus $\widehat{G}_p(\Theta_p|X_0 = 1)$. With that caveat in mind, we use a bootstrapped adaptation of the KS test for stochastic dominance. Our dominance testing procedure follows closely one developed by Marmer et al. (2017).

First, we compute weighted empirical CDFs as described above (and formally in Appendix B) to mitigate partially the measurement error problem. Second, we define a grid of points in quantile space by $\boldsymbol{r} = \{r_1, r_2, \ldots, r_K\} = \{0.010, 0.011, \ldots, 0.990\}$ and $\boldsymbol{\theta}_j = \{\theta_{j1}, \ldots, \theta_{jK}\} = \widehat{G}_j^{-1}(\boldsymbol{r})$, $j = p, a, l$. We then evaluate the bootstrapped empirical CDFs $\widehat{G}^*_{js}(\theta_{jk}|X_0 = 0)$ and $\widehat{G}^*_{js}(\theta_{jk}|X_0 = 0)$ $j = p, a, l$, $k = 1, \ldots, K$ (conditional on recruitment status) at each of the points in these grids. Similarly as in previous methods, for each $j = p, a, l$ 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 $H_3 : G_j(\cdot|X_0 = 0) <> G_j(\cdot|X_0 = 1)$. Finally, we compute a p-value for H_1 by $p^*_{1j} = 1 - \max_{k=1,\ldots,K} \frac{\sum_{s=1}^{S} \mathbb{1}\left(\widehat{G}^*_{js}(\theta_{jk}|X_0=0) - \widehat{G}^*_{js}(\theta_{jk}|X_0=1) > 0\right)}{S}$, and symmetrically for H_2 by

$p^*_{2j} = 1 - \max_{k=1,\ldots,K} \frac{\sum_{s=1}^{S} \mathbb{1}\left(\widehat{G}^*_{js}(\theta_{jk}|X_0=1) - \widehat{G}^*_{js}(\theta_{jk}|X_0=0) > 0\right)}{S}$.[22] 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. As with other previous methods, 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.

In our empirical results section, we will also explore worker selection on the basis of wage offer, using this same process.

5 Empirical Results

Our discussion of results begins with a summary of insights gained from the hiring stage. We first look at how different wage rates and non-pecuniary incentives from CSR affect application rates at the hiring stage. We then present insights from the work stage on how wages and information about CSR affect labor supply and output. We then explore unobserved characteristics and selection effects. We conclude the results summary with counterfactual model simulations that capture the impacts of differing recruitment strategies on a firm's cost structure, through their influence on the pool of worker characteristics.

5.1 Application Rates and Descriptive Statistics

The recruitment stage of the experiment was conducted during the fall of 2016. The cities included in the experiment were 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 (containing randomized wage offer and CSR/Neutral framing) 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 no information about CSR (n=266), (2) \$11 wage rate and information about CSR (n=309), (3) \$15 wage rate and no information about CSR (n=277), and (4) \$15 wage rate and information about CSR (n=251). The top panel of Table 2 presents mean application rates, or the share of subjects proceeding to apply for the job, in each treatment cell. The bottom panel explores differences in application rates across treatments.

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. Interestingly, the data also show that including information about CSR in the job description increases the application rate 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 their

[22] We also bias correct our bootstrap testing procedure for each θ by re-centering the boostrapped 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 $\theta \in supp(\Theta_j)$.

	Mean	Estimate		SD	90% CI
APPLICATION RATES					
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	$[0.432, 0.556]$
APPLICATION RATE DIFFERENCES					
within wage offer group					
	Mean Difference	*Estimate*	*P-value*	*SD*	*90% CI*
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	$[0.106, 0.113]$
within nonpecuniary recruitment group					
	Mean Difference	*Estimate*	*P-value*	*SD*	*90% CI*
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

neutral-recruited counterparts, an increase of roughly 24% in total applications received. Both shifts are statistically significant, as are the effects broken down by the four stage 1 treatment bins.[23] In addition, it is worth noting that the observed application rates and wage elasticities are roughly similar to those found in comparable studies (see for example Flory et al. (2014); Leibbrandt and List (2014)). Thus we see that the volume of workers who self select into the labor pool increases with CSR, but the question that remains is whether the composition of characteristics within the worker pool changes as well. This represents the crux of our focus in the structural model of worker fixed effects.

Table 3 presents summary statistics for the work stage of the experiment. Subjects were randomly sampled from each of the four hiring stage treatment cells to achieve a balanced sample of types and genders in the work stage. A total of 170 subjects participated in the work task stage of the experiment. Each subject was observed during a period of 10 days, which implies 1700 individual-day observations. Some subjects did not supply a positive amount of hours every day. The mean number of days worked during the experiment across all subjects was roughly 5 out of 10. The mean daily shift time was roughly 2 hours, for mean earnings of about \$27 per work day. On average, workers would process roughly 54 images on a given shift, implying a per-unit average cost of \$0.498.

A striking feature of the table is the considerable variation across workers. On a given day, many workers could not find the time to work, while others occasionally logged shifts upwards of 16 hours. There is considerable heterogeneity along the dimensions of 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. Taking both measures into account, 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 as well, ranging between 30% and 86%.

As with application rates, we see substantial differences in the raw data by worker recruitment bins. High-wage recruits logged substantially more hours per day, though this raw change

[23] Robustness checks indicate no statistically significant difference in these results across cities.

WITHIN-WORKER SUMMARY STATISTICS						
Variable	*Median*	*Mean*	*SD*	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

WAGE-OFFER GROUP DIFFERENCES: *Group Outcome*$\|\$15 -$ *Group Outcome*$\|\$11$					
Outcome	*Median Difference*	*Mean Difference*	*P-Value for H_0: Equal Means*	*90% CI*	*DF*
Hours worked per day	0.1636	0.4579	1.3×10^{-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

NON-PECUNIARY GROUP DIFFERENCES: *Group Outcome*$\|CSR -$ *Group Outcome*$\|Ntr$					
Outcome	*Median Difference*	*Mean Difference*	*P-Value for H_0: Equal Means*	*90% CI*	*DF*
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.

confounds several underlying factors that will be disentangled by the structural model. On one hand there is a mechanical increase in their marginal incentives to supply time, but their underlying labor supply costs may also be different because of selection. The raw mean shifts also do not account for the role of within-person, day-to-day variation in labor supply. For CSR-recruited workers, we see a raw mean reduction in hours worked, which is more suggestive of a shift in unobserved costs since monetary incentives to supply time are the same as neutral counterparts. We also see non-trivial increases in productivity and accuracy among high-wage recruits and CSR recruits, though once again 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 Model Estimates: General Parameters

By estimating unobserved worker characteristics we can gain a more complete understanding into the workings of pecuniary and non-pecuniary job characteristics. We have seen that the number of applicants willing to accept a given wage offer increases with the presence of CSR, but are these new additions to the worker pool more or less productive? Do they produce better output? Are they more or less willing to supply time to the firm? In a nutshell, how much of an impact does CSR have on a firm's supply costs and output quality, and to what extent does the effect come through selection versus treatment effects? Given the popularity of CSR amongst profit-driven firms, a viable hypothesis is that there are meaningful benefits.

A key concern in answering these questions is how the composition of the labor force changes when CSR is used as a recruiting tool. That is why estimating a model of underlying worker characteristics is necessary. In addition, we also seek to understand whether there is a quality-quantity tradeoff: are there correlations between productivity, work quality, and willingness to supply time, and if so, what is their sign. Also, how is the relation between these characteristics affected by selection on wage offer and/or CSR?

	Parameter	*Estimate*	*Std Error*	*P-value*	*90% CI*
Learning:	δ	0.1467	0.1347 $(H_0{:}\delta{=}0)$	0.2760	$[-0.075, 0.368]$
Mean Initial Production Time (minutes):	τ_1	3.7016	2.8851	—	$[1.022, 13.411]$
Active Productivity Model Fit:	R^2 $(N = 62,138)$	0.7314	—	—	—
Baseline Mean Down-Time (minutes):	D_0	0.3587	0.0425	—	$[0.289, 0.429]$
Passive Productivity Model Fit:	R^2 $(N = 62,138)$	0.3342	—	—	—
Labor Supply Shock Standard Deviation:	σ_l	0.7658	0.4216	—	$[0.5135, 2.0928]$

Table 4: Basic Model Parameter Estimates

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 δ 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 δ 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.[24]

Figure 12 in Appendix A provides a graphical depiction of active productivity heterogeneity. It displays estimated learning curves (with confidence bands) for the 10$^{\text{th}}$ percentile, median, and 90$^{\text{th}}$ percentile individuals. After learning has largely ceased, the mean difference between the 10$^{\text{th}}$ and 90$^{\text{th}}$ 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 the R^2 for passive productivity is 0.33. Still, worker heterogeneity in passive productivity remains significant: the 10^{th} and 90^{th} percentiles of Θ_{di} differ by a factor of 4.

For the accuracy model and the labor supply model, goodness of fit is graphically represented in Figure 13 in Appendix A. 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. To provide an idea of the relative importance of transitory shocks versus permanent, idiosyncratic variation, we present some additional numbers as well. 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.[25] In other words, within the estimated model, permanent worker heterogeneity is quite important, but transitory, within-worker

[24] 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.

[25] Within-person standard deviation in accuracy rates is calculated from raw data by taking a weighted mean

quality shocks also play a substantial role.

For labor supply, this idea is summarized by $\widehat{\sigma}_l$, which implies a standard deviation of 1.1974 in transitory labor supply cost shocks. Contrast this to the (weighted) standard deviation of Θ_l, which is 7.0736. Though at first glance this seems like a large difference, the two standard deviations are much closer when expressed as fractions of their means. To illustrate, consider a thought experiment where the average worker type $E[\Theta_l]$ experiences a mean labor supply 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% on that day. Alternatively, if we maintain her shock at the mean value, but reduce her permanent cost type $E[\Theta_d]$ by one standard deviation, her supply costs fall by 58.58%. In other words, permanent cross-worker variation in labor supply costs play roughly the same size role as day-to-day, within-worker variation of determining aggregate variation in labor supply to the firm. Figure 14 in Appendix A displays the estimated common supply cost curve and marginal cost curve for the median Θ_l in the sample.

5.3 Model Estimates: Work Stage Treatment Effects

Table 5 displays point estimates and standard errors on parameters which represent 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.

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

Table 5: Parameter estimates: Treatment Effects

Our first result is that the firm saw substantial productivity gains from the internal CSR advertisement. In particular, active production times reduced by 24.2% under treatment, while

across individuals of the following quantity $\sigma_{Ai} = \frac{Q_{Ci}\sigma^C_{Ai} + Q_{Ni}\sigma^N_{Ai}}{Q_{Ci}+Q_{Ni}}$, where $\sigma^j_{Ai} = \sqrt{\frac{\sum_{q_i=1}^{Q_i}(A_{q_i} - \mu^j_{Ai})^2 \mathbb{1}(X_{1q_i} = \mathbb{1}(j=C))}{Q_{ji}-1}}$, $j = C, N$, and where $\mu^j_{Ai} = \frac{\sum_{q_i=1}^{Q_i} A_{q_i}\mathbb{1}(X_{1q_i} = \mathbb{1}(j=C))}{Q_{ji}}$, $j = C, N$. In the above formula for σ_{Ai}, σ^j_{Ai} is excluded whenever $Q_{ji} < 2$, and when taking the weighted mean of σ_{Ai} we exclude the single test subject who only produced one unit of output. The cross-person standard deviation in baseline accuracy rates is calculated as the square root of a weighted mean of the following quantity $\left(\Phi(\Theta_{ai}) - \mu_{\hat{A}i}\right)^2$, where $\mu_{\hat{A}i}$ is a weighted mean of $\Phi(\Theta_{ai})$ (once again excluding the single test subject who only produced one unit of output). For all weighted means, weights on individual i are calculated as $\sqrt{Q_i}$, and normalized so that they sum to one.

non-productive downtime dropped by 42.5%. Both of these shifts are statistically significant. Given the large speed-up in production of each unit of output, it may be less obvious *a priori* what the expected sign of the treatment effects for work quality should be. On one hand, the CSR treatment may inspire workers to exert themselves in producing accurate output. However, since they are producing units of output more quickly and taking less rest time in between, it is also possible, through burnout or mental fatigue, that work quality may fall. Indeed, in the third row of Table 5 we see a negative point estimate. However, the effect is economically insignificant, being only enough to reduce mean predicted accuracy for the average worker by about one half of a percentage point. Taking this result in combination with the previous estimates, we find that the work-stage CSR treatment substantially increases productivity and that this speed-up of production is virtually all valuable to the firm.

5.4 Model Estimates: Unobserved Characteristics and Selection Effects

In this section 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. We first compute the empirical CDFs of worker types separately for the subsamples of CSR recruits and neutral recruits. We then run our bootstrap testing procedure for stochastic dominance, repeating the same process for the subsamples of high-wage recruits and low-wage recruits. This facilitates a comparison of worker selection on both pecuniary and non-pecuniary job characteristics.

A summary of all stochastic dominance tests appear in Table 9 in Appendix A.[26] Whenever the null hypothesis of equality is rejected, we also report the estimated mean percent change in the relevant characteristic both in levels and as a fraction of a standard deviation within the subsamples involved in the test. Finally, to assess the magnitudes of combined effects, we summarize two additional stochastic dominance tests in Table 9. 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 of workers who were offered a high wage *and* CSR recruited. The first comparison informs us about the combined effects of CSR selection during the hiring stage and CSR treatment at the work stage, and the latter comparison explores the combined selection on pecuniary and non-pecuniary job characteristics. We also examine correlations between unobserved worker characteristics to show how productivity, work quality, and the value of one's time are related.

5.4.1 Productivity

Figure 4 displays the distributions of estimated active productivity types Θ_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 unconditional median fixed effect is roughly one, it follows that the 10^{th} percentile worker averages production times at roughly half of the median worker, while the 90^{th} averages production times at roughly twice those of the median worker.

[26]While Table 9 in Appendix A reports the minimum point-specific p-values used to determine the outcomes of the stochastic dominance tests, the interested reader is directed to Supplemental Online Appendix D for a more complete picture, where Tables 12 – 19 display point-specific p-values over a fine grid of quantiles.

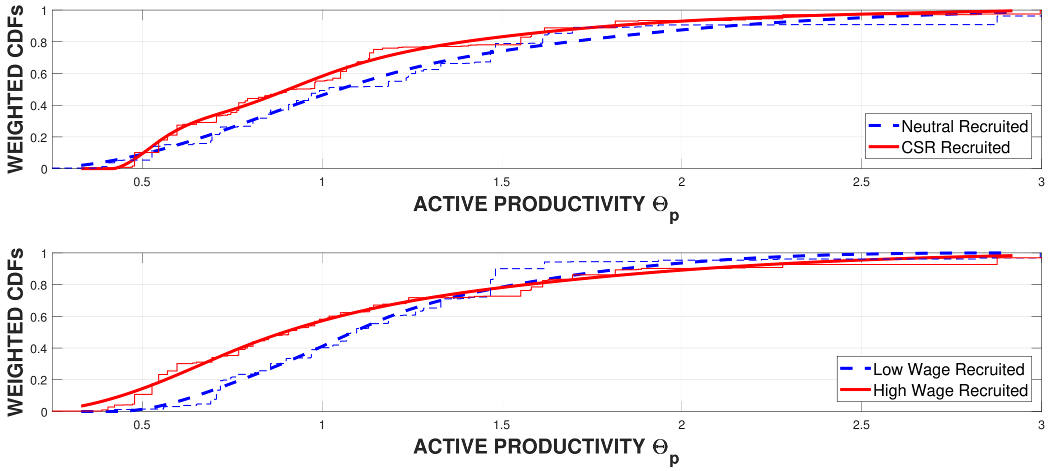

Figure 4: Productivity: Selection on Wage and CSR

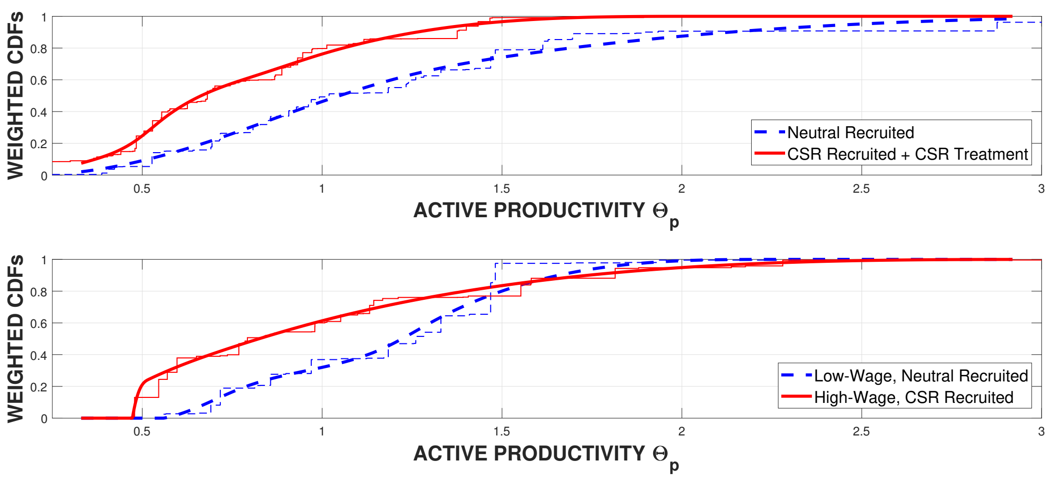

Figure 5: Productivity: Combined Effects

Figure 4 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 the alternative hypothesis of first-order dominance. Since lower values of Θ_p imply more output per unit of time, this means that including a CSR advertisement in the firm's job announcement led to an advantageously selected worker pool.

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 4) are similar though somewhat less decisive. There is a crossing of the two empirical CDFs at about the 80^{th} percentile, which results in the null of equality being 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 5 shows the combined effects of CSR selection and CSR work-stage

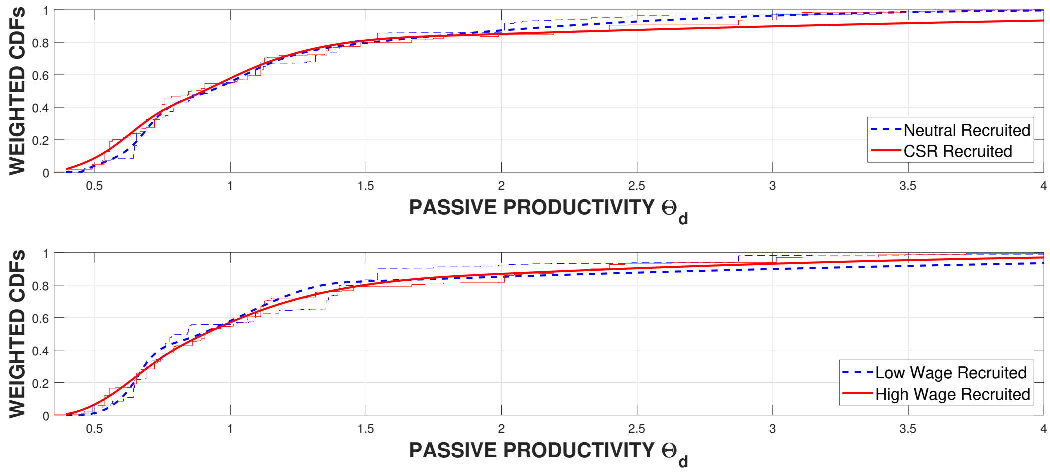

Figure 6: Productivity: Selection on Wage and CSR

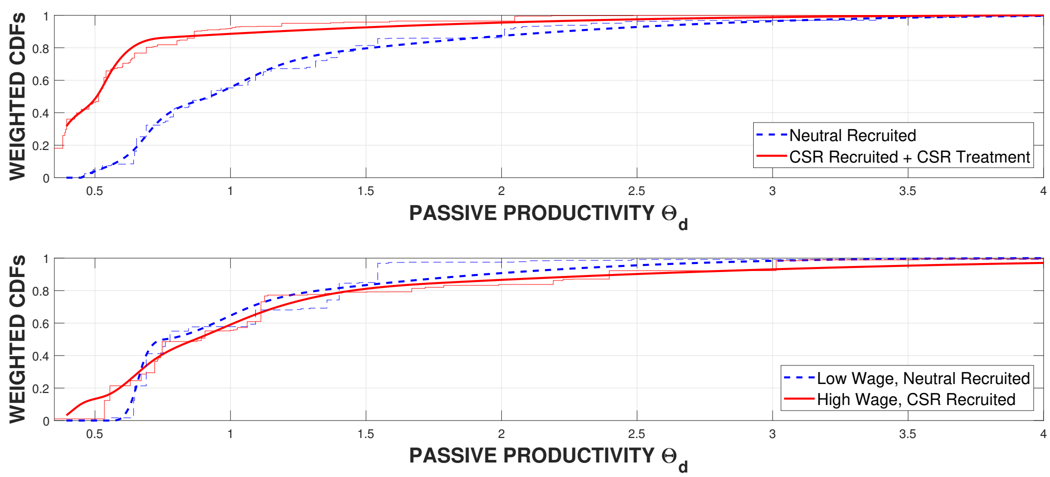

Figure 7: Productivity: Combined Effects

treatment, relative to the neutral-neutral baseline. Not surprisingly, 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 against the baseline of 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 that pecuniary and non-pecuniary job characteristics are complementary recruitment tools.

Estimated CDFs of passive productivity characteristics are shown in Figures 6 and 7. The level of cross-worker heterogeneity remains substantial, but in this case, the expansive upper tail of the distribution is particularly striking. 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. The selection effects for passive productivity types Θ_d are far less conclusive: the estimated CDFs have multiple crossing points and achieve enough separation in some small regions to reject the null hypothesis of equality, but no clear ordering arises.

5.4.2 Accuracy/Work Quality

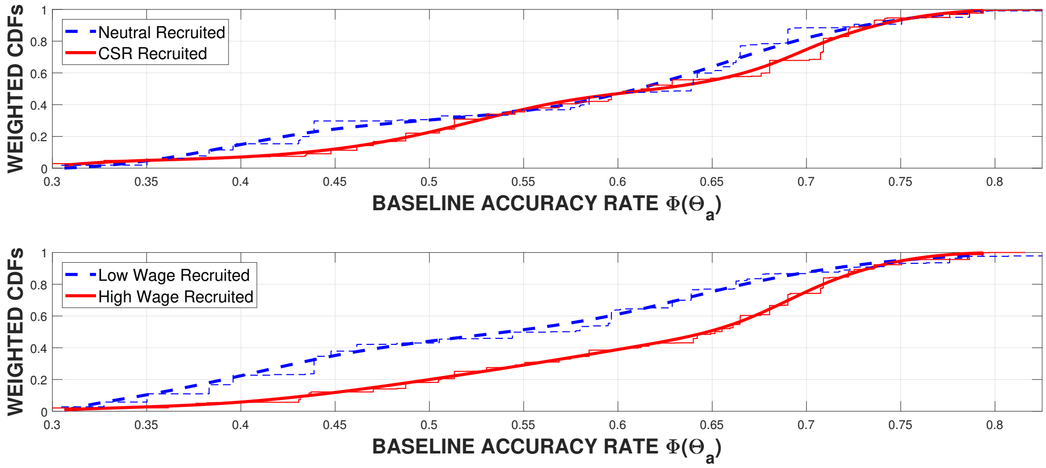

Figure 8: Accuracy Rates: Selection on Wage and CSR

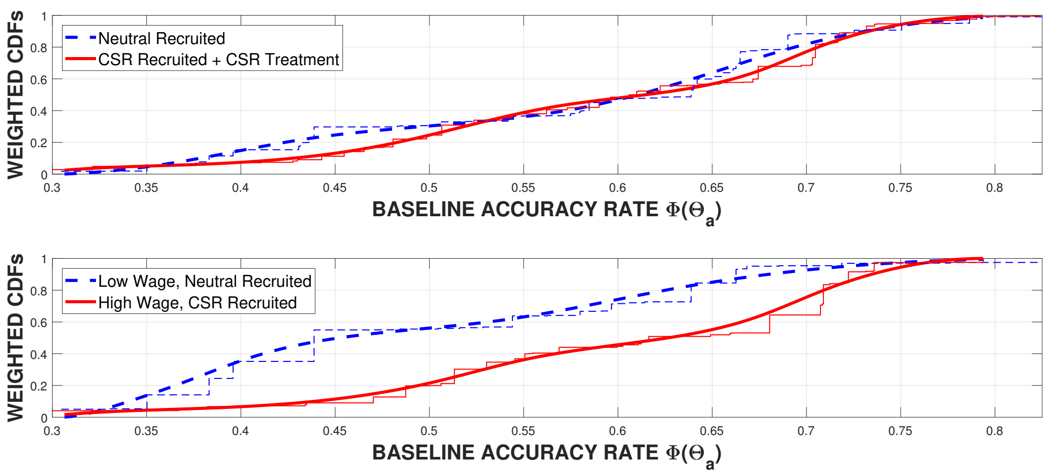

Figure 9: Accuracy Rates: Combined Effects

Figure 8 and the lower-left panel of Table 9 display results on individual work quality 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 on worker fixed effects point to a substantial degree of heterogeneity. Figure 8 indicates that recruiting with either high wage offers or CSR advertisement leads to a first-order dominance shift toward workers who produce superior output, which is 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 shall see through our model simulations that they can have a big impact on the firm's bottom line through quality control costs.

In Figure 9 (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 estimated mean accuracy rate improvement is about 21% (or 0.785 of standard deviations).

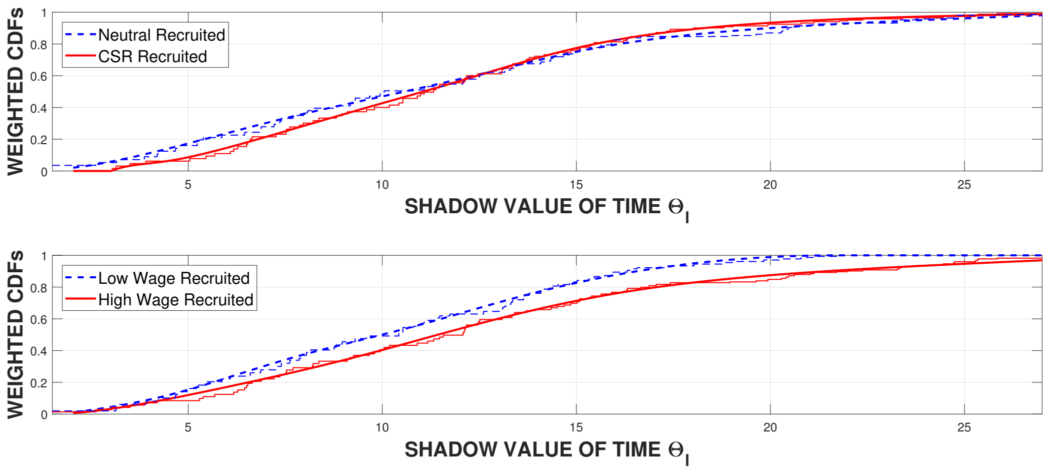

Figure 10: Shadow Value of Time: Selection on Wage and CSR

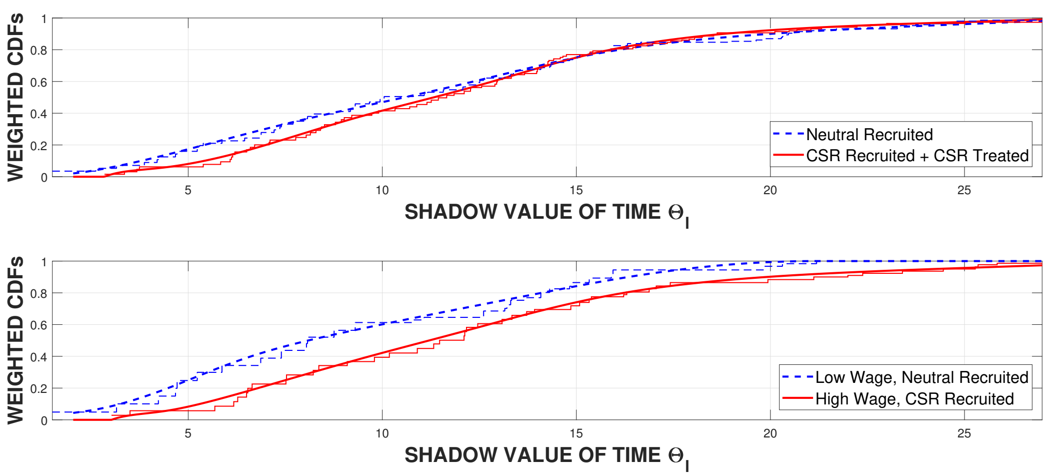

Figure 11: Shadow Value of Time: Combined Effects

5.4.3 Worker Labor Supply Costs

Figure 10 and the lower-right section of Table 9 display summary results on type distributions of Θ_l, the shadow value that workers place on their time when making labor-leisure decisions. Once again, a striking feature of the estimates is a high level of cross-worker heterogeneity. Relative to the median worker, the 10^{th} percentile worker's cost of supplying time to the firm is 60% lower, and the 90^{th} percentile worker's supply cost is 80% greater than the median. This would account for the long upper tail of observed positive work times in Figure 13 (see appendix) where the 90^{th} percentile (6.46 hours) is 4 times larger than the median (1.61 hours).

Figure 10 depicts selective entry into the worker pool based on pecuniary and non-pecuniary characteristics. 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 11 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 $(w, X_0, \boldsymbol{Z})$, if w is increased to some $w' > w$, we see an influx of new workers with more valuable time who are willing to accept the offer under w' but not under 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, worker non-pecuniary CSR status seems to play a role similar to money in attracting workers to the firm.

At the end of the day, it may not be surprising that CSR-selected and high-wage-selected workers have more valuable time, given our observations above that these same workers 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 Appendix A.4.1 where we compute various measures of correlations among our estimates of unobserved worker traits. There we find that variation in $\widehat{\Theta}_p$, $\widehat{\Theta}_d$, and $\widehat{\Theta}_a$ explains 34% of variation in $\widehat{\Theta}_l$.[27]

6 Bottom-Line Model Simulations

To conclude our empirical analysis, 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.[28] In order to assess the final impact of selection on wage offer

[27] Residual variation in $\widehat{\Theta}_l$ is partly sampling variability, so the 34% number is an estimated lower bound on the strength of the relationship between the actual value of time Θ and worker productivity (Θ_p, Θ_d) and quality (Θ_a).

[28] We only consider ex-post wage raises because labor laws constrain firms from reducing wages after offering a contract at level w during the hiring stage. Moreover, since wage offer increases at the hiring stage induce composition effects through entry of extramarginal worker types, an ex-post wage decrease would likely induce attrition.

and CSR on the firm's level of profit, we simulated 400 blocks of 10 workdays for each worker in our sample according to the following 3-step process:

1. We assume the worker is inexperienced before their first day of the ten, and that odd-numbered days are under Neutral task framing while even-numbered days are under CSR framing. 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 then iteratively simulate iid draws from the (empirical) joint distribution of productivity shocks (u^p, u^d) to obtain production times, until the simulated labor supply time for that day is exhausted.[29,30] 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 quality 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 industry contexts, work quality is not simply a second dimension of the production process, but rather, it feeds directly into production costs as well. This can occur 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 quality of final output—as is the case in our data-entry context.[31] 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.

Specifically, for a hypothetical set of K workers chosen at random from subsample j to evaluate a fixed field for the same image, we begin by assuming for simplicity that the objectively correct response for the field-image pair is 3 out of 5 on the Likert scale. For various values of K we then simulated 4×10^6 sets of $k = 1, \ldots, K$ random draws, $\{a_1, \ldots, a_K\}$, from the single-worker accuracy distribution $F_{acc,j}(a)$. For each set of K, we then simulate K uniform random numbers ν^{acc} and code worker k's response as correct if $\nu^{acc,k} \leq a_k$. If the simulated response is incorrect, then we randomly select the answer from among the remaining options according to the empirical distribution of incorrect responses in our raw data.[32] We then computed the mode within each set of K responses and recorded a binary variable for whether the modal response was the correct one (a value of 3), and averaged across all 4×10^6 observations to get a K-worker modal

[29] Within a simulated workday, we assume a worker will not exceed her planned worktime from step 1. Therefore, at the end of her shift, if she finds that she has a few minutes remaining, but not enough to accomplish another unit of output, we assume that she spends the remaining minutes in downtime, and we add them to the production time of the last unit of that day. Dropping those final minutes instead made very little difference.

[30] We found that prolonged learning effects led to unusually low simulated production times for the highest output workers, toward the end of their employment history. Therefore, after the 200^{th} unit of simulated output we hold the learning term $q^{-\delta}$ fixed at a value of $200^{-\delta}$ for all subsequent units of output. This learning cutoff allowed for a close match between the simulated distribution of per-unit production times and its analog in the raw data.

[31] As mentioned previously, our field experimental application of data entry with an element of ambiguity is most closely related to standardized exam scoring by testing firms such as Pearson, ACT, or the College Board.

[32] 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. We followed this convention when determining the values of incorrect responses in our redundancy simulations.

response accuracy rate. Since different subsamples of workers (e.g., CSR recruits versus 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 our results for both simulation exercises below.

6.1 Single-Worker Production Costs and Daily Supply (no 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—i.e., who produce more output per hour—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?

Recruitment Group:	*\$15*	*\$11*	*CSR*	*NTR*	*CSR15*	*NTR11*
MEAN DAILY OUTPUT:	52.71	38.10	44.83	48.15	56.17	45.45
% Differences:	%Δ(\$15,\$11):	+38.35%	$\%\Delta(CSR,NTR)$:	−6.89%	$\%\Delta(CSR15,NTR11)$:	+23.60%
MEAN PER-PAGE COST:	\$0.5552	\$0.4356	\$0.4732	\$0.5530	\$0.5000	\$0.4462
% Differences:	%Δ(\$15,\$11):	+27.46%	$\%\Delta(CSR,NTR)$:	−14.44%	$\%\Delta(CSR15,NTR11)$:	+12.04%
CF OUTPUT-EQUIV. RAISE:	− − −	$R^* = \$2.87$	$R^* = \$0.63$	− − −	− − −	$R^* = \$1.89$
CF MEAN PER-PAGE COST:	− − −	\$0.5384	\$0.4945	− − −	− − −	\$0.5192
CF % Differences:	$\%\Delta(\$15,\$11{+}R^*)$:	+3.12%	$\%\Delta(CSR{+}R^*,NTR)$:	−10.58%	$\%\Delta(CSR15,NTR11{+}R^*)$:	−3.71%

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

To answer these questions, for each cross-group comparison in our study—i.e., \$11 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 suplemental online appendix display simulated CDFs of daily output, per-unit costs, and per-unit accuracy rates.[33] In Table 6, percent changes $\%\Delta(k, j) = 100 \times (outcome_k - outcome_j)/outcome_j$ are reported, with the convention that group j (the second argument) is used as the baseline group.

[33]Note that all daily output means reported here are unconditional. That is, 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 appendix.

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 just over a 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, it is interesting to note that R^* would nearly equalize per-page costs to the firm, relative to high-wage workers under the baseline contract, due to cross-group productivity differences.

The comparison between CSR recruits and NTR recruits is somewhat different. 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 do not 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 (\$2.62) per worker, per day, under the baseline wage contracts (counterfactual wage raise). To put these numbers into context, a constant-returns-to-scale CSR 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 (1,046) workers or more under the status quo contract (counterfactual wage raise), which we refer to as *minimum effective scale*.[34] 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 per-worker profitability of CSR may be higher as well.

6.1.1 Redundancy Simulations

Our discussion in the previous section implicitly 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 many other applications, including janitorial work, data entry, and various service professions, the quality of an individual worker's output may directly impact the firm's cost structure. For example, if

[34] This simple calculation proceeds as follows:

$$minimum\ effective\ scale = \left\lceil \left(\frac{\$1,000,000}{\$3.58 \times 365} \right) \right\rceil = 766.$$

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. If there are non-trivial human resources costs involved in hiring and maintaining one worker (e.g., payroll, management, etc.), then minimum effective scale serves as a lower bound on minimum efficient scale under no redundancy, since CSR workers process fewer unique pages per day, requiring a larger workforce to meet production goals. When redundancy measures are required for quality control (discussed in the following section), the reverse is true: CSR workers produce more effective output per day, so minimum effective scale is an upper bound on minimum efficient scale.

a janitorial worker fails to adequately clean an area of a building, the firm may need to dispatch a second employee to redo the job. Or, 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 data-entry application, this concept is naturally represented: our Likert scale variables contain some element of ambiguity, requiring nontrivial cognitive effort to process training materials and thoughtfully evaluate each field for each Google Street View image. Following industry standard practice, we have adopted the modal response across K independent evaluations as our criterion for determining final output.

	Recruitment Group:	***$15***	***$11***	***CSR***	***NTR***	***CSR15***	***NTR11***
Mode Accuracy Rate $=\frac{2}{3}$	**REQUIRED REDUNDANCY:**	$K^*=5$	$K^*=6$	$K^*=5$	$K^*=6$	$K^*=5$	$K^*=7$
	ADJ. MEAN DAILY OUTPUT:	10.54	6.35	8.97	8.03	11.23	6.49
	% Differences:	%Δ($15,$11):	+66.03%	%Δ(CSR,NTR):	+11.73%	%Δ($CSR15,NTR11$):	+73.04%
	ADJ. MEAN PER-UNIT COST:	$2.7762	$2.6138	$2.3659	$3.3180	$2.4998	$3.1237
	% Differences:	%Δ($15,$11):	+6.22%	%Δ(CSR,NTR):	−28.70%	%Δ($CSR15,NTR11$):	−19.97%
	CF OUTPUT-EQUIV. RAISE:	— — —	$R^*=\$5.03$	— — —	$R^*=\$0.93$	— — —	$R^*=\$6.52$
	CF MEAN PER-UNIT COST:	— — —	$3.7020	— — —	$3.5208	— — —	$4.8905
	CF % Differences:	%Δ($15,$11+R^*):	−25.01%	%Δ($CSR,NTR+R^*$):	−32.80%	%Δ($CSR15,NTR11+R^*$):	−48.88%
Mode Accuracy Rate $=\frac{4}{5}$	**REQUIRED REDUNDANCY:**	$K^*=8$	$K^*=10$	$K^*=8$	$K^*=9$	$K^*=9$	$K^*=13$
	ADJ. MEAN DAILY OUTPUT:	6.59	3.81	5.60	5.35	6.24	3.50
	% Differences:	%Δ($15,$11):	+72.94%	%Δ(CSR,NTR):	+4.75%	%Δ($CSR15,NTR11$):	+78.54%
	ADJ. MEAN PER-UNIT COST:	$4.4420	$4.3563	$3.7854	$4.9770	$4.4996	$5.8012
	% Differences:	%Δ($15,$11):	+1.97%	%Δ(CSR,NTR):	−23.94%	%Δ($CSR15,NTR11$):	−22.44%
	CF OUTPUT-EQUIV. RAISE:	— — —	$R^*=\$5.33$	— — —	$R^*=\$0.14$	— — —	$R^*=\$6.89$
	CF MEAN PER-UNIT COST:	— — —	$6.2766	— — —	$5.0143	— — —	$9.2791
	CF % Differences:	%Δ($15,$11+R^*):	−29.23%	%Δ($CSR,NTR+R^*$):	−24.51%	%Δ($CSR15,NTR11+R^*$):	−51.51%
Mode Accuracy Rate $=\frac{9}{10}$	**REQUIRED REDUNDANCY:**	$K^*=13$	$K^*=16$	$K^*=13$	$K^*=15$	$K^*=14$	$K^*=22$
	ADJ. MEAN DAILY OUTPUT:	4.06	2.38	3.45	3.21	4.01	2.07
	% Differences:	%Δ($15,$11):	+70.28%	%Δ(CSR,NTR):	+7.43%	%Δ($CSR15,NTR11$):	+94.23%
	ADJ. MEAN PER-UNIT COST:	$7.2182	$6.9700	$6.1513	$8.2951	$6.9994	$9.8174
	% Differences:	%Δ($15,$11):	+3.56%	%Δ(CSR,NTR):	−25.84%	%Δ($CSR15,NTR11$):	−28.70%
	CF OUTPUT-EQUIV. RAISE:	— — —	$R^*=\$5.23$	— — —	$R^*=\$0.62$	— — —	$R^*=\$8.47$
	CF MEAN PER-UNIT COST:	— — —	$9.9963	— — —	$8.6129	— — —	$17.0441
	CF % Differences:	%Δ($15,$11+R^*):	−27.79%	%Δ($CSR,NTR+R^*$):	−28.58%	%Δ($CSR15,NTR11+R^*$):	−58.93%

Table 7: Redundancy Simulations

Table 7 summarizes the results of our redundancy simulations. We began by defining *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 one and three extra independent evaluations to ensure a fixed mode accuracy, relative to $15 recruits. For NTR recruits, the requirement is one or two extra evaluations relative to CSR recruits. This comparison is even more striking for the combined selection effect: among NTR11 recruits, the firm requires between two and seven extra independent evaluations to maintain the same mode accuracy as CSR15 recruits.

Naturally, these extra redundancy requirements shift the cost structure substantially 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, on average, each day. Since one employee hired at a $15 wage offer can on average accomplish what 1.66 employees (or more) hired at an $11 wage offer can, with nearly the same per-unit costs, if fixed costs involved in hiring workers are moderate or high, then it would follow that 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 from these numbers is that, if work quality has a direct influence on the firm's cost structure, it may be more advantageous to recruit a higher-productivity, higher-quality workforce with high initial wage contracts, than to accept lower-productivity, lower-quality workers in exchange for promising lower wages.

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 also expands the scope for CSR expenses to be profitable: under the status quo contracts (counterfactual wage raises) a firm can profitably operate a CSR program on a budget of between $6.67–$8.54 ($6.88–$10.36) per worker, per day. This brings the viable firm size for a profitable $1,000,000 annual CSR budget down considerably (see Table 8), with a conservative estimate being 411 workers or more.

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

Table 8: Minimum Effective Scale 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 have been undermining the basis of a free society.” Today, over 90% of major businesses have specific programs dedicated to corporate social responsibility, and many CEOs trumpet their organization's commitment to social engagement. Critics view such facts as strong evidence that Friedman was wrong. But was he? A narrow interpretation of Friedman certainly has proven incorrect, as the business world has evolved in a much more social manner than Friedman likely anticipated. This does not, however, mean that CSR is necessarily at odds with maximizing profits.

This study takes a systematic approach to exploring the supply-side effects of CSR. A great deal of previous empirical scholarly exercises have 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 its labor supply and subsequent effort levels and productivities of workers. In this way, 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. For instance, advertising the firm's CSR endeavors during recruiting increases the application probability by nearly the same amount 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. Conditional on being employed by the firm, a work-stage internal CSR advertisement campaign increases productivity of all workers by speeding up production outside of break time, and reducing workers' propensity to take elective breaks between productive tasks. Importantly, there is no meaningful concomitant decrease in product quality.

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 to its potential work force during the recruitment process. In particular, 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: as Wren (2005) points out, the earliest known instances of CSR date back to the 19^{th}-century industrial welfare movement which aimed to improve labor conditions for workers. 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. Indeed, sprinkled throughout nearly every aspect of 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 in uence. *Working paper, University of British Columbia Department of Economics*.

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. (2017). 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 working paper, 22043.

D'Haultfoeuille, X. and Février, P. (2011). The provision of wage incentives: A structural estimation using contracts variation. *CREST Working Paper Series*, 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. *The 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. (2016). Identification and estimation of a bidding model for electronic auctions. *Quantitative Economics, forthcoming*.

Huang, J. Z. (2003). Local asymptotics for polynomial spline regression. *Annals of Statistics*, 31(5):1600–1635.

Kolmogorov, A. (1933). Sulla determinazione empirica di une 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. *Working paper, 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. *Working paper, University of British Columbia Economics Department*.

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. *Matematiceskii 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*. Cambrige: MIT Press.

Wren, D. A. (2005). *The history of management thought*. John Wiley & Sons.

A Appendix: Additional Figures and Tables

A.1 Estimated Learning Curves

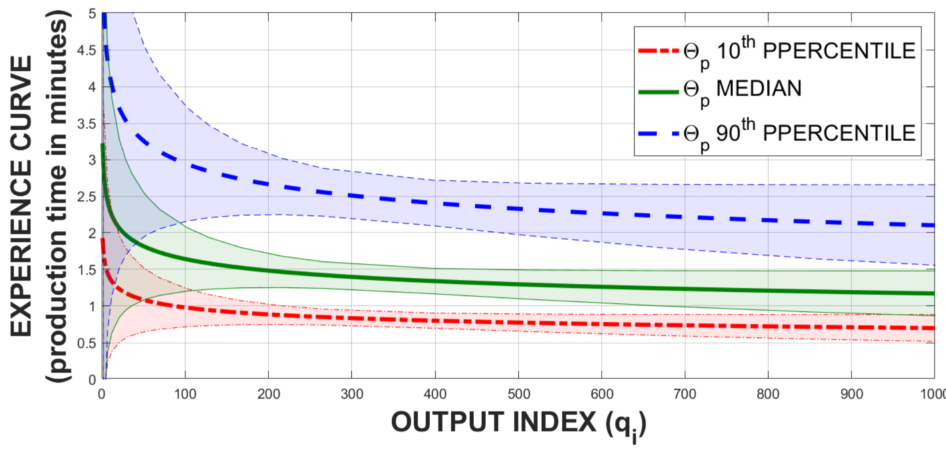

Figure 12: Experience Curves with 90% Confidence Bounds

A.2 Model Fit

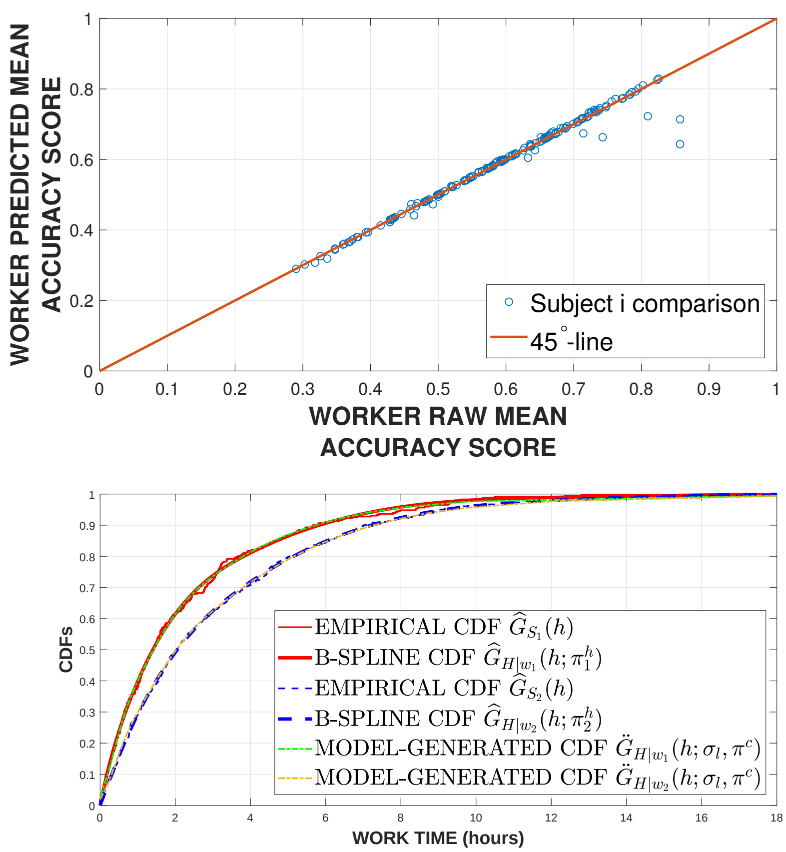

Figure 13: Model Fit: Accuracy and Labor Supply

A.3 Common Labor Supply Cost Function

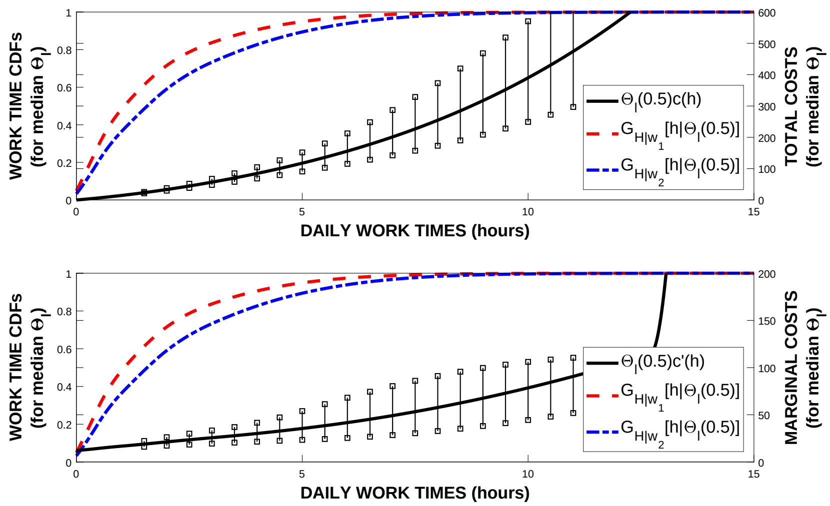

Figure 14: Labor Supply Costs and Marginal Costs (w/90% confidence bounds)

A.4 Stochastic Dominance Tests: Summary Table

TEST	HYPOTHESIS	BOOTSTRAP P-VALUE	RESULT	TEST	HYPOTHESIS	BOOTSTRAP P-VALUE	RESULT
Θ_p Selection on CSR FOSD	$H_0 : G_{p\|CSR} = G_{p\|Ntr}$	—	*reject*	**Θ_d Selection on CSR FOSD**	$H_0 : G_{a\|CSR} = G_{a\|Ntr}$	—	*reject*
	$H_1 : G_{p\|CSR} > G_{p\|Ntr}$	$< 1 \times 10^{-5}$	**adopt**		$H_1 : G_{d\|CSR} > G_{d\|Ntr}$	$< 1 \times 10^{-5}$	*reject*
	$H_2 : G_{p\|CSR} < G_{p\|Ntr}$	0.1561	*reject*		$H_2 : G_{d\|CSR} < G_{d\|Ntr}$	0.0577	*reject*
	$H_3 : G_{p\|CSR} \neq G_{p\|Ntr}$	—	*reject*		$H_3 : G_{d\|CSR} \neq G_{d\|Ntr}$	—	**adopt**
	Mean Production Time Change: −15.14% (26.11% St. Dev.)				**Mean Production Time Change:** +4.02% (6.35% St. Dev.)		
Θ_p Selection on WAGE FOSD	$H_0 : G_{p\|w=\$15} = G_{p\|w=\$11}$	—	*reject*	**Θ_d Selection on WAGE FOSD**	$H_0 : G_{d\|w=\$15} = G_{d\|w=\$11}$	—	*reject*
	$H_1 : G_{p\|w=\$15} > G_{p\|w=\$11}$	$< 1 \times 10^{-5}$	*reject*		$H_1 : G_{d\|w=\$15} > G_{d\|w=\$11}$	$< 1 \times 10^{-5}$	*reject*
	$H_2 : G_{p\|w=\$15} < G_{p\|w=\$11}$	0.1003	*reject*		$H_2 : G_{d\|w=\$15} < G_{d\|w=\$11}$	0.0068	*reject*
	$H_3 : G_{p\|w=\$15} \neq G_{p\|w=\$11}$	—	**adopt**		$H_3 : G_{d\|w=\$15} \neq G_{d\|w=\$11}$	—	**adopt**
	Mean Production Time Change: −3.62% (8.70% St. Dev.)				**Mean Production Time Change:** +5.73% (8.54% St. Dev.)		
	COMBINED EFFECTS						
Θ_p CSR Selection + CSR Treatment FOSD	$H_0 : G_{\Theta_p \mathcal{T}_p\|CSR} = G_{p\|Ntr}$	—	*reject*	**Θ_d CSR Selection + CSR Treatment FOSD**	$H_0 : G_{\Theta_d \mathcal{T}_d\|CSR} = G_{d\|Ntr}$	—	*reject*
	$H_1 : G_{\Theta_p \mathcal{T}_p\|CSR} > G_{p\|Ntr}$	$< 1 \times 10^{-5}$	**adopt**		$H_1 : G_{\Theta_d \mathcal{T}_d\|CSR} < G_{d\|Ntr}$	$< 1 \times 10^{-5}$	**adopt**
	$H_2 : G_{\Theta_p \mathcal{T}_p\|CSR} < G_{p\|Ntr}$	0.7927	*reject*		$H_2 : G_{\Theta_d \mathcal{T}_d\|CSR} > G_{d\|Ntr}$	0.9376	*reject*
	$H_3 : G_{\Theta_p \mathcal{T}_p\|CSR} \neq G_{p\|Ntr}$	—	*reject*		$H_3 : G_{\Theta_d \mathcal{T}_d\|CSR} \neq G_{d\|Ntr}$	—	*reject*
	Mean Production Time Change: −38.49% (66.35% St. Dev.)				**Mean Production Time Change:** −48.07% (75.99% St. Dev.)		
Θ_p CSR Selection + Wage Selection FOSD	$H_0 : G_{p\|CSR,w=\$15} = G_{p\|Ntr,w=\$11}$	—	*reject*	**Θ_d CSR Selection + Wage Selection FOSD**	$H_0 : G_{a\|CSR,w=\$15} = G_{a\|Ntr,w=\$11}$	—	*reject*
	$H_1 : G_{p\|CSR,w=\$15} > G_{p\|Ntr,w=\$11}$	$< 1 \times 10^{-5}$	*reject*		$H_1 : G_{d\|CSR,w=\$15} < G_{d\|Ntr,w=\$11}$	$< 1 \times 10^{-5}$	*reject*
	$H_2 : G_{p\|CSR,w=\$15} < G_{p\|Ntr,w=\$11}$	0.0376	*reject*		$H_2 : G_{d\|CSR,w=\$15} > G_{d\|Ntr,w=\$11}$	0.0068	*reject*
	$H_3 : G_{p\|CSR,w=\$15} \neq G_{p\|Ntr,w=\$11}$	—	**adopt**		$H_3 : G_{d\|CSR,w=\$15} \neq G_{d\|Ntr,w=\$11}$	—	**adopt**
	Mean Production Time Change: −15.28% (54.79% St. Dev.)				**Mean Production Time Change:** +13.32% (21.32% St. Dev.)		
Θ_a Selection on CSR FOSD	$H_0 : G_{a\|CSR} = G_{a\|Ntr}$	—	*reject*	**Θ_l Selection on CSR FOSD**	$H_0 : G_{l\|CSR} = G_{l\|Ntr}$	—	*reject*
	$H_1 : G_{a\|CSR} > G_{a\|Ntr}$	0.0087	**adopt**		$H_1 : G_{l\|CSR} > G_{l\|Ntr}$	$< 1 \times 10^{-5}$	*reject*
	$H_2 : G_{a\|CSR} < G_{a\|Ntr}$	0.1870	*reject*		$H_2 : G_{l\|CSR} < G_{l\|Ntr}$	0.0365	*reject*
	$H_3 : G_{a\|CSR} \neq G_{a\|Ntr}$	—	*reject*		$H_3 : G_{l\|CSR} \neq G_{l\|Ntr}$	—	**adopt**
	Mean Accuracy Change: +3.75% (16.50% St. Dev.)				**Mean Time Value Change:** +2.68% (4.21% St. Dev.)		
Θ_a Selection on WAGE FOSD	$H_0 : G_{a\|w=\$15} = G_{a\|w=\$11}$	—	*reject*	**Θ_l Selection on WAGE FOSD**	$H_0 : G_{l\|w=\$15} = G_{l\|w=\$11}$	—	*reject*
	$H_1 : G_{a\|w=\$15} > G_{a\|w=\$11}$	0.0035	**adopt**		$H_1 : G_{l\|w=\$15} > G_{l\|w=\$11}$	$< 1 \times 10^{-5}$	**adopt**
	$H_2 : G_{a\|w=\$15} < G_{a\|w=\$11}$	0.2790	*reject*		$H_2 : G_{l\|w=\$15} < G_{l\|w=\$11}$	0.3458	*reject*
	$H_3 : G_{a\|w=\$15} \neq G_{a\|w=\$11}$	—	*reject*		$H_3 : G_{l\|w=\$15} \neq G_{l\|w=\$11}$	—	*reject*
	Mean Accuracy Change: +13.23% (51.83% St. Dev.)				**Mean Time Value Change:** +18.13% (33.83% St. Dev.)		
	COMBINED EFFECTS						
Θ_a CSR Selection + CSR Treatment FOSD	$H_0 : G_{\Theta_a + \mathcal{T}_a\|CSR} = G_{a\|Ntr}$	—	*reject*	**Θ_l CSR Selection + CSR Treatment FOSD**	$H_0 : G_{\Theta_l \mathcal{T}_l\|CSR} = G_{l\|Ntr}$	—	*reject*
	$H_1 : G_{\Theta_a + \mathcal{T}_a\|CSR} > G_{a\|Ntr}$	0.0087	**adopt**		$H_1 : G_{\Theta_l \mathcal{T}_l\|CSR} < G_{l\|Ntr}$	$< 1 \times 10^{-5}$	*reject*
	$H_2 : G_{\Theta_a + \mathcal{T}_a\|CSR} < G_{a\|Ntr}$	0.1481	*reject*		$H_2 : G_{\Theta_l \mathcal{T}_l\|CSR} > G_{l\|Ntr}$	0.0677	*reject*
	$H_3 : G_{\Theta_a + \mathcal{T}_a\|CSR} \neq G_{a\|Ntr}$	—	*reject*		$H_3 : G_{\Theta_l \mathcal{T}_l\|CSR} \neq G_{l\|Ntr}$	—	**adopt**
	Mean Accuracy Change: +2.77% (12.19% St. Dev.)				**Mean Time Value Change:** +4.60% (7.24% St. Dev.)		
Θ_a CSR Selection + Wage Selection FOSD	$H_0 : G_{a\|CSR,w=\$15} = G_{a\|Ntr,w=\$11}$	—	*reject*	**Θ_l CSR Selection + Wage Selection FOSD**	$H_0 : G_{l\|CSR,w=\$15} = G_{l\|Ntr,w=\$11}$	—	*reject*
	$H_1 : G_{a\|CSR,w=\$15} > G_{a\|Ntr,w=\$11}$	$< 1 \times 10^{-5}$	**adopt**		$H_1 : G_{l\|CSR,w=\$15} < G_{l\|Ntr,w=\$11}$	$< 1 \times 10^{-5}$	**adopt**
	$H_2 : G_{a\|CSR,w=\$15} < G_{a\|Ntr,w=\$11}$	0.3911	*reject*		$H_2 : G_{l\|CSR,w=\$15} > G_{l\|Ntr,w=\$11}$	0.5111	*reject*
	$H_3 : G_{a\|CSR,w=\$15} \neq G_{a\|Ntr,w=\$11}$	—	*reject*		$H_3 : G_{l\|CSR,w=\$15} \neq G_{l\|Ntr,w=\$11}$	—	*reject*
	Mean Accuracy Change: +20.61% (78.52% St. Dev.)				**Mean Time Value Change:** +29.04% (46.99% St. Dev.)		

Table 9: Stochastic Dominance Tests for Worker Characteristics

A.4.1 Correlations Among Worker Characteristics

We conclude our discussion on unobserved worker characteristics with an examination of the correlation between productivity, work quality, and the value of one's leisure time. Table 10 contains a summary of the estimated pairwise linear correlations between $\log(\Theta_p)$, $\log(\Theta_d)$, Θ_a, and $\log(\Theta_l)$. In order to aid interpretation of the magnitudes in the table, we also report two additional measures of the relationships between worker characteristics. Since Pearson's correlation measures linear co-movement between two random variables, we also computed Spearman's rank correlation coefficient, which is essentially Pearson's correlation measure applied to the quantile ranks, or $Corr\left[G_j(\Theta_j), G_k(\Theta_k)\right]$.[35] This alternative measure allows for non-linear co-movement between Θ_j and Θ_k.

We also compute several measures relating to the predictive power of productivity and work quality on the value of one's time. Specifically, we first report the combined R_l^2 of a weighted regression of the following form:

$$\begin{aligned} log\left(\Theta_{li}\right) =& \pi_0 + \pi_1 log\left(\Theta_{pi}\right) + \pi_2 log\left(\Theta_{pi}\right)^2 + \pi_3 log\left(\Theta_{di}\right) + \pi_4 log\left(\Theta_{di}\right)^2 + \pi_5\Theta_{ai} + \pi_6\Theta_{ai}^2 \\ &+ \pi_7 log\left(\Theta_{pi}\right) log\left(\Theta_{di}\right) + \pi_8 log\left(\Theta_{pi}\right)\Theta_{ai} + \pi_9 log\left(\Theta_{di}\right)\Theta_{ai} + \nu_i. \end{aligned} \tag{18}$$

In addition, we report individual partial-$R_{l,j}^2$ measures, $j = p,\ d,\ a$, using only the linear, quadratic, and intercept terms in the j^{th} control, one at a time. Finally, we report partial-$R_{l,pda}^2$ which is the goodness of fit measure from a regression of $log\left(\Theta_{li}\right)$ on only the interaction terms from (18) and an intercept. When interpreting the numbers in the table, it is important to keep in mind that, due to first-stage sample variation—that is, due to the fact that we must estimate correlations and predictive power using *estimated* worker types $\widehat{\Theta}_{ji},\ j = p, d, a, l$, rather than a sample of direct observations of worker types $\Theta_{ji},\ j = p, d, a, l$—we must confront a measurement error problem which attenuates correlations and R^2 measures toward zero. Thus, one can think of the figures in the table as providing an approximate lower bound on the strengths of the various relationships they represent.

	PEARSON'S LINEAR CORRELATIONS					**SPEARMAN'S RANK CORRELATIONS**				**PREDICTIVE POWER**	
	$\log(\Theta_p)$	$\log(\Theta_d)$	Θ_a	$\log(\Theta_l)$		Θ_p	Θ_d	Θ_a	Θ_l	Measure	Value
$\log(\Theta_p)$	1	0.5818	-0.1571	-0.4395	Θ_p	1	0.5357	-0.1966	-0.2153	R_l^2 (combined)	0.3407
$\log(\Theta_d)$	—	1	0.0683	0.0223	Θ_d	—	1	0.0767	0.1214	$R_{l,p}^2$ (partial)	0.1494
Θ_a	—	—	1	0.0727	Θ_a	—	—	1	0.0469	$R_{l,d}^2$ (partial)	0.0202
$\log(\Theta_l)$	—	—	—	1	Θ_l	—	—	—	1	$R_{l,a}^2$ (partial)	0.0093
										$R_{l,pda}^2$ (partial)	0.1220

Table 10: Correlations and Predictive Power

The table shows that active productivity is strongly correlated with the other three characteristics, both in levels and in ranks. Recall that one's productivity in the active (passive) sense is *decreasing* in Θ_p (Θ_d), whereas accuracy and the value of time are *increasing* in Θ_a and Θ_l, respectively. Therefore, the signs of the estimated correlations suggest that workers who are more productive in the active sense also tend to be more productive in the passive sense. They also tend to produce higher quality work, and have more highly valued time. The remaining pairwise correlations are smaller and weaker. In terms of predictive power, a quadratic polynomial

[35] Appendix B.4 contains explicit formulae for our correlation measures.

in the productivity and accuracy variables is able to explain 34% of the variation in the value of one's time.[36] The partial R^2 numbers indicate that active productivity is the most important single variable for predicting the value of worker time, but that interactions among the productivity and work quality characteristics play an important role as well.

B Estimation Technical Details

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.[37] We chose gender as an illustrative example partly because it was one of few demographic variables that were easily verifiable, and partly 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 straightforwardly 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 $j = p, d, l$, β_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, β_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.

B.1 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:

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

[36] Adding cubic terms only increases R^2_l by about 2%.

[37] The interested reader is directed to a supplemental online appendix D where we present model treatment effects allowing for heterogeneity by gender. There our overall results are similar, but we 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 substantially less productive than their female counterparts at baseline. One possible explanation for this is that male workers are further from their production possibility frontier at baseline. Productivity treatment effects are advantageous to the firm and statistically significant for both genders, but under treatment the gender productivity gap (with males lagging behind) is no longer statistically significant.

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 δ:

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

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 $Q_{Ci} \equiv \sum_{q_i=1}^{Q_i}\mathbb{1}(X_{1q_i} = 1)$ and $Q_{Ni} \equiv \sum_{q_i=1}^{Q_i}\mathbb{1}(X_{1q_i} = 0)$ 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

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

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

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

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:

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

where we use the standard inverse variance sampling weights given by[38]

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

Next we estimate the baseline initial production time τ_1 and fixed effects $\{\Theta_{pi}\}_{i=1}^{I}$. If we let $Z_{q_i}^{p} \equiv \ln(\tau_{q_i}) + \hat{\delta}\ln(q_i)\mathbb{1}(q_i \geq 2) - X_{1q_i}\left[\ln\left(\hat{\beta}_p\right) + \ln\left(\hat{\beta}_{pf}\right)F_i\right]$, then since $E[\varepsilon_{q_i}^{p}] = 0$ we have :

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

and

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

[38]Here we use the standard pooled variance formula to calculate $\widehat{var}\left(\overline{Y}_{Ci}^{p} - \overline{Y}_{Ni}^{p}\right)$ in the sampling weights.

B.2 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 β_a and β_{af}. We then back out individual fixed effects Θ_{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

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

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

$$\mathrm{E}\left[\Phi^{-1}\left(\overline{Y}^a_{Ci}\right) + BC_{Ci} - \Phi^{-1}\left(\overline{Y}^a_{Ni}\right) - BC_{Ni}\right] = \beta_a + \beta_{af}F_i,$$

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 $\overline{Y}^a_{ji}$. Thus, accuracy treatment effects are estimable via a simple least squares routine:

$$[\hat{\beta}_a, \hat{\beta}_{af}]^\top = \underset{[\beta_a, \beta_{af}]^\top}{\arg\min} \left\{ \sum_{i=1}^{I} \omega^a_i \left(\Phi^{-1}\left[\overline{Y}^a_{Ci}\right] + BC_{Ci} - \Phi^{-1}\left[\overline{Y}^a_{Ni}\right] - BC_{Ni} - \beta_a - \beta_{af}F_i\right)^2 \times \mathbb{1}\left[(Q_{Ci} > 0) \ \& \ (Q_{Ni} > 0)\right] \right\}, \tag{24}$$

where the ω^a_i's are inverse variance sampling weights given by[39]

$$\omega^a_i \equiv \begin{cases} \frac{1}{\widehat{var\left[\Phi^{-1}\left(\overline{Y}^a_{Ci}\right) - \Phi^{-1}\left(\overline{Y}^a_{Ni}\right)\right]}} & \text{if } Q_{Ci} > 0 \ \& \ Q_{Ni} > 0 \\ 0 & \text{otherwise.} \end{cases}$$

As for fixed effects and selections, let $\overline{\Omega}^{ak}_{rji}$ 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

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

---

[39] Here we use the standard pooled variance formula to calculate $\widehat{var\left[\Phi^{-1}\left(\overline{Y}^a_{Ci}\right) - \Phi^{-1}\left(\overline{Y}^a_{Ni}\right)\right]}$ in the sampling weights. Moreover, we calculate the variance of $\Phi^{-1}\left(\overline{Y}^a_{ji}\right)$ using the delta method, the inverse function theorem, and standard properties of the sample mean for Bernoulli random samples: $var\left[\Phi^{-1}\left(\overline{Y}^a_{ji}\right)\right] \approx \left(\frac{1}{\varphi\left(\overline{Y}^a_{ji}\right)^2}\right)\left(\frac{\overline{Y}^a_{ji}(1-\overline{Y}^a_{ji})}{7Q_{ji}}\right), j = C, N.$

where $Q^k_{rji} \equiv \sum_{q_i=1}^{Q_i} \mathbb{1}\left[(F_i = \mathbb{1}[k=f]) \;\&\; (X_{0i} = \mathbb{1}[r=C]) \;\&\; (X_{1q_i} = \mathbb{1}[j=C])\right]$ 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 Θ_{ai} for individual i with gender k and hiring treatment r using a simple weighted average for each $i = 1, \ldots, I$:

$$\hat{\Theta}_{ai} = \left[\frac{Q^k_{rCi}}{Q^k_{rCi}+Q^k_{rNi}} \left(\Phi^{-1}\left[\overline{\Omega}^{ak}_{rCi}\right] + BC^k_{rCi} - \hat{\beta}_a - \hat{\beta}_{af}\mathbb{1}[k=f] \right) + \frac{Q^k_{rNi}}{Q^k_{rCi}+Q^k_{rNi}} \left(\Phi^{-1}\left[\overline{\Omega}^{ak}_{rNi}\right] + BC^k_{rNi} \right) \right]. \tag{25}$$

B.2.1 Bias Correction

A problem that remains for the above estimator is the fact that the empirical accuracy rates (the $\overline{Y}$'s and $\overline{\Omega}$'s) above are subject to finite-sample variability, and $\mathrm{E}\left[\Phi^{-1}(Z)\right] \neq \Phi^{-1}\left(\mathrm{E}[Z]\right)$ for any random variable Z, due to the non-linearity in Φ. However, one can use the known functional form of Φ 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 ρ, then the sample mean $\overline{A} = \sum_{n=1}^N a_n/N$ is (approximately) normal with mean ρ and standard deviation $\rho(1-\rho)/\sqrt{N}$. Using this information, one can compute the expectation

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

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

B.3 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 $G_{H|w_1}$, $G_{H|w_2}$ 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. (2016) 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 $[\underline{h},\overline{h}] = [\min_{i,t}(h_{it}), \max_{i,t}(h_{it})]$. 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 $\mathbf{k}_c = \{k_{c1} = \underline{h} < k_{c2} < \cdots < k_{c,K_c+1} = \overline{h}\}$ and for the conditional work time CDFs we choose a partition of K_h subintervals, denoted $\mathbf{k}_h = \{k_{h1} = \underline{h} < k_{h2} < \cdots < k_{h,K_h+1} = \overline{h}\}$. These in combination with the Cox-de Boor recursion formula (see de Boor (2001)) uniquely pin down a set of globally-defined basis functions $\mathcal{B}_{jk} : [\underline{h},\overline{h}] \to \mathbb{R}$, $k = 1,\ldots,K_j+3$, $j = c,h$, which are twice continuously differentiable and are mathematically equivalent to cubic piecewise splines with continuity and differentiability conditions imposed at the knots.[40] The parameterized B-spline functions themselves exist as linear combinations of their respective bases: $\widehat{G}_{H|w_1}(h;\boldsymbol{\pi}_1^h) \equiv \sum_{k=1}^{K_h+3} \pi_{1k}^h \mathcal{B}_{1k}^h(h)$; $\widehat{G}_{H|w_2}(h;\boldsymbol{\pi}_2^h) \equiv \sum_{k=1}^{K_h+3} \pi_{2k}^h \mathcal{B}_{2k}^h(h)$; $\widehat{c}(h;\boldsymbol{\pi}^c) \equiv \sum_{k=1}^{K_c+3} \pi_k^c \mathcal{B}_k^c(h)$.

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 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: $\pi_{jk}^h \leq \pi_{j(k+1)}^h, \quad \forall k = 1,2,...,K_h+2, \;\; j=1,2$ and $\pi_k^c \leq \pi_{(k+1)}^c, \quad \forall k = 1,2,...,K_c+2$. 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 $\pi_{2k}^h \leq \pi_{1k}^h, \quad \forall k = 1,2,...,K_h+3$. 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 then we use these to recover the remaining labor supply parameters which are all nonlinear functions of σ_l and $c(\cdot)$.

B.3.1 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 $\mathcal{S}_2 = \widehat{\widehat{\mathcal{S}}}_{N2} \cup \widehat{\widehat{\mathcal{S}}}_{C2}$, where

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

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

[40] 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.

jk), and

$$\widehat{G}_{\overline{H}^{jk}}(x) = \sum_{i=1}^{I} \frac{\mathbb{1}\left(\overline{H}_i^{jk} \le x\right)}{\sum_{i=1}^{I} \mathbb{1}\left(X_{0i} = \mathbb{1}(j = C)\ \&\ W_i = w_k\right)}$$

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 $\mathcal{S}_1 = \{h_{11}, h_{12}, \ldots, h_{1S_1}\}$ and $\mathcal{S}_2 = \{h_{21}, h_{22}, \ldots, h_{2S_2}\}$, 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 $\widetilde{\Theta}_{js} = \Theta_{li} \mathcal{T}_l^{X_{1it}} u_{it}^l$ 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 $\mathcal{S}_2$, $\widehat{G}_{\mathcal{S}_j}(x) = \sum_{s=1}^{S_j} \mathbb{1}\left(h_{js} \le x\right) / S_j$.[41] This idea implies the following estimator for the B-spline CDF parameters:

$$\begin{aligned}
\begin{bmatrix} \hat{\boldsymbol{\pi}}_1^h \\ \hat{\boldsymbol{\pi}}_2^h \end{bmatrix} \equiv \arg\min & \left\{ \sum_{h_{1s} \in \mathcal{S}_1} \left[\widehat{G}_{H|w_1}(h_{1s}; \boldsymbol{\pi}_1^h) - \widehat{G}_{\mathcal{S}_1}(h_{1s})\right]^2 + \sum_{h_{2s} \in \mathcal{S}_2} \left[\widehat{G}_{H|w_2}(h_{2s}; \boldsymbol{\pi}_2^h) - \widehat{G}_{\mathcal{S}_2}(h_{2s})\right]^2 \right\} \\
& \textit{subject to}: \\
& \pi_{jk}^h \le \pi_{j(k+1)}^h, \quad \forall k = 1, 2, ..., K_h + 2, \;\; j = 1, 2 \;\; (\textit{monotonicity}) \\
& \pi_{2k}^h \le \pi_{1k}^h, \quad \forall k = 1, 2, ..., K_h + 3 \;\; (\textit{first-order dominance}) \\
& \pi_{11}^h = \pi_{21}^h = 0, \;\; (\textit{initial conditions}) \quad \text{and} \\
& \pi_{1,K_h+3}^h = \pi_{2,K_h+3}^h = 1 \;\; (\textit{terminal conditions}).
\end{aligned} \tag{27}$$

B.3.2 Stage II Estimation

With the work time CDF parameters in hand, the second stage estimator uses this information to recover the cost function parameters $\boldsymbol{\pi}^c$, the labor-supply shock variance σ_l, the treatment parameters $\boldsymbol{\beta}_l = [\beta_l, \beta_{fl}]^\top$, and the worker-specific supply cost types $\boldsymbol{\Theta}_l = [\Theta_{l1}, \ldots, \Theta_{lI}]^\top$. To do so, we construct what we refer to as a "counterfactual estimator" which derives directly from equations (11) – (16) above.

To begin, let $\widetilde{\boldsymbol{\Theta}}_l = [\Theta_{l1}, \ldots, \Theta_{lI-q}]^\top$ denote the subset of test subjects whose work-time data survived the trimming rule in equation (26), and assume without loss of 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 $\ddot{G}_{H|w_j}(h; \sigma_l, \boldsymbol{\pi}^c, \boldsymbol{\beta}_l, \boldsymbol{\Theta}_l)$, $j = 1, 2$, and we construct them through the following process:

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

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

[41] 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 cannot directly carry the empirical CDFs through without smoothing them using the B-spline forms.

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

$$\ddot{G}_{H_i|w_j}(h_i;\sigma_l,\boldsymbol{\pi}^c,\boldsymbol{\beta}_l,\Theta_{li}) = \frac{\ddot{G}_{H_i|w_j,X_1=1}(h_i|X_1=1;\sigma_l,\boldsymbol{\pi}^c,\boldsymbol{\beta}_l,\Theta_{li}) + G_{H_i|w_j,X_1=0}(h_i|X_1=0;\sigma_l,\boldsymbol{\pi}^c,\boldsymbol{\beta}_l,\Theta_{li})}{2}, \quad j=1,2. \tag{29}$$

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:

$$\begin{aligned}
\ddot{G}_{H|w_1}(h;\sigma_l,\boldsymbol{\pi}^c,\boldsymbol{\beta}_l,\widetilde{\boldsymbol{\Theta}}_l) &= \frac{\sum_{i=1}^{I-q} G_{H_i|w_1}(h;\sigma_l,\boldsymbol{\pi}^c,\boldsymbol{\beta}_l,\Theta_{li})}{I} \\
\ddot{G}_{H|w_2}(h;\sigma_l,\boldsymbol{\pi}^c,\boldsymbol{\beta}_l,\widetilde{\boldsymbol{\Theta}}_l) &= \frac{\sum_{i=1}^{I-q} G_{H_i|w_2}(h;\sigma_l,\boldsymbol{\pi}^c,\boldsymbol{\beta}_l,\Theta_{li})}{I}.
\end{aligned} \tag{30}$$

Wit that definition, the last thing we must do is define a grid of L domain points, $\{h_{c1}, h_{c2}, \ldots, h_{cL}\}$, 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 (12) and to optimize fit between the empirical and model-generated work-time CDFs:

$$\begin{aligned}
\begin{bmatrix} \hat{\sigma}_l \\ \hat{\boldsymbol{\pi}}^c \\ \hat{\boldsymbol{\beta}}_l \\ \widehat{\boldsymbol{\Theta}}_l \end{bmatrix} \equiv \underset{\mathbb{R}^{3+L+K_c+3}_{++}}{\arg\min} \Bigg\{ & \sum_{h_s \in \mathcal{S}_1 \cup \mathcal{S}_2} \left[\widehat{G}_{H|w_1}(h_s;\hat{\boldsymbol{\pi}}^h_1) - \ddot{G}_{H|w_1}(h_s;\sigma_l,\boldsymbol{\pi}^c,\boldsymbol{\beta}_l,\widetilde{\boldsymbol{\Theta}}_l) \right]^2 \\
& + \sum_{h_s \in \mathcal{S}_1 \cup \mathcal{S}_2} \left[\widehat{G}_{H|w_2}(h_s;\hat{\boldsymbol{\pi}}^h_2) - \ddot{G}_{H|w_2}(h_s;\sigma_l,\boldsymbol{\pi}^c,\boldsymbol{\beta}_l,\widetilde{\boldsymbol{\Theta}}_l) \right]^2 \\
& + \sum_{i=1}^{I} \sum_{j=1}^{5} \left(\mathbb{E}_{(j:5)}(\sigma_l) - \ln\left(\frac{W_i}{c'\left(H_{Ci}(j:5);\boldsymbol{\pi}^c\right)} \right) + \ln\left(\Theta_{li}\right) + \ln\left(\mathcal{T}_l\right) \right)^2 \mathbb{1}\left(H_{Ci}(j:5) > 0\right) \\
& + \sum_{i=1}^{I} \sum_{j=1}^{5} \left(\mathbb{E}_{(j:5)}(\sigma_l) - \ln\left(\frac{W_i}{c'\left(H_{Ni}(j:5);\boldsymbol{\pi}^c\right)} \right) + \ln\left(\Theta_{li}\right) \right)^2 \mathbb{1}\left(H_{Ni}(j:5) > 0\right) \Bigg\} \\
& \textit{subject to}: \\
& \pi^c_k \leq \pi^c_{(k+1)}, \quad \forall k = 1,2,...,K_c+2, \ \ (\textit{monotonicity}) \\
& \hat{c}''(h_j;\boldsymbol{\pi}^c) > 0, \quad \forall j = 1,2,...,J \ \ (\textit{convexity}) \\
& \hat{c}'(0;\boldsymbol{\pi}^c) = 1 \ \ (\textit{boundary condition}) \\
& \Theta_{li} > 0, \ i = 1,\ldots,I; \ \ \beta_l > 0; \ \ \beta_{lf} > 0; \ \ \sigma_l > 0 \ \ (\textit{positivity}).
\end{aligned} \tag{31}$$

In the objective function above, note that only test subjects who survived the trimming rule (26) enter into the CDF fitting criterion (first two lines), whereas all test subjects' data contribute to the moment conditions of shock order statistics (third and fourth lines). Intuitively, this is because matching the shape of the positive work time CDFs for the trimmed sample of individuals primarily pins down σ_l and the common cost function $\hat{c}(\cdot; \boldsymbol{\pi}^c)$, but if these objects are known, then all workers' cost types Θ_{li} are identified and estimable by matching the moments of the shock order statistics in equation (12).

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 σ_l and cost function $\boldsymbol{\pi}^c$ during runtime, and then for each guess at these values, solving for the treatment parameters $\boldsymbol{\beta}_l$ and fixed effects $\boldsymbol{\Theta}_l$ as functions of the first two using equations (12). 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. These benefits are due to the fact that the solver need only search through a $(1 + K_c + 3)$-dimensional space, and the shape constraints on the cost function now only indirectly interact with optimal choice of the treatment and fixed effect parameters. Econometrically, the simultaneous and sequential approaches are the same, as they both optimize an identical set of empirical moment conditions.

B.3.3 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 $G_{H|w_1}(h; \boldsymbol{\pi}_1^h)$, $G_{H|w_2}(h; \boldsymbol{\pi}_2^h)$, and the cost function, $c(h; \boldsymbol{\pi}^c)$. 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 points 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 $100 \times \frac{j}{K+1}^{\text{th}}$ empirical quantiles, for each $j = 0, 1, 2, \ldots, K + 1$. 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 $K_h = 5$—that is, knots chosen 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 $\boldsymbol{\pi}_1^h$ and $\boldsymbol{\pi}_2^h$ (eight parameters total, with two boundary conditions), and nine free parameters in the vector $\boldsymbol{\pi}^c$ (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.

B.4 Type Distributions

The final step in our empirical implementation is to obtain estimates of the distributions of worker characteristics G_p, G_a, G_l. The main difficulty here is that the sample of observations to be used for this part $\{\widehat{\Theta}_{pi}, \widehat{\Theta}_{ai}, \widehat{\Theta}_{li}\}_{i=1}^{I}$ are estimates themselves. In particular, due to the unbalanced nature of our panel data, some fixed effect estimates are more precisely known than others. For that reason, we construct weighted empirical CDF estimators $\widehat{G}_p(x) = \sum_{i=1}^{I} \eta_i^p \mathbb{1}(\widehat{\Theta}_{pi} \leq x)$, $\widehat{G}_a(x) = \sum_{i=1}^{I} \eta_i^a \mathbb{1}(\widehat{\Theta}_{ai} \leq x)$, and $\widehat{G}_l(x) = \sum_{i=1}^{I} \eta_i^l \mathbb{1}(\widehat{\Theta}_{li} \leq x)$, using sample weights given by

$$\eta_i^p = \eta_i^a \equiv \frac{Q_{Ci} + Q_{Ni}}{\sum_{j=1}^{I} Q_{Cj} + Q_{Nj}}, \text{ and } \eta_i^l \equiv \frac{\sum_{t=1}^{T} \mathbb{1}(H_i t > 0)}{\sum_{j=1}^{I} \sum_{t=1}^{T} \mathbb{1}(H_j t > 0)}. \tag{32}$$

We also wish to learn about the correlation structure in the joint distribution of worker characteristics, G_{pal}. Using the weights defined above we can also define weighted Pearson's linear correlation coefficient estimators as follows:

$$\begin{aligned} \widehat{Corr}[\log(\Theta_p), \Theta_a] &= \frac{\widehat{Cov}[\log(\Theta_p), \Theta_a]}{\sqrt{\widehat{Var}[\log(\Theta_p)]\widehat{Var}[\Theta_a]}}, \\ \widehat{Corr}[\log(\Theta_p), \log(\Theta_l)] &= \frac{\widehat{Cov}[\log(\Theta_p), \log(\Theta_l)]}{\sqrt{\widehat{Var}[\log(\Theta_p)]\widehat{Var}[\log(\Theta_l)]}}, \text{ and} \\ \widehat{Corr}[\Theta_a, \log(\Theta_l)] &= \frac{\widehat{Cov}[\Theta_a, \log(\Theta_l)]}{\sqrt{\widehat{Var}[\Theta_a]\widehat{Var}[\log(\Theta_l)]}}, \end{aligned} \tag{33}$$

where

$$\begin{aligned} \widehat{Cov}[\log(\Theta_p), \Theta_a] &= \sum_{i=1}^{I} \frac{\eta_i^p + \eta_i^a}{2} \left(\log(\widehat{\Theta}_p) - \widehat{E}[\log(\Theta_p)]\right) \left(\widehat{\Theta}_a - \widehat{E}[\Theta_a]\right), \\ \widehat{Cov}[\log(\Theta_p), \log(\Theta_l)] &= \sum_{i=1}^{I} \frac{\eta_i^p + \eta_i^l}{2} \left(\log(\widehat{\Theta}_p) - \widehat{E}[\log(\Theta_p)]\right) \left(\log(\widehat{\Theta}_l) - \widehat{E}[\log(\Theta_l)]\right), \\ \widehat{Cov}[\Theta_a, \log(\Theta_l)] &= \sum_{i=1}^{I} \frac{\eta_i^l + \eta_i^a}{2} \left(\log(\widehat{\Theta}_l) - \widehat{E}[\log(\Theta_l)]\right) \left(\widehat{\Theta}_a - \widehat{E}[\Theta_a]\right), \end{aligned} \tag{34}$$

and where the $\widehat{E}[\cdot]$'s and $\widehat{Var}[\cdot]$'s are the familiar weighted expectation and variance estimators.[42] Alternatively, we may wish to examine rank correlations among the various worker characteristics, using Spearman's rank correlation coefficient. This we do in the same way as above, except that in equations (33) –(34) we compute weighted correlations of (weighted) quantile ranks $\left(\widehat{G}_p\left(\widehat{\Theta}_{pi}\right), \widehat{G}_d\left(\widehat{\Theta}_{di}\right), \widehat{G}_a\left(\widehat{\Theta}_{ai}\right), \widehat{G}_l\left(\widehat{\Theta}_{li}\right)\right)$ instead.

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.

[42] Because the correlation coefficient seeks to measure the strength of a linear relationship between two random variables, we take the logs of productivity and leisure preferences in order to put them into a linear scale first.

B.5 Asymptotic Theory and Inference

One nice feature of our estimator is that it falls within the class of GMM estimators, since each parameter is essentially chosen to match specific empirical moments from the data. Therefore, established GMM asymptotic theory applies. 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 therefore the basis functions as well) will be held fixed—as the sample size grows. Under this view, the same panel-data asymptotic theory once again applies. Alternatively, one could think of them as *sieve type estimators*, meaning that as the sample size grows, the researcher will gradually add additional knots until in the limit as the sample size becomes infinite the knot vectors will become a dense set on the relevant domains and the B-spline functions will therefore become arbitrarily flexible everywhere. This view complicates matters somewhat, but a recent econometric literature on sieve estimators has established asymptotic normality of the model parameters and pointwise asymptotic normality of the sieve functionals (see Chen (2007) and Huang (2003)).

The main caveat to recognize, given the unbalanced panel structure of the data, is which dimension governs the asymptotics of a given parameter estimate. In particular, parameter estimates that are common to all subjects—e.g., treatment effects, δ, τ_1, work-time CDFs, and the common cost function—will be consistent at a rate of either $\sqrt{\sum_{i=1}^{I} Q_i}$ or $\sqrt{I \times T}$, 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., Θ_{pi}, Θ_{ai}, and Θ_{li}—are consistent at a rate of either $\sqrt{Q_i}$ or $\sqrt{T}$. Thus, although increasing the width of the panel (I) will aid finite-sample precision of the type distribution estimates, the only way to make them arbitrarily precise is to increase the length of the panel or number of worker days (T) without bound.

ONLINE SUPPLEMENTAL APPENDIX TO ACCOMPANY:
Toward an Understanding of Corporate Social Responsibility: Theory and Field Experimental Evidence,
by Daniel Hedblom, Brent R. Hickman, and John A. List

C Hiring and Work Stage Additional Experimental Details

In this section we provide additional details of the experimental design to illustrate various aspects of the test subject experience during our study.

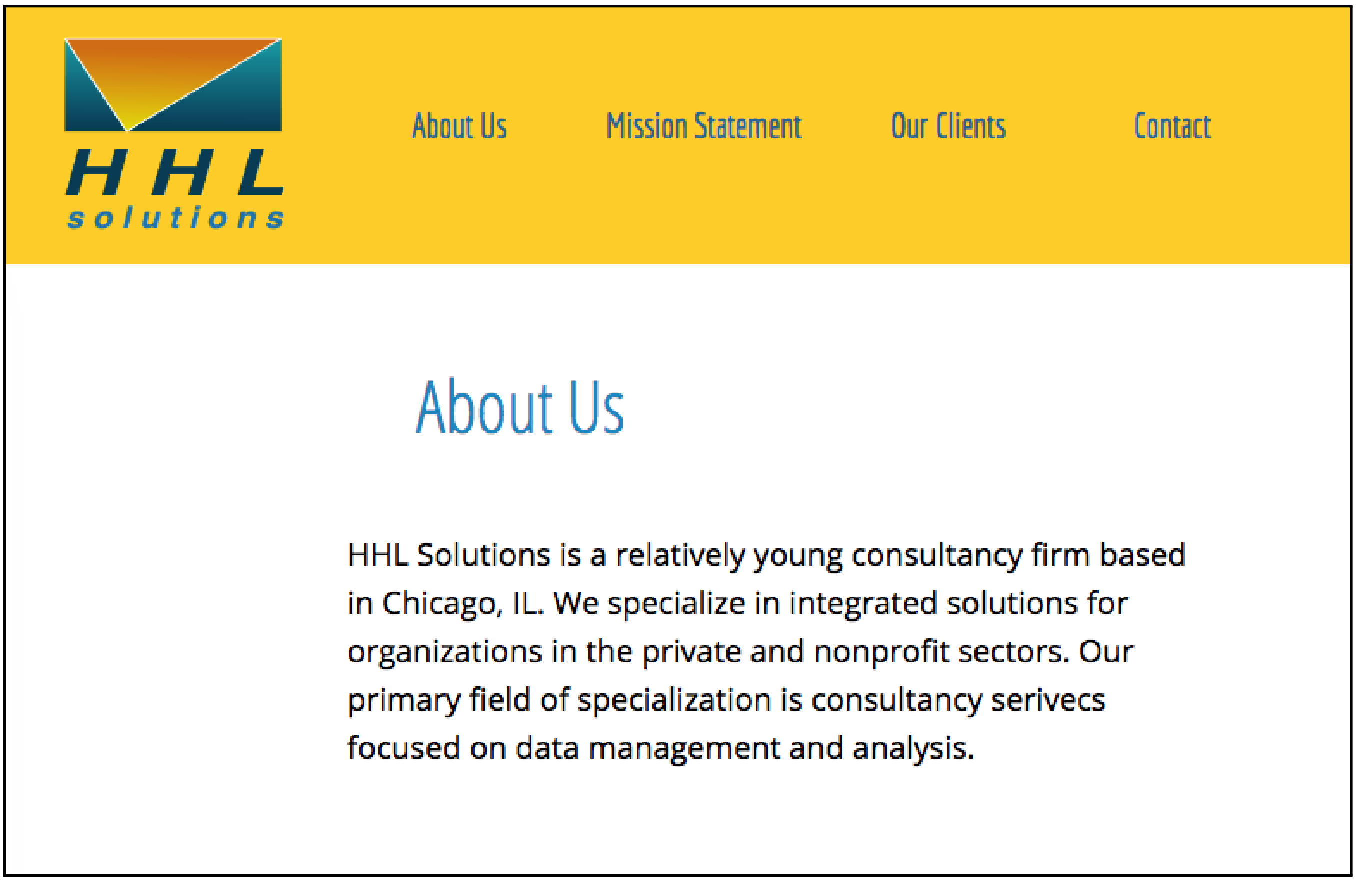

Figure 15: Screenshot of firm website

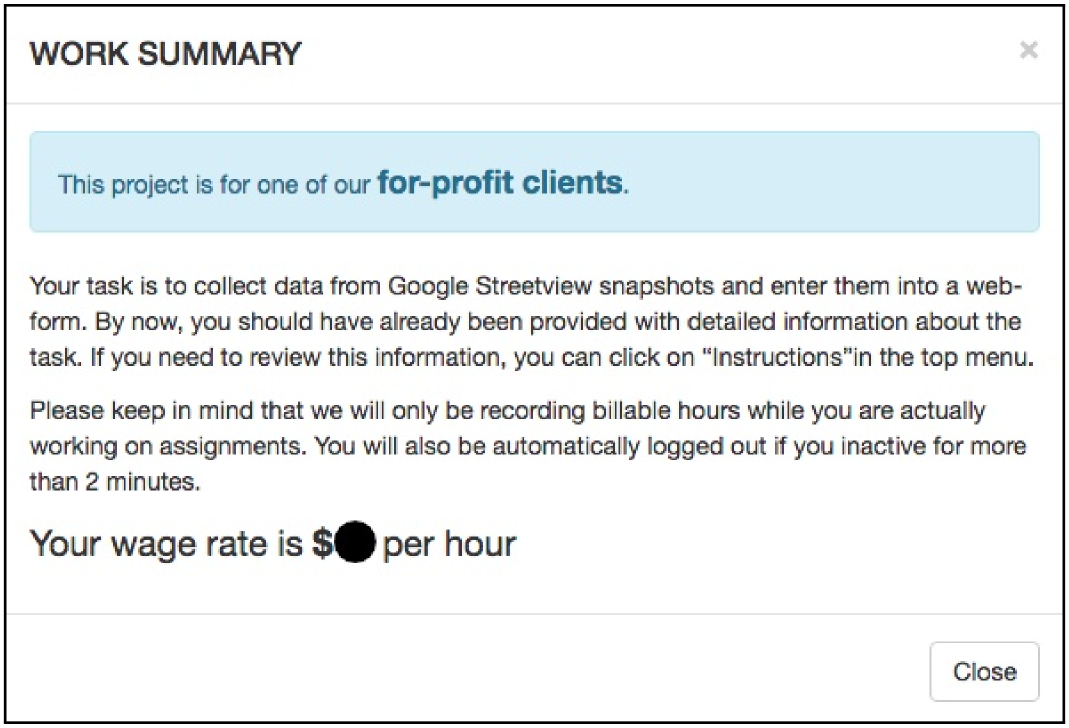

Figure 16: Example login information: for-profit client

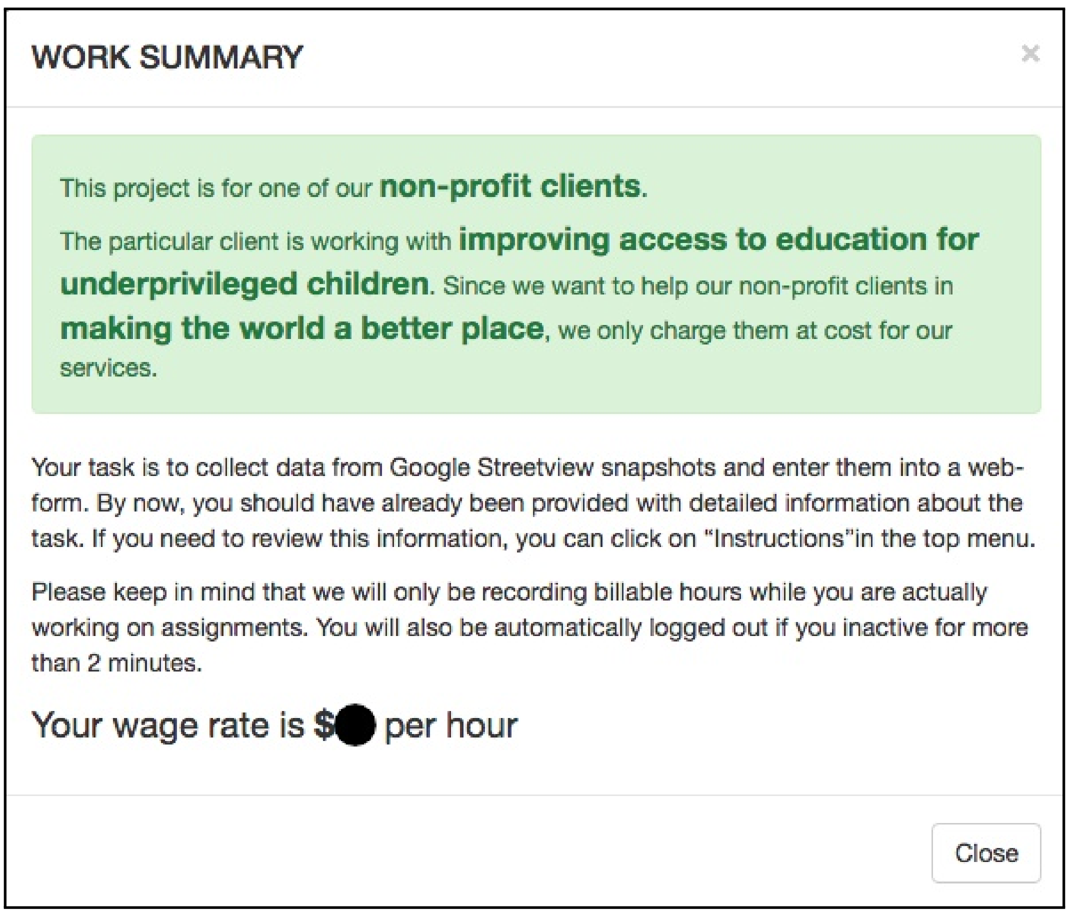

Figure 17: Example login-information: CSR client

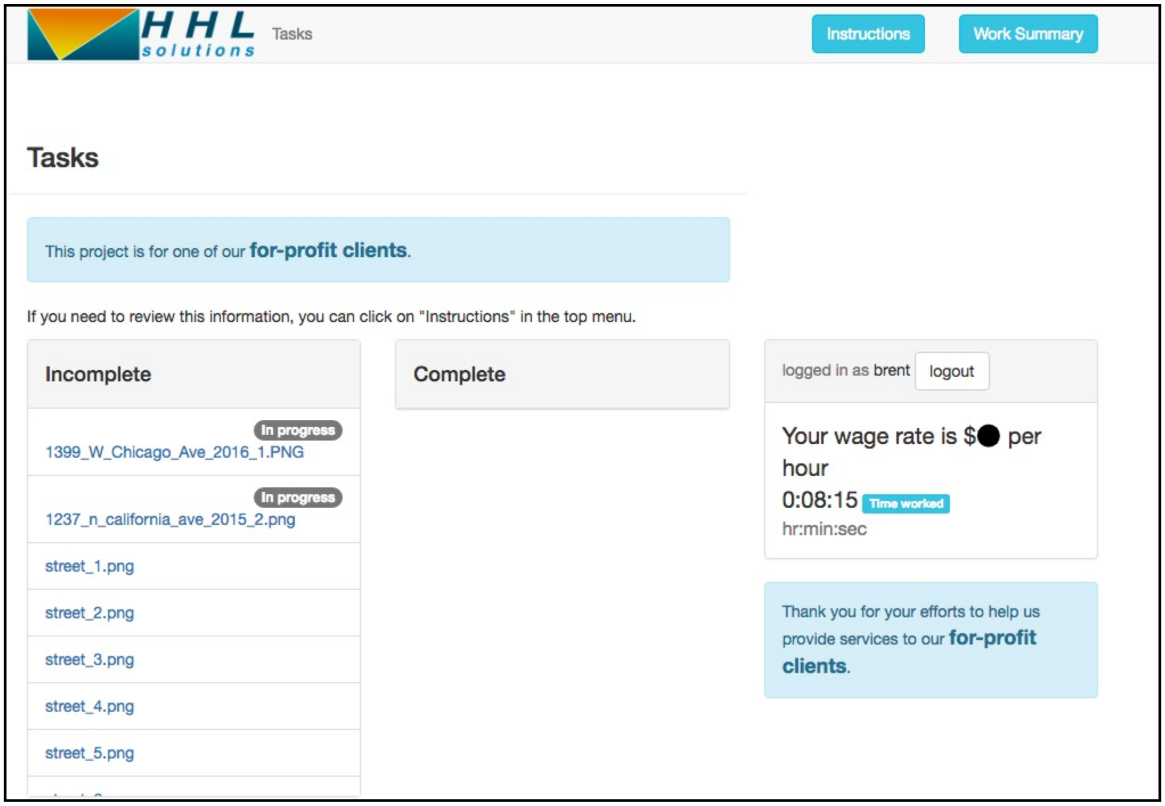

Figure 18: Example main screen: for-profit client

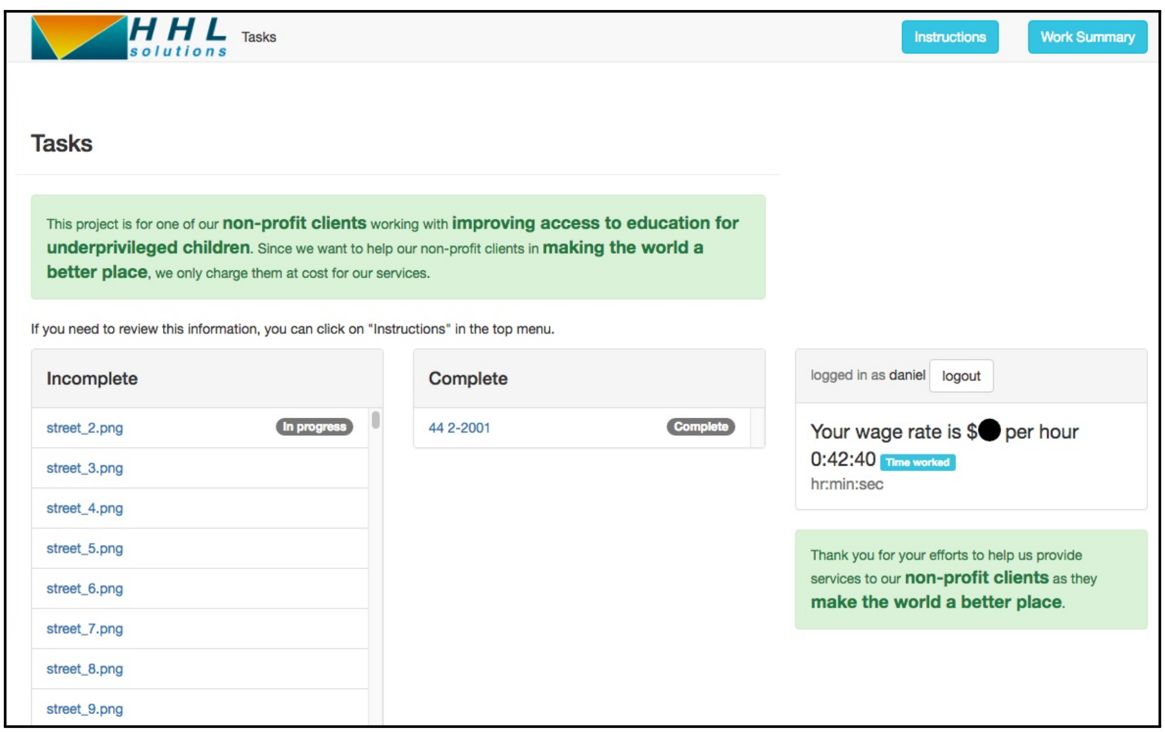

Figure 19: Example main screen: CSR client

Figure 20: Example data entry screen (top half)

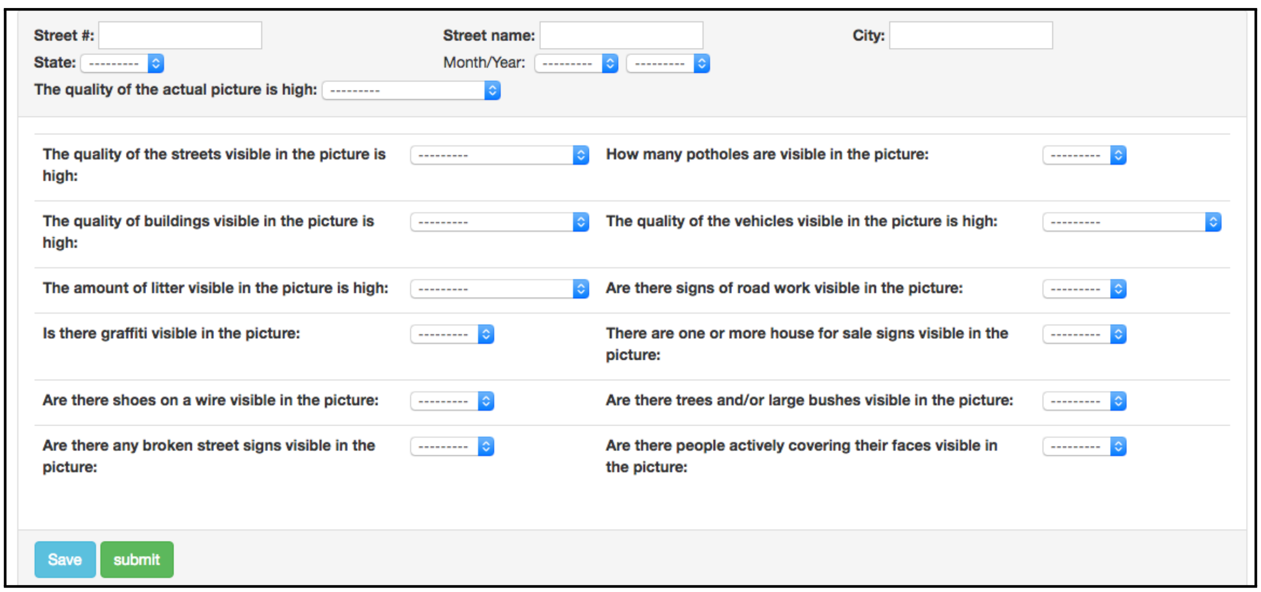

Figure 21: Example data entry screen (bottom half)

Variable Name	Response Type	Mean Response	StDev	Mean Correct	#Obs 62,138
v_1 : *Road Work Visible*	Binary (0,1,N/A)	0.07	0.29	0.96	62,138
v_2 : *Graffiti Visible*	Binary (0,1,N/A)	0.05	0.25	0.96	62,138
v_3 : *Trees/Shrubs Visible*	Binary (0,1,N/A)	0.91	0.29	0.93	62,138
v_4 : *For-Sale Signs Visible*	Binary (0,1,N/A)	0.08	0.37	0.96	62,138
v_5 : *Broken Street Signs Visible*	Binary (0,1,N/A)	0.03	0.23	0.98	62,138
v_6 : *People Covering Faces*	Binary (0,1,N/A)	0.16	0.53	0.92	62,138
v_7 : *Shoes hanging from wires*	Binary (0,1,N/A)	0.04	0.26	0.98	62,138
v_8: *Street Number*	Integer (open)	1,663	2,188	0.95	62,138
v_9 : *Month*	Categorical (drop-down)	7.31	2.26	0.99	62,138
v_{10} : *Year*	Categorical (drop-down)	2,012	3.02	0.99	62,138
v_{11} : *City*	String (open)	–	–	0.99	62,138
v_{12} : *State*	Categorical (drop-down)	–	–	0.99	62,138
v_{13} : *Building Quality*	Likert scale (1-5)	3.64	1.15	0.48	62,138
v_{14} : *Quality of Visible Cars*	Likert scale (1-5)	3.73	1.15	0.50	62,138
v_{15} : *Litter*	Likert scale (1-5)	2.37	1.57	0.47	62,138
v_{16} : *Picture Quality*	Likert scale (1-5)	3.83	1.10	0.55	62,138
v_{17} : *Street Quality*	Likert scale (1-5)	3.60	1.15	0.48	62,138
v_{18} *Number of Visible Potholes*	Integer (drop-down)	0.52	1.19	0.76	62,138
v_{19} : *Street Name*	String (open)	–	–	0.85	62,138
Accuracy Index: (across worker-image pairs)	Derived $(\min A_{q_i}, \max A_{q_i}) = (0,1)$	0.586	0.226	–	62,138

Table 11: Summary of Data Entry Variables

D Supplemental Tables and Figures

D.1 Model Simulations and Counterfactual Results:

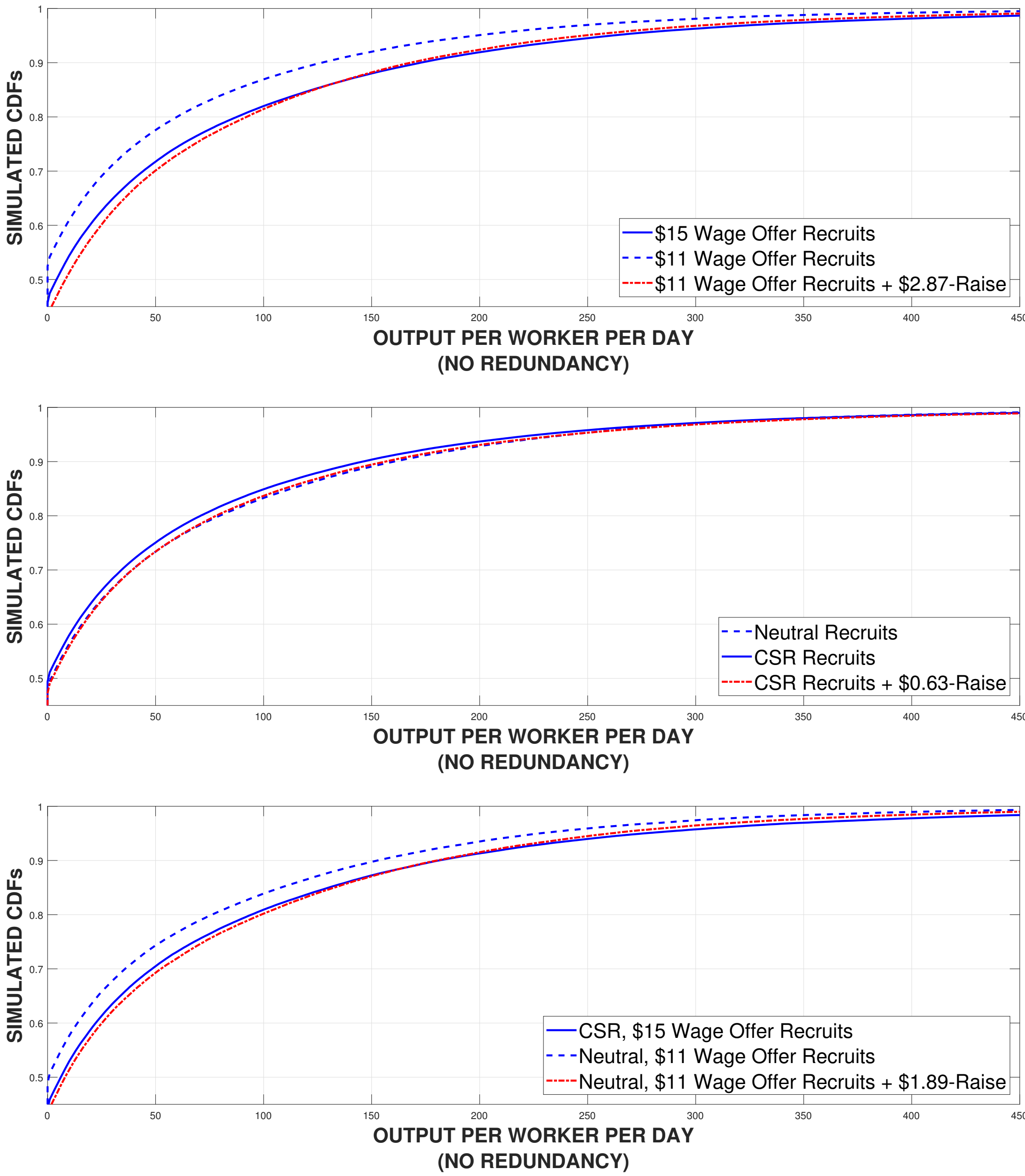

Figure 22: Simulated Per-Worker Daily Output Supply

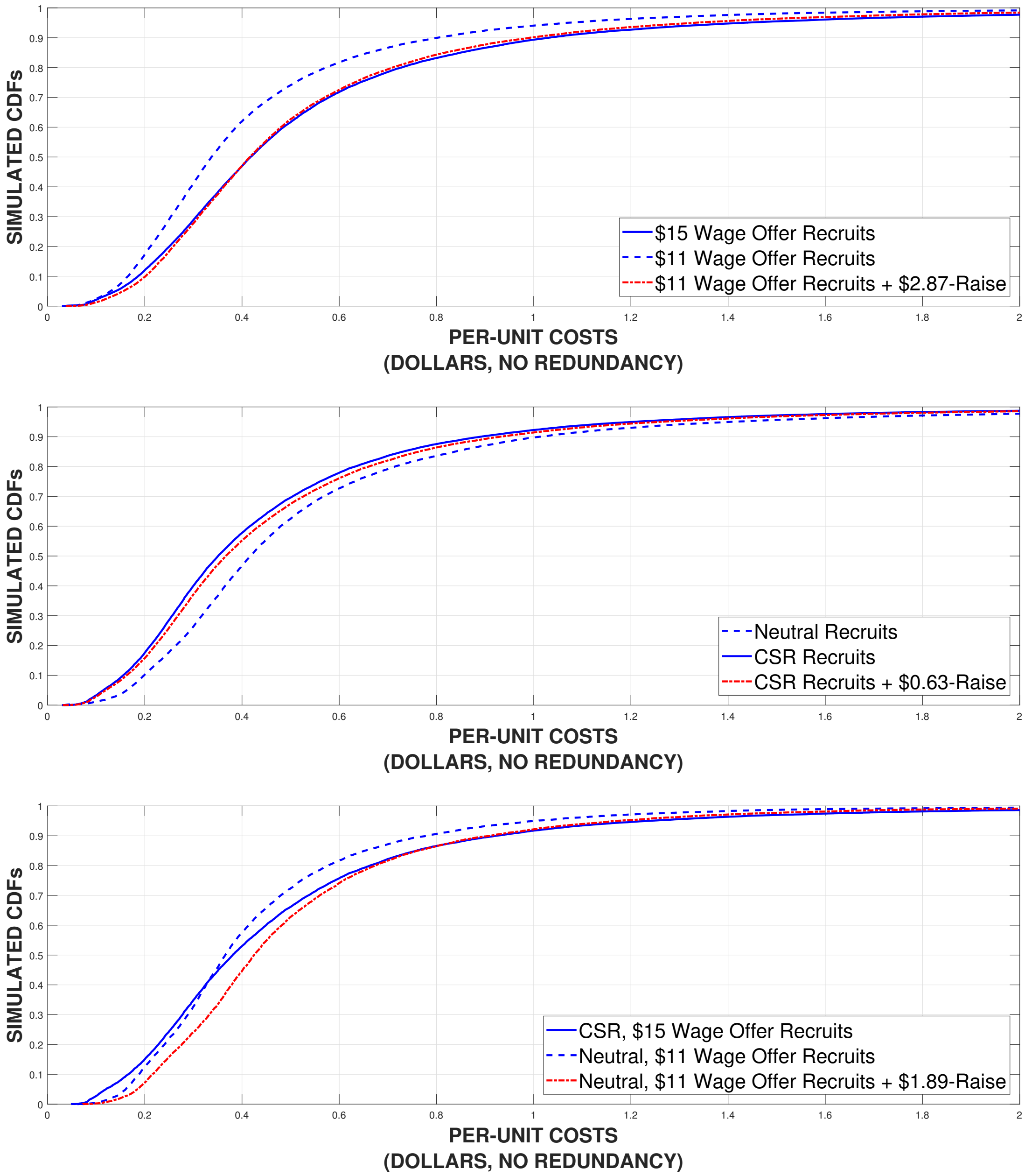

Figure 23: Simulated Per-Unit Production Costs

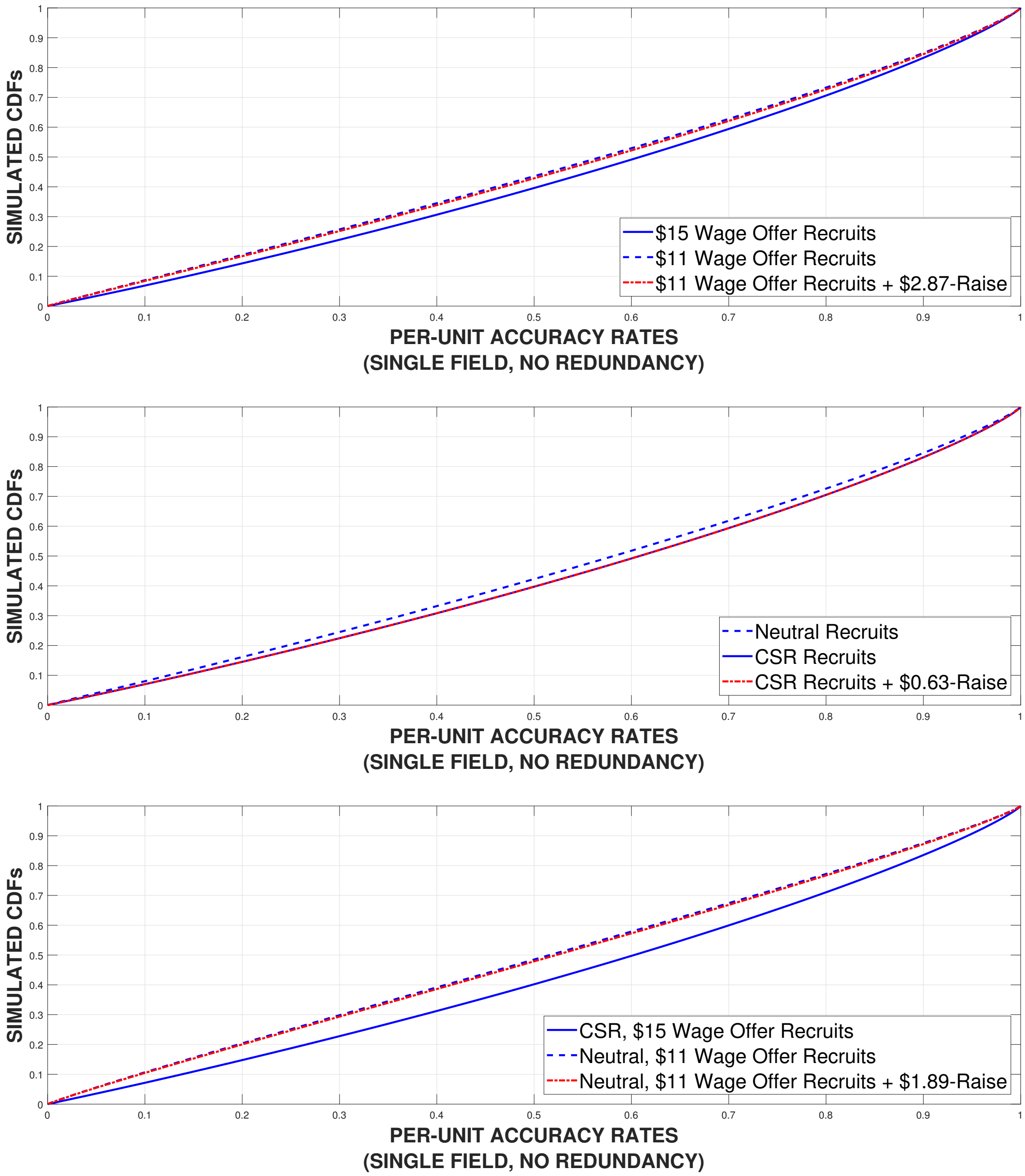

Figure 24: Simulated Per-Unit, Single-Worker Accuracy Rates

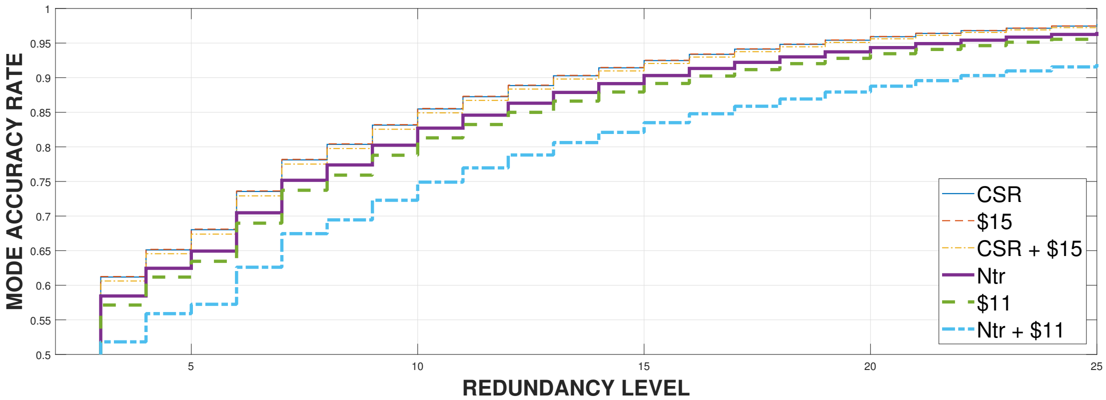

Figure 25: Redundancy Simulations

D.2 Point-Specific P-Values for Stochastic Dominance Tests

D.3 Heterogeneous Treatment Effects By Gender

Recall that for productivity measures, the treatment parameters are directly interpretable as a scaling factor for mean production time. For active productivity, our estimates show that in response to our CSR treatment, male workers reduce on-task production times by roughly 51.1%, while females reduce on-task production times by about 11.4%. In terms of passive productivity, the estimated effects are large as well: male workers reduce their consumption of paid, off-task time by 71.3%, while females reduce it by 28.7%. One possible reason for the treatment effects being larger for males than for females is the fact that our male workers were less productive than their female counterparts at baseline. In particular, mean on-task production times for males (in absence of work-stage treatment) were 44.7% higher than for their female counterparts, and mean off-task times per unit of output were 89.7% higher. A two-sided bootstrap test reveals that these productivity differences by gender are statistically significant at the 10% level at baseline. However, a similar test in the presence of CSR work-stage treatment results in no statistically significant productivity difference (active or passive) between male and female workers.[43]

	Parameter	*Estimate*	*P-value*	*Std Error*	*90% CI*
ACTIVE PRODUCTIVITY:					
Treatment, Male:	β_p	0.4894	1.16×10^-5	0.0842	$[0, 0.608]$
			$(H_0 : \beta_p = 1)$		(one-sided)
Female Treatment Differential:	β_{pf}	1.8102	0.0019	0.3440	$[1.321, 2.480]$
			$(H_0 : \beta_{pf} = 1)$		
Treatment, Female:	$\beta_p \times \beta_{pf}$	0.8860	0.0799	0.0771	$[0, 0.989]$
			$(H_0 : \beta_p \times \beta_{pf} = 1)$		(one-sided)
PASSIVE PRODUCTIVITY:					
Treatment, Male:	β_d	0.2875	0.0026	0.1340	$[0, 0.510]$
			$(H_0 : \beta_d = 1)$		(one-sided)
Female Treatment Differential:	β_{df}	2.4787	0.0679	1.2093	$[1.109, 5.616]$
			$(H_0 : \beta_{df} = 1)$		
Treatment, Female:	$\beta_d \times \beta_{df}$	0.7127	0.0571	0.1565	$[0, 0.938]$
			$(H_0 : \beta_d \times \beta_{df} = 1)$		(one-sided)
ACCURACY/WORK QUALITY:					
Treatment, Male:	β_a	-0.0124	0.1742	0.0091	$[-0.027, 0.003]$
			$(H_0 : \beta_a = 0)$		
Female Treatment Differential:	β_{af}	-0.0045	0.6774	0.0108	$[-0.022, 0.013]$
			$(H_0 : \beta_{af} = 0)$		
Treatment, Female:	$\beta_a + \beta_{af}$	-0.0168	0.0031	0.0057	$[-0.026, -0.008]$
			$(H_0 : \beta_a + \beta_{af} = 0)$		
LABOR SUPPLY COSTS:					
Treatment, Male:	β_l	0.9438	0.1582	0.0398	$[0.878, 1.009]$
			$(H_0 : \beta_l = 1)$		
Female Treatment Differential:	β_{lf}	1.1199	0.1807	0.0896	$[0.9726, 1.2673]$
			$(H_0 : \beta_{lf} = 1)$		
Treatment, Female:	$\beta_l \times \beta_{lf}$	1.0570	0.3578	0.062	$[0.955, 1.159]$
			$(H_0 : \beta_l \times \beta_{lf} = 1)$		

Table 20: Parameter estimates: Treatment Effects By Gender.

[43] These productivity differences are measured in terms of workers' estimated fixed effects, which we discus in the body of the paper. The baseline test compares weighted means of $\widehat{\Theta}_{ji}$ within gender group for $j = p, d$, and the second test compares weighted means of $\widehat{\Theta}_{ji}\widehat{\mathcal{T}}_j$ within gender group for $j = p, d$.

Given the large speed-up in production of each unit of output, it may be less obvious *a priori* what the expected sign of the treatment effects for work quality should be. On the one hand, the CSR treatment may inspire workers to exert themselves in producing accurate output. On the other hand, since they are producing units of output more quickly and taking less rest time in between, it is also possible, through burnout or mental fatigue, that work quality may fall. In the third section of Table 5 we see that the point estimates are all negative. Estimated effects for male and female workers are of roughly the same magnitude, but both are economically insignificant. To place the units in context, note that the accuracy treatment effect for females—the one which is statistically different from zero—is only enough to reduce mean predicted accuracy for the average female worker by two thirds of one percentage point. Taking this result in combination with the previous estimates, we find that the work-stage CSR treatment substantially increases productivity in both the active and passive senses, and that this speed-up of production is virtually all valuable to the firm, coming at little or no cost in terms of diminished work quality.

For labor supply, we do not find strong evidence of a work-stage treatment effect. For males we find weak evidence that receiving a CSR treatment at the work stage causes them to increase their within-day labor supply as if their supply costs had dropped by 5.6%. However, for females, the point estimate for work-stage treatment is in the opposite direction but less precisely estimated.

Percentile	Θ_p	**P-val** $H_2 : G_{p\|Ntr} < G_{p\|CSR}$	**P-val** $H_3 : G_{p\|Ntr} > G_{p\|CSR}$	Percentile	Θ_p	**P-val** $H_2 : G_{p\|w=\$11} < G_{p\|w=\$15}$	**P-val** $H_3 : G_{p\|w=\$11} > G_{p\|w=\$15}$
0.01	0.330	0.000	1.000	0.01	0.330	0.000	1.000
0.02	0.399	0.000	1.000	0.02	0.399	0.000	1.000
0.03	0.433	0.822	0.178	0.03	0.433	0.158	0.842
0.04	0.449	0.790	0.210	0.04	0.449	0.218	0.781
0.05	0.462	0.760	0.240	0.05	0.462	0.249	0.751
0.07	0.474	0.712	0.288	0.07	0.474	0.264	0.736
0.08	0.485	0.189	0.811	0.08	0.485	0.028	0.972
0.09	0.495	0.205	0.795	0.09	0.495	0.045	0.955
0.10	0.505	0.220	0.780	0.10	0.505	0.074	0.926
0.11	0.515	0.239	0.761	0.11	0.515	0.114	0.886
0.12	0.524	0.258	0.742	0.12	0.524	0.165	0.835
0.13	0.534	0.630	0.370	0.13	0.534	0.026	0.974
0.14	0.543	0.625	0.375	0.14	0.543	0.045	0.955
0.15	0.553	0.358	0.642	0.15	0.553	0.009	0.991
0.16	0.563	0.399	0.601	0.16	0.563	0.025	0.975
0.18	0.574	0.304	0.696	0.18	0.574	0.020	0.980
0.19	0.585	0.310	0.690	0.19	0.585	0.029	0.971
0.20	0.596	0.314	0.686	0.20	0.596	0.039	0.961
0.21	0.608	0.155	0.845	0.21	0.608	0.017	0.983
0.22	0.621	0.152	0.848	0.22	0.621	0.026	0.974
0.23	0.634	0.158	0.842	0.23	0.634	0.033	0.967
0.24	0.647	0.175	0.825	0.24	0.647	0.035	0.965
0.25	0.661	0.158	0.842	0.25	0.661	0.041	0.959
0.26	0.675	0.163	0.837	0.26	0.675	0.042	0.958
0.27	0.690	0.167	0.833	0.27	0.690	0.041	0.959
0.29	0.705	0.289	0.711	0.29	0.705	0.039	0.961
0.30	0.720	0.306	0.694	0.30	0.720	0.149	0.851
0.31	0.735	0.300	0.700	0.31	0.735	0.149	0.851
0.32	0.750	0.259	0.741	0.32	0.750	0.171	0.829
0.33	0.764	0.233	0.767	0.33	0.764	0.197	0.803
0.34	0.779	0.152	0.848	0.34	0.779	0.130	0.870
0.35	0.793	0.115	0.885	0.35	0.793	0.101	0.899
0.36	0.806	0.116	0.884	0.36	0.806	0.099	0.901
0.37	0.820	0.194	0.806	0.37	0.820	0.075	0.925
0.38	0.833	0.184	0.816	0.38	0.833	0.084	0.916
0.40	0.846	0.167	0.833	0.40	0.846	0.075	0.925
0.41	0.859	0.246	0.754	0.41	0.859	0.101	0.899
0.42	0.872	0.235	0.765	0.42	0.872	0.096	0.904
0.43	0.885	0.234	0.766	0.43	0.885	0.097	0.903
0.44	0.898	0.224	0.776	0.44	0.898	0.095	0.905
0.45	0.911	0.258	0.742	0.45	0.911	0.102	0.898
0.46	0.924	0.257	0.743	0.46	0.924	0.103	0.897
0.47	0.936	0.305	0.695	0.47	0.936	0.087	0.913
0.48	0.949	0.304	0.696	0.48	0.949	0.089	0.911
0.49	0.962	0.341	0.659	0.49	0.962	0.080	0.920
0.51	0.976	0.417	0.583	0.51	0.976	0.146	0.854
0.52	0.989	0.295	0.705	0.52	0.989	0.125	0.875
0.53	1.002	0.336	0.664	0.53	1.002	0.112	0.888
0.54	1.016	0.337	0.663	0.54	1.016	0.105	0.895
0.55	1.030	0.371	0.629	0.55	1.030	0.096	0.904
0.56	1.044	0.342	0.658	0.56	1.044	0.117	0.883
0.57	1.059	0.250	0.750	0.57	1.059	0.124	0.876
0.58	1.073	0.182	0.818	0.58	1.073	0.201	0.799
0.59	1.089	0.178	0.822	0.59	1.089	0.208	0.792
0.60	1.105	0.136	0.864	0.60	1.105	0.256	0.744
0.61	1.121	0.135	0.865	0.61	1.121	0.253	0.747
0.63	1.138	0.068	0.932	0.63	1.138	0.262	0.738
0.64	1.155	0.048	0.952	0.64	1.155	0.232	0.768
0.65	1.173	0.043	0.957	0.65	1.173	0.223	0.777
0.66	1.191	0.065	0.935	0.66	1.191	0.323	0.677
0.67	1.211	0.064	0.936	0.67	1.211	0.328	0.672
0.68	1.231	0.058	0.942	0.68	1.231	0.328	0.672
0.69	1.252	0.110	0.890	0.69	1.252	0.256	0.744
0.70	1.275	0.132	0.868	0.70	1.275	0.309	0.691
0.71	1.298	0.140	0.860	0.71	1.298	0.350	0.650
0.72	1.323	0.138	0.862	0.72	1.323	0.350	0.650
0.74	1.350	0.206	0.794	0.74	1.350	0.483	0.517
0.75	1.378	0.199	0.801	0.75	1.378	0.482	0.518
0.76	1.408	0.180	0.820	0.76	1.408	0.482	0.518
0.77	1.439	0.182	0.818	0.77	1.439	0.489	0.511
0.78	1.473	0.380	0.620	0.78	1.473	0.762	0.238
0.79	1.508	0.530	0.470	0.79	1.508	0.884	0.116
0.80	1.546	0.531	0.469	0.80	1.546	0.897	0.103
0.81	1.585	0.253	0.747	0.81	1.585	0.835	0.165
0.82	1.626	0.323	0.677	0.82	1.626	0.863	0.137
0.83	1.669	0.355	0.645	0.83	1.669	0.873	0.127
0.85	1.714	0.514	0.486	0.85	1.714	0.843	0.157
0.86	1.761	0.515	0.485	0.86	1.761	0.865	0.135
0.87	1.812	0.515	0.485	0.87	1.812	0.883	0.117
0.88	1.865	0.301	0.699	0.88	1.865	0.797	0.203
0.89	1.923	0.320	0.680	0.89	1.923	0.766	0.234
0.90	1.985	0.342	0.658	0.90	1.985	0.812	0.188
0.91	2.053	0.338	0.662	0.91	2.053	0.819	0.181
0.92	2.128	0.333	0.667	0.92	2.128	0.824	0.176
0.93	2.211	0.285	0.715	0.93	2.211	0.802	0.198
0.94	2.306	0.137	0.863	0.94	2.306	0.744	0.256
0.96	2.415	0.122	0.878	0.96	2.415	0.747	0.253
0.97	2.545	0.100	0.900	0.97	2.545	0.757	0.243
0.98	2.704	0.073	0.927	0.98	2.704	0.772	0.228
0.99	2.899	0.369	0.631	0.99	2.899	0.483	0.517

Table 12: Point-Specific P-Values for Θ_p (First-Order Dominance Test)

Percentile	Θ_p	P-val $H_2 : G_{p\|Ntr} < G_{p\|CSR+Trtmt}$	P-val $H_3 : G_{p\|Ntr} > G_{p\|CSR+Trtmt}$	Percentile	Θ_p	P-val $H_2 : G_{p\|Ntr,w=\$11} < G_{p\|CSR,w=\$15}$	P-val $H_3 : G_{p\|Ntr,w=\$11} > G_{p\|CSR,w=\$15}$
0.01	0.330	0.099	0.901	0.01	0.330	0.000	1.000
0.02	0.399	0.078	0.922	0.02	0.399	0.000	1.000
0.03	0.433	0.204	0.796	0.03	0.433	0.000	1.000
0.04	0.449	0.165	0.835	0.04	0.449	0.000	1.000
0.05	0.462	0.176	0.824	0.05	0.462	0.000	1.000
0.07	0.474	0.180	0.820	0.07	0.474	0.000	1.000
0.08	0.485	0.039	0.961	0.08	0.485	0.000	1.000
0.09	0.495	0.045	0.955	0.09	0.495	0.000	1.000
0.10	0.505	0.032	0.968	0.10	0.505	0.000	1.000
0.11	0.515	0.037	0.963	0.11	0.515	0.000	1.000
0.12	0.524	0.042	0.958	0.12	0.524	0.000	1.000
0.13	0.534	0.065	0.935	0.13	0.534	0.000	1.000
0.14	0.543	0.070	0.930	0.14	0.543	0.000	1.000
0.15	0.553	0.057	0.943	0.15	0.553	0.000	1.000
0.16	0.563	0.042	0.958	0.16	0.563	0.041	0.959
0.18	0.574	0.041	0.959	0.18	0.574	0.020	0.980
0.19	0.585	0.037	0.963	0.19	0.585	0.034	0.966
0.20	0.596	0.039	0.961	0.20	0.596	0.050	0.950
0.21	0.608	0.040	0.960	0.21	0.608	0.008	0.992
0.22	0.621	0.040	0.960	0.22	0.621	0.013	0.987
0.23	0.634	0.023	0.977	0.23	0.634	0.018	0.982
0.24	0.647	0.026	0.974	0.24	0.647	0.023	0.977
0.25	0.661	0.025	0.975	0.25	0.661	0.025	0.975
0.26	0.675	0.024	0.976	0.26	0.675	0.029	0.971
0.27	0.690	0.008	0.992	0.27	0.690	0.032	0.968
0.29	0.705	0.013	0.987	0.29	0.705	0.059	0.941
0.30	0.720	0.026	0.974	0.30	0.720	0.147	0.853
0.31	0.735	0.025	0.975	0.31	0.735	0.147	0.853
0.32	0.750	0.021	0.979	0.32	0.750	0.140	0.860
0.33	0.764	0.018	0.982	0.33	0.764	0.138	0.862
0.34	0.779	0.016	0.984	0.34	0.779	0.076	0.924
0.35	0.793	0.016	0.984	0.35	0.793	0.054	0.946
0.36	0.806	0.015	0.985	0.36	0.806	0.055	0.945
0.37	0.820	0.032	0.968	0.37	0.820	0.065	0.935
0.38	0.833	0.032	0.968	0.38	0.833	0.067	0.933
0.40	0.846	0.033	0.967	0.40	0.846	0.060	0.940
0.41	0.859	0.065	0.935	0.41	0.859	0.104	0.896
0.42	0.872	0.044	0.956	0.42	0.872	0.099	0.901
0.43	0.885	0.044	0.956	0.43	0.885	0.099	0.901
0.44	0.898	0.022	0.978	0.44	0.898	0.096	0.904
0.45	0.911	0.037	0.963	0.45	0.911	0.100	0.900
0.46	0.924	0.030	0.970	0.46	0.924	0.100	0.900
0.47	0.936	0.023	0.977	0.47	0.936	0.101	0.899
0.48	0.949	0.013	0.987	0.48	0.949	0.100	0.900
0.49	0.962	0.016	0.984	0.49	0.962	0.101	0.899
0.51	0.976	0.019	0.981	0.51	0.976	0.189	0.811
0.52	0.989	0.018	0.982	0.52	0.989	0.128	0.872
0.53	1.002	0.023	0.977	0.53	1.002	0.128	0.872
0.54	1.016	0.017	0.983	0.54	1.016	0.120	0.880
0.55	1.030	0.021	0.979	0.55	1.030	0.121	0.879
0.56	1.044	0.019	0.981	0.56	1.044	0.121	0.879
0.57	1.059	0.019	0.981	0.57	1.059	0.093	0.907
0.58	1.073	0.017	0.983	0.58	1.073	0.097	0.903
0.59	1.089	0.015	0.985	0.59	1.089	0.100	0.900
0.60	1.105	0.014	0.986	0.60	1.105	0.099	0.901
0.61	1.121	0.008	0.992	0.61	1.121	0.104	0.896
0.63	1.138	0.007	0.993	0.63	1.138	0.079	0.921
0.64	1.155	0.007	0.993	0.64	1.155	0.069	0.931
0.65	1.173	0.006	0.994	0.65	1.173	0.070	0.930
0.66	1.191	0.010	0.990	0.66	1.191	0.127	0.873
0.67	1.211	0.009	0.991	0.67	1.211	0.136	0.864
0.68	1.231	0.009	0.991	0.68	1.231	0.137	0.863
0.69	1.252	0.020	0.980	0.69	1.252	0.141	0.859
0.70	1.275	0.024	0.976	0.70	1.275	0.186	0.814
0.71	1.298	0.027	0.973	0.71	1.298	0.213	0.787
0.72	1.323	0.026	0.974	0.72	1.323	0.209	0.791
0.74	1.350	0.048	0.952	0.74	1.350	0.339	0.661
0.75	1.378	0.018	0.982	0.75	1.378	0.328	0.672
0.76	1.408	0.009	0.991	0.76	1.408	0.296	0.704
0.77	1.439	0.005	0.995	0.77	1.439	0.265	0.735
0.78	1.473	0.011	0.989	0.78	1.473	0.643	0.357
0.79	1.508	0.023	0.977	0.79	1.508	0.814	0.186
0.80	1.546	0.017	0.983	0.80	1.546	0.841	0.159
0.81	1.585	0.012	0.988	0.81	1.585	0.758	0.242
0.82	1.626	0.027	0.973	0.82	1.626	0.798	0.202
0.83	1.669	0.024	0.976	0.83	1.669	0.833	0.167
0.85	1.714	0.053	0.947	0.85	1.714	0.871	0.129
0.86	1.761	0.037	0.963	0.86	1.761	0.902	0.098
0.87	1.812	0.023	0.977	0.87	1.812	0.926	0.074
0.88	1.865	0.016	0.984	0.88	1.865	0.753	0.247
0.89	1.923	0.015	0.985	0.89	1.923	0.734	0.266
0.90	1.985	0.012	0.988	0.90	1.985	0.898	0.102
0.91	2.053	0.000	1.000	0.91	2.053	0.927	0.073
0.92	2.128	0.000	1.000	0.92	2.128	0.958	0.042
0.93	2.211	0.000	1.000	0.93	2.211	0.956	0.044
0.94	2.306	0.000	1.000	0.94	2.306	0.000	1.000
0.96	2.415	0.000	1.000	0.96	2.415	0.000	1.000
0.97	2.545	0.000	1.000	0.97	2.545	0.000	1.000
0.98	2.704	0.000	1.000	0.98	2.704	0.000	1.000
0.99	2.899	0.000	1.000	0.99	2.899	0.000	1.000

Table 13: Point-Specific P-Values for Θ_p (First-Order Dominance Test): Combined Effects

Percentile	Θ_d	P-val $H_2 : G_{d\|Ntr} < G_{d\|CSR}$	P-val $H_3 : G_{d\|Ntr} > G_{d\|CSR}$	Percentile	Θ_d	P-val $H_2 : G_{d\|w=\$11} < G_{d\|w=\$15}$	P-val $H_3 : G_{d\|w=\$11} > G_{d\|w=\$15}$
0.01	0.403	0.000	1.000	0.01	0.403	0.000	1.000
0.02	0.430	0.000	1.000	0.02	0.430	0.708	0.292
0.03	0.450	0.370	0.630	0.03	0.450	0.499	0.501
0.04	0.464	0.705	0.295	0.04	0.464	0.276	0.724
0.05	0.477	0.679	0.321	0.05	0.477	0.304	0.696
0.07	0.489	0.647	0.353	0.07	0.489	0.334	0.666
0.08	0.503	0.522	0.478	0.08	0.503	0.682	0.318
0.09	0.516	0.513	0.487	0.09	0.516	0.649	0.351
0.10	0.530	0.656	0.344	0.10	0.530	0.486	0.514
0.11	0.544	0.176	0.824	0.11	0.544	0.173	0.827
0.12	0.558	0.045	0.955	0.12	0.558	0.080	0.920
0.13	0.571	0.063	0.937	0.13	0.571	0.162	0.838
0.14	0.584	0.096	0.904	0.14	0.584	0.178	0.822
0.15	0.596	0.119	0.881	0.15	0.596	0.195	0.805
0.16	0.607	0.108	0.892	0.16	0.607	0.272	0.728
0.18	0.618	0.123	0.877	0.18	0.618	0.281	0.720
0.19	0.627	0.136	0.864	0.19	0.627	0.286	0.714
0.20	0.636	0.111	0.889	0.20	0.636	0.248	0.752
0.21	0.644	0.187	0.813	0.21	0.644	0.437	0.563
0.22	0.652	0.283	0.717	0.22	0.652	0.586	0.414
0.23	0.659	0.539	0.461	0.23	0.659	0.370	0.630
0.24	0.666	0.538	0.462	0.24	0.666	0.373	0.627
0.25	0.673	0.455	0.545	0.25	0.673	0.318	0.682
0.26	0.680	0.448	0.552	0.26	0.680	0.332	0.668
0.27	0.687	0.449	0.551	0.27	0.687	0.334	0.666
0.29	0.693	0.634	0.366	0.29	0.693	0.648	0.352
0.30	0.700	0.633	0.367	0.30	0.700	0.644	0.356
0.31	0.707	0.633	0.367	0.31	0.707	0.643	0.357
0.32	0.714	0.631	0.369	0.32	0.714	0.639	0.361
0.33	0.720	0.496	0.504	0.33	0.720	0.537	0.463
0.34	0.727	0.478	0.522	0.34	0.727	0.564	0.436
0.35	0.734	0.434	0.566	0.35	0.734	0.536	0.464
0.36	0.741	0.434	0.566	0.36	0.741	0.536	0.464
0.37	0.749	0.254	0.746	0.37	0.749	0.395	0.605
0.38	0.757	0.291	0.709	0.38	0.757	0.463	0.537
0.40	0.765	0.221	0.779	0.40	0.765	0.656	0.344
0.41	0.773	0.219	0.781	0.41	0.773	0.657	0.343
0.42	0.782	0.276	0.724	0.42	0.782	0.750	0.250
0.43	0.792	0.345	0.655	0.43	0.792	0.720	0.280
0.44	0.802	0.397	0.603	0.44	0.802	0.685	0.315
0.45	0.813	0.395	0.605	0.45	0.813	0.686	0.314
0.46	0.825	0.392	0.608	0.46	0.825	0.693	0.307
0.47	0.838	0.390	0.610	0.47	0.838	0.694	0.306
0.48	0.852	0.339	0.661	0.48	0.852	0.815	0.185
0.49	0.867	0.429	0.571	0.49	0.867	0.771	0.229
0.51	0.884	0.422	0.578	0.51	0.884	0.766	0.234
0.52	0.903	0.412	0.588	0.52	0.903	0.740	0.260
0.53	0.923	0.324	0.676	0.53	0.923	0.687	0.313
0.54	0.945	0.475	0.525	0.54	0.945	0.573	0.427
0.55	0.968	0.520	0.480	0.55	0.968	0.539	0.461
0.56	0.991	0.520	0.480	0.56	0.991	0.538	0.462
0.57	1.013	0.547	0.453	0.57	1.013	0.500	0.500
0.58	1.034	0.500	0.500	0.58	1.034	0.506	0.494
0.59	1.053	0.500	0.500	0.59	1.053	0.507	0.493
0.60	1.071	0.512	0.488	0.60	1.071	0.419	0.581
0.61	1.089	0.514	0.486	0.61	1.089	0.416	0.584
0.63	1.106	0.608	0.392	0.63	1.106	0.564	0.436
0.64	1.123	0.417	0.583	0.64	1.123	0.329	0.671
0.65	1.140	0.334	0.666	0.65	1.140	0.267	0.733
0.66	1.157	0.387	0.613	0.66	1.157	0.231	0.769
0.67	1.174	0.386	0.614	0.67	1.174	0.231	0.769
0.68	1.192	0.348	0.652	0.68	1.192	0.273	0.727
0.69	1.210	0.348	0.652	0.69	1.210	0.273	0.727
0.70	1.229	0.344	0.656	0.70	1.229	0.277	0.723
0.71	1.249	0.345	0.655	0.71	1.249	0.281	0.719
0.72	1.269	0.350	0.650	0.72	1.269	0.286	0.714
0.74	1.291	0.356	0.644	0.74	1.291	0.268	0.732
0.75	1.314	0.484	0.516	0.75	1.314	0.198	0.802
0.76	1.340	0.491	0.509	0.76	1.340	0.192	0.808
0.77	1.369	0.407	0.593	0.77	1.369	0.418	0.582
0.78	1.400	0.400	0.600	0.78	1.400	0.428	0.572
0.79	1.437	0.514	0.486	0.79	1.437	0.610	0.390
0.80	1.479	0.567	0.433	0.80	1.479	0.639	0.361
0.81	1.527	0.565	0.435	0.81	1.527	0.643	0.357
0.82	1.582	0.728	0.272	0.82	1.582	0.846	0.154
0.83	1.644	0.742	0.258	0.83	1.644	0.870	0.130
0.85	1.714	0.697	0.303	0.85	1.714	0.871	0.129
0.86	1.797	0.626	0.374	0.86	1.797	0.884	0.116
0.87	1.895	0.646	0.354	0.87	1.895	0.893	0.107
0.88	2.014	0.834	0.166	0.88	2.014	0.808	0.192
0.89	2.155	0.899	0.101	0.89	2.155	0.807	0.193
0.90	2.310	0.874	0.126	0.90	2.310	0.742	0.258
0.91	2.481	0.853	0.147	0.91	2.481	0.523	0.477
0.92	2.672	0.904	0.096	0.92	2.672	0.500	0.500
0.93	2.892	0.786	0.214	0.93	2.892	0.810	0.190
0.94	3.151	0.316	0.684	0.94	3.151	0.681	0.319
0.96	3.473	0.545	0.455	0.96	3.473	0.450	0.550
0.97	3.909	0.579	0.421	0.97	3.909	0.269	0.731
0.98	4.610	0.462	0.538	0.98	4.610	0.318	0.682
0.99	5.840	0.401	0.599	0.99	5.840	0.205	0.795

Table 14: Point-Specific P-Values for Θ_d (First-Order Dominance Test)

Percentile	Θ_d	P-val $H_2 : G_{d\|Ntr} < G_{d\|CSR+Trtmt}$	P-val $H_3 : G_{d\|Ntr} > G_{d\|CSR+Trtmt}$	Percentile	Θ_d	P-val $H_2 : G_{d\|Ntr,w=\$11} < G_{d\|CSR,w=\$15}$	P-val $H_3 : G_{d\|Ntr,w=\$11} > G_{d\|CSR,w=\$15}$
0.01	0.403	0.000	1.000	0.01	0.403	0.000	1.000
0.02	0.430	0.000	1.000	0.02	0.430	0.000	1.000
0.03	0.450	0.000	1.000	0.03	0.450	0.000	1.000
0.04	0.464	0.004	0.996	0.04	0.464	0.000	1.000
0.05	0.477	0.013	0.987	0.05	0.477	0.000	1.000
0.07	0.489	0.007	0.993	0.07	0.489	0.000	1.000
0.08	0.503	0.021	0.979	0.08	0.503	0.000	1.000
0.09	0.516	0.003	0.997	0.09	0.516	0.000	1.000
0.10	0.530	0.004	0.996	0.10	0.530	0.000	1.000
0.11	0.544	0.000	1.000	0.11	0.544	0.000	1.000
0.12	0.558	0.001	0.999	0.12	0.558	0.000	1.000
0.13	0.571	0.002	0.998	0.13	0.571	0.034	0.966
0.14	0.584	0.003	0.997	0.14	0.584	0.051	0.949
0.15	0.596	0.004	0.996	0.15	0.596	0.072	0.928
0.16	0.607	0.003	0.997	0.16	0.607	0.094	0.906
0.18	0.618	0.003	0.997	0.18	0.618	0.114	0.886
0.19	0.627	0.003	0.997	0.19	0.627	0.131	0.869
0.20	0.636	0.003	0.997	0.20	0.636	0.118	0.882
0.21	0.644	0.005	0.995	0.21	0.644	0.262	0.738
0.22	0.652	0.006	0.994	0.22	0.652	0.424	0.576
0.23	0.659	0.016	0.984	0.23	0.659	0.433	0.567
0.24	0.666	0.015	0.985	0.24	0.666	0.438	0.562
0.25	0.673	0.015	0.985	0.25	0.673	0.368	0.632
0.26	0.680	0.015	0.985	0.26	0.680	0.371	0.629
0.27	0.687	0.014	0.986	0.27	0.687	0.372	0.628
0.29	0.693	0.020	0.980	0.29	0.693	0.714	0.286
0.30	0.700	0.019	0.981	0.30	0.700	0.710	0.290
0.31	0.707	0.018	0.982	0.31	0.707	0.707	0.293
0.32	0.714	0.017	0.983	0.32	0.714	0.706	0.294
0.33	0.720	0.016	0.984	0.33	0.720	0.585	0.415
0.34	0.727	0.015	0.985	0.34	0.727	0.585	0.415
0.35	0.734	0.013	0.987	0.35	0.734	0.546	0.454
0.36	0.741	0.013	0.987	0.36	0.741	0.547	0.453
0.37	0.749	0.013	0.987	0.37	0.749	0.364	0.636
0.38	0.757	0.014	0.986	0.38	0.757	0.443	0.557
0.40	0.765	0.014	0.986	0.40	0.765	0.489	0.511
0.41	0.773	0.014	0.986	0.41	0.773	0.489	0.511
0.42	0.782	0.016	0.984	0.42	0.782	0.621	0.379
0.43	0.792	0.021	0.979	0.43	0.792	0.622	0.378
0.44	0.802	0.019	0.981	0.44	0.802	0.623	0.377
0.45	0.813	0.017	0.983	0.45	0.813	0.623	0.377
0.46	0.825	0.016	0.984	0.46	0.825	0.623	0.377
0.47	0.838	0.014	0.986	0.47	0.838	0.624	0.376
0.48	0.852	0.012	0.988	0.48	0.852	0.672	0.328
0.49	0.867	0.010	0.990	0.49	0.867	0.672	0.328
0.51	0.884	0.009	0.991	0.51	0.884	0.664	0.336
0.52	0.903	0.008	0.992	0.52	0.903	0.637	0.363
0.53	0.923	0.006	0.994	0.53	0.923	0.553	0.447
0.54	0.945	0.009	0.991	0.54	0.945	0.553	0.447
0.55	0.968	0.008	0.992	0.55	0.968	0.552	0.448
0.56	0.991	0.006	0.994	0.56	0.991	0.552	0.448
0.57	1.013	0.005	0.995	0.57	1.013	0.540	0.460
0.58	1.034	0.004	0.996	0.58	1.034	0.511	0.489
0.59	1.053	0.003	0.997	0.59	1.053	0.512	0.488
0.60	1.071	0.004	0.996	0.60	1.071	0.468	0.532
0.61	1.089	0.004	0.996	0.61	1.089	0.468	0.532
0.63	1.106	0.005	0.995	0.63	1.106	0.640	0.360
0.64	1.123	0.007	0.993	0.64	1.123	0.400	0.600
0.65	1.140	0.007	0.993	0.65	1.140	0.318	0.682
0.66	1.157	0.008	0.992	0.66	1.157	0.318	0.682
0.67	1.174	0.007	0.993	0.67	1.174	0.317	0.683
0.68	1.192	0.005	0.995	0.68	1.192	0.315	0.685
0.69	1.210	0.004	0.996	0.69	1.210	0.315	0.685
0.70	1.229	0.004	0.996	0.70	1.229	0.314	0.686
0.71	1.249	0.003	0.997	0.71	1.249	0.319	0.681
0.72	1.269	0.003	0.997	0.72	1.269	0.326	0.674
0.74	1.291	0.003	0.997	0.74	1.291	0.315	0.685
0.75	1.314	0.006	0.994	0.75	1.314	0.322	0.678
0.76	1.340	0.006	0.994	0.76	1.340	0.317	0.683
0.77	1.369	0.007	0.993	0.77	1.369	0.407	0.593
0.78	1.400	0.006	0.994	0.78	1.400	0.404	0.596
0.79	1.437	0.012	0.988	0.79	1.437	0.631	0.369
0.80	1.479	0.030	0.970	0.80	1.479	0.646	0.354
0.81	1.527	0.025	0.975	0.81	1.527	0.658	0.342
0.82	1.582	0.077	0.923	0.82	1.582	0.893	0.107
0.83	1.644	0.059	0.941	0.83	1.644	0.921	0.079
0.85	1.714	0.051	0.949	0.85	1.714	0.930	0.070
0.86	1.797	0.040	0.960	0.86	1.797	0.941	0.059
0.87	1.895	0.031	0.969	0.87	1.895	0.960	0.040
0.88	2.014	0.151	0.849	0.88	2.014	0.972	0.028
0.89	2.155	0.069	0.931	0.89	2.155	0.986	0.014
0.90	2.310	0.051	0.949	0.90	2.310	0.988	0.012
0.91	2.481	0.092	0.908	0.91	2.481	0.924	0.076
0.92	2.672	0.070	0.930	0.92	2.672	0.994	0.006
0.93	2.892	0.037	0.963	0.93	2.892	0.999	0.001
0.94	3.151	0.019	0.981	0.94	3.151	0.575	0.425
0.96	3.473	0.090	0.910	0.96	3.473	0.605	0.395
0.97	3.909	0.074	0.926	0.97	3.909	0.335	0.665
0.98	4.610	0.052	0.948	0.98	4.610	0.274	0.726
0.99	5.840	0.000	1.000	0.99	5.840	0.183	0.817

Table 15: Point-Specific P-Values for Θ_d (First-Order Dominance Test): Combined Effects

Percentile	Θ_a	P-val $H_2 : G_{a\|Ntr} < G_{a\|CSR}$	P-val $H_3 : G_{a\|Ntr} > G_{a\|CSR}$	Percentile	Θ_a	P-val $H_2 : G_{a\|w=\$11} < G_{a\|w=\$15}$	P-val $H_3 : G_{a\|w=\$11} > G_{a\|w=\$15}$
0.01	−0.506	0.709	0.291	0.01	−0.506	0.390	0.610
0.02	−0.468	0.694	0.306	0.02	−0.468	0.401	0.599
0.03	−0.431	0.811	0.189	0.03	−0.431	0.169	0.831
0.04	−0.398	0.759	0.241	0.04	−0.398	0.235	0.765
0.05	−0.368	0.498	0.502	0.05	−0.368	0.053	0.947
0.07	−0.341	0.414	0.586	0.07	−0.341	0.082	0.918
0.08	−0.316	0.509	0.491	0.08	−0.316	0.130	0.870
0.09	−0.294	0.220	0.780	0.09	−0.294	0.063	0.937
0.10	−0.272	0.237	0.763	0.10	−0.272	0.085	0.915
0.11	−0.252	0.113	0.887	0.11	−0.252	0.022	0.978
0.12	−0.232	0.116	0.884	0.12	−0.232	0.023	0.977
0.13	−0.213	0.126	0.874	0.13	−0.213	0.021	0.979
0.14	−0.194	0.130	0.870	0.14	−0.194	0.022	0.978
0.15	−0.175	0.148	0.852	0.15	−0.175	0.030	0.970
0.16	−0.156	0.058	0.942	0.16	−0.156	0.149	0.851
0.18	−0.137	0.011	0.989	0.18	−0.137	0.025	0.975
0.19	−0.118	0.023	0.977	0.19	−0.118	0.013	0.987
0.20	−0.099	0.028	0.972	0.20	−0.099	0.014	0.986
0.21	−0.081	0.058	0.942	0.21	−0.081	0.006	0.994
0.22	−0.064	0.099	0.901	0.22	−0.064	0.009	0.991
0.23	−0.047	0.107	0.893	0.23	−0.047	0.011	0.990
0.24	−0.031	0.220	0.780	0.24	−0.031	0.021	0.979
0.25	−0.016	0.219	0.781	0.25	−0.016	0.022	0.978
0.26	−0.001	0.222	0.778	0.26	−0.001	0.023	0.977
0.27	0.014	0.207	0.793	0.27	0.014	0.021	0.979
0.29	0.028	0.239	0.761	0.29	0.028	0.026	0.974
0.30	0.041	0.426	0.574	0.30	0.041	0.054	0.946
0.31	0.055	0.424	0.576	0.31	0.055	0.054	0.946
0.32	0.068	0.421	0.579	0.32	0.068	0.056	0.944
0.33	0.081	0.520	0.480	0.33	0.081	0.075	0.925
0.34	0.095	0.519	0.481	0.34	0.095	0.076	0.924
0.35	0.108	0.503	0.497	0.35	0.108	0.080	0.920
0.36	0.121	0.462	0.538	0.36	0.121	0.054	0.946
0.37	0.134	0.536	0.464	0.37	0.134	0.068	0.932
0.38	0.147	0.544	0.456	0.38	0.147	0.070	0.930
0.40	0.161	0.538	0.462	0.40	0.161	0.068	0.932
0.41	0.175	0.621	0.379	0.41	0.175	0.087	0.913
0.42	0.188	0.617	0.383	0.42	0.188	0.088	0.912
0.43	0.202	0.567	0.433	0.43	0.202	0.070	0.930
0.44	0.216	0.404	0.596	0.44	0.216	0.119	0.881
0.45	0.230	0.405	0.595	0.45	0.230	0.118	0.882
0.46	0.244	0.436	0.564	0.46	0.244	0.094	0.906
0.47	0.257	0.509	0.491	0.47	0.257	0.027	0.973
0.48	0.270	0.525	0.475	0.48	0.270	0.027	0.973
0.49	0.283	0.535	0.465	0.49	0.283	0.029	0.971
0.51	0.296	0.568	0.432	0.51	0.296	0.030	0.970
0.52	0.308	0.632	0.368	0.52	0.308	0.039	0.961
0.53	0.320	0.615	0.385	0.53	0.320	0.040	0.960
0.54	0.331	0.706	0.294	0.54	0.331	0.017	0.983
0.55	0.342	0.702	0.298	0.55	0.342	0.017	0.983
0.56	0.353	0.702	0.298	0.56	0.353	0.015	0.985
0.57	0.364	0.474	0.526	0.57	0.364	0.006	0.994
0.58	0.374	0.381	0.619	0.58	0.374	0.009	0.991
0.59	0.384	0.401	0.599	0.59	0.384	0.010	0.990
0.60	0.394	0.356	0.644	0.60	0.394	0.012	0.988
0.61	0.403	0.292	0.708	0.61	0.403	0.016	0.984
0.63	0.412	0.304	0.696	0.63	0.412	0.016	0.984
0.64	0.421	0.182	0.818	0.64	0.421	0.008	0.992
0.65	0.430	0.071	0.929	0.65	0.430	0.022	0.978
0.66	0.438	0.064	0.936	0.66	0.438	0.017	0.983
0.67	0.446	0.065	0.935	0.67	0.446	0.014	0.986
0.68	0.454	0.064	0.936	0.68	0.454	0.013	0.987
0.69	0.462	0.073	0.927	0.69	0.462	0.007	0.993
0.70	0.470	0.201	0.799	0.70	0.470	0.023	0.977
0.71	0.478	0.195	0.805	0.71	0.478	0.020	0.980
0.72	0.485	0.186	0.814	0.72	0.485	0.017	0.983
0.74	0.493	0.163	0.837	0.74	0.493	0.016	0.984
0.75	0.500	0.028	0.972	0.75	0.500	0.079	0.921
0.76	0.508	0.023	0.977	0.76	0.508	0.076	0.924
0.77	0.515	0.022	0.978	0.77	0.515	0.075	0.925
0.78	0.523	0.021	0.979	0.78	0.523	0.075	0.925
0.79	0.531	0.023	0.977	0.79	0.531	0.060	0.940
0.80	0.539	0.019	0.981	0.80	0.539	0.055	0.945
0.81	0.547	0.043	0.957	0.81	0.547	0.098	0.902
0.82	0.556	0.218	0.782	0.82	0.556	0.312	0.688
0.83	0.565	0.223	0.777	0.83	0.565	0.321	0.679
0.85	0.574	0.214	0.786	0.85	0.574	0.316	0.684
0.86	0.584	0.207	0.793	0.86	0.584	0.253	0.747
0.87	0.594	0.495	0.505	0.87	0.594	0.342	0.658
0.88	0.605	0.406	0.594	0.88	0.605	0.405	0.595
0.89	0.617	0.402	0.598	0.89	0.617	0.402	0.598
0.90	0.629	0.412	0.588	0.90	0.629	0.409	0.591
0.91	0.643	0.674	0.326	0.91	0.643	0.583	0.417
0.92	0.657	0.772	0.228	0.92	0.657	0.435	0.565
0.93	0.673	0.800	0.200	0.93	0.673	0.429	0.571
0.94	0.691	0.483	0.517	0.94	0.691	0.699	0.301
0.96	0.712	0.482	0.518	0.96	0.712	0.704	0.296
0.97	0.737	0.507	0.493	0.97	0.737	0.692	0.308
0.98	0.768	0.693	0.307	0.98	0.768	0.407	0.593
0.99	0.814	0.194	0.806	0.99	0.814	0.711	0.289

Table 16: Point-Specific P-Values for Θ_a (First-Order Dominance Test)

Percentile	Θ_a	P-val $H_2 : G_{a\|Ntr} < G_{a\|CSR+Trtmt}$	P-val $H_3 : G_{a\|Ntr} > G_{a\|CSR+Trtmt}$	Percentile	Θ_a	P-val $H_2 : G_{a\|Ntr,w=\$11} < G_{a\|CSR,w=\$15}$	P-val $H_3 : G_{a\|Ntr,w=\$11} > G_{a\|CSR,w=\$15}$
0.01	−0.506	0.639	0.361	0.01	−0.506	0.410	0.590
0.02	−0.468	0.694	0.306	0.02	−0.468	0.360	0.640
0.03	−0.431	0.811	0.189	0.03	−0.431	0.348	0.652
0.04	−0.398	0.758	0.242	0.04	−0.398	0.457	0.543
0.05	−0.368	0.498	0.502	0.05	−0.368	0.142	0.858
0.07	−0.341	0.487	0.513	0.07	−0.341	0.149	0.851
0.08	−0.316	0.550	0.450	0.08	−0.316	0.225	0.775
0.09	−0.294	0.221	0.779	0.09	−0.294	0.081	0.919
0.10	−0.272	0.254	0.746	0.10	−0.272	0.102	0.898
0.11	−0.252	0.113	0.887	0.11	−0.252	0.028	0.972
0.12	−0.232	0.116	0.884	0.12	−0.232	0.029	0.971
0.13	−0.213	0.126	0.874	0.13	−0.213	0.029	0.971
0.14	−0.194	0.130	0.870	0.14	−0.194	0.030	0.970
0.15	−0.175	0.205	0.795	0.15	−0.175	0.034	0.966
0.16	−0.156	0.059	0.941	0.16	−0.156	0.069	0.931
0.18	−0.137	0.020	0.980	0.18	−0.137	0.007	0.993
0.19	−0.118	0.023	0.977	0.19	−0.118	0.007	0.993
0.20	−0.099	0.053	0.947	0.20	−0.099	0.007	0.993
0.21	−0.081	0.096	0.904	0.21	−0.081	0.007	0.993
0.22	−0.064	0.100	0.900	0.22	−0.064	0.012	0.988
0.23	−0.047	0.230	0.770	0.23	−0.047	0.014	0.986
0.24	−0.031	0.221	0.779	0.24	−0.031	0.033	0.967
0.25	−0.016	0.221	0.779	0.25	−0.016	0.034	0.966
0.26	−0.001	0.292	0.708	0.26	−0.001	0.034	0.966
0.27	0.014	0.230	0.770	0.27	0.014	0.035	0.965
0.29	0.028	0.429	0.571	0.29	0.028	0.043	0.957
0.30	0.041	0.426	0.574	0.30	0.041	0.096	0.904
0.31	0.055	0.424	0.576	0.31	0.055	0.096	0.904
0.32	0.068	0.525	0.475	0.32	0.068	0.096	0.904
0.33	0.081	0.520	0.480	0.33	0.081	0.132	0.868
0.34	0.095	0.520	0.480	0.34	0.095	0.131	0.869
0.35	0.108	0.553	0.447	0.35	0.108	0.129	0.871
0.36	0.121	0.536	0.464	0.36	0.121	0.083	0.917
0.37	0.134	0.543	0.457	0.37	0.134	0.110	0.890
0.38	0.147	0.546	0.454	0.38	0.147	0.112	0.888
0.40	0.161	0.621	0.379	0.40	0.161	0.109	0.891
0.41	0.175	0.621	0.379	0.41	0.175	0.146	0.854
0.42	0.188	0.655	0.345	0.42	0.188	0.144	0.856
0.43	0.202	0.575	0.425	0.43	0.202	0.117	0.883
0.44	0.216	0.404	0.596	0.44	0.216	0.115	0.885
0.45	0.230	0.562	0.438	0.45	0.230	0.114	0.886
0.46	0.244	0.600	0.400	0.46	0.244	0.112	0.888
0.47	0.257	0.530	0.470	0.47	0.257	0.078	0.922
0.48	0.270	0.570	0.430	0.48	0.270	0.081	0.919
0.49	0.283	0.638	0.362	0.49	0.283	0.086	0.914
0.51	0.296	0.631	0.369	0.51	0.296	0.095	0.905
0.52	0.308	0.632	0.368	0.52	0.308	0.123	0.877
0.53	0.320	0.708	0.292	0.53	0.320	0.124	0.876
0.54	0.331	0.707	0.293	0.54	0.331	0.123	0.877
0.55	0.342	0.703	0.297	0.55	0.342	0.119	0.881
0.56	0.353	0.709	0.291	0.56	0.353	0.111	0.889
0.57	0.364	0.483	0.517	0.57	0.364	0.028	0.972
0.58	0.374	0.402	0.598	0.58	0.374	0.026	0.974
0.59	0.384	0.401	0.599	0.59	0.384	0.027	0.973
0.60	0.394	0.356	0.644	0.60	0.394	0.026	0.974
0.61	0.403	0.318	0.682	0.61	0.403	0.023	0.977
0.63	0.412	0.321	0.679	0.63	0.412	0.021	0.979
0.64	0.421	0.184	0.816	0.64	0.421	0.004	0.996
0.65	0.430	0.071	0.929	0.65	0.430	0.003	0.997
0.66	0.438	0.066	0.934	0.66	0.438	0.001	0.999
0.67	0.446	0.086	0.914	0.67	0.446	0.001	0.999
0.68	0.454	0.223	0.777	0.68	0.454	0.001	0.999
0.69	0.462	0.205	0.795	0.69	0.462	0.000	1.000
0.70	0.470	0.201	0.799	0.70	0.470	0.006	0.994
0.71	0.478	0.196	0.804	0.71	0.478	0.005	0.995
0.72	0.485	0.186	0.814	0.72	0.485	0.004	0.996
0.74	0.493	0.163	0.837	0.74	0.493	0.004	0.996
0.75	0.500	0.028	0.972	0.75	0.500	0.004	0.996
0.76	0.508	0.023	0.977	0.76	0.508	0.004	0.996
0.77	0.515	0.022	0.978	0.77	0.515	0.004	0.996
0.78	0.523	0.024	0.976	0.78	0.523	0.003	0.997
0.79	0.531	0.022	0.978	0.79	0.531	0.003	0.997
0.80	0.539	0.246	0.754	0.80	0.539	0.001	0.999
0.81	0.547	0.233	0.767	0.81	0.547	0.004	0.996
0.82	0.556	0.234	0.766	0.82	0.556	0.092	0.908
0.83	0.565	0.236	0.764	0.83	0.565	0.083	0.917
0.85	0.574	0.463	0.537	0.85	0.574	0.061	0.939
0.86	0.584	0.496	0.504	0.86	0.584	0.029	0.971
0.87	0.594	0.501	0.499	0.87	0.594	0.195	0.805
0.88	0.605	0.406	0.594	0.88	0.605	0.166	0.834
0.89	0.617	0.416	0.584	0.89	0.617	0.136	0.864
0.90	0.629	0.668	0.332	0.90	0.629	0.120	0.880
0.91	0.643	0.757	0.243	0.91	0.643	0.543	0.457
0.92	0.657	0.772	0.228	0.92	0.657	0.562	0.438
0.93	0.673	0.800	0.200	0.93	0.673	0.571	0.429
0.94	0.691	0.483	0.517	0.94	0.691	0.577	0.423
0.96	0.712	0.508	0.492	0.96	0.712	0.584	0.416
0.97	0.737	0.507	0.493	0.97	0.737	0.602	0.398
0.98	0.768	0.690	0.310	0.98	0.768	0.509	0.491
0.99	0.814	0.601	0.399	0.99	0.814	0.503	0.497

Table 17: Point-Specific P-Values for Θ_a (First-Order Dominance Test):Combined Effects

Percentile	Θ_l	P-val $H_2 : G_{l\|Ntr} < G_{l\|CSR}$	P-val $H_3 : G_{l\|Ntr} > G_{l\|CSR}$	Percentile	Θ_l	P-val $H_2 : G_{l\|w=\$11} < G_{l\|w=\$15}$	P-val $H_3 : G_{l\|w=\$11} > G_{l\|w=\$15}$
0.01	1.515	0.000	1.000	0.01	1.515	0.000	1.000
0.02	1.828	0.000	1.000	0.02	1.828	0.141	0.859
0.03	2.503	0.000	1.000	0.03	2.503	0.623	0.377
0.04	2.892	0.020	0.980	0.04	2.892	0.023	0.977
0.05	3.135	0.056	0.944	0.05	3.135	0.156	0.844
0.07	3.416	0.035	0.965	0.07	3.416	0.464	0.536
0.08	3.691	0.037	0.963	0.08	3.691	0.443	0.557
0.09	3.935	0.025	0.975	0.09	3.935	0.162	0.838
0.10	4.148	0.054	0.946	0.10	4.148	0.379	0.621
0.11	4.338	0.020	0.980	0.11	4.338	0.037	0.963
0.12	4.514	0.023	0.977	0.12	4.514	0.066	0.934
0.13	4.681	0.042	0.958	0.13	4.681	0.014	0.986
0.14	4.842	0.014	0.986	0.14	4.842	0.010	0.990
0.15	4.998	0.017	0.983	0.15	4.998	0.006	0.994
0.16	5.152	0.033	0.967	0.16	5.152	0.025	0.975
0.18	5.304	0.037	0.963	0.18	5.304	0.029	0.971
0.19	5.455	0.042	0.958	0.19	5.455	0.032	0.968
0.20	5.606	0.045	0.955	0.20	5.606	0.030	0.970
0.21	5.758	0.076	0.924	0.21	5.758	0.073	0.927
0.22	5.912	0.166	0.834	0.22	5.912	0.191	0.809
0.23	6.067	0.247	0.753	0.23	6.067	0.091	0.909
0.24	6.224	0.235	0.765	0.24	6.224	0.233	0.767
0.25	6.384	0.242	0.758	0.25	6.384	0.231	0.769
0.26	6.546	0.177	0.823	0.26	6.546	0.131	0.869
0.27	6.710	0.131	0.869	0.27	6.710	0.197	0.803
0.29	6.878	0.178	0.822	0.29	6.878	0.237	0.763
0.30	7.048	0.119	0.881	0.30	7.048	0.292	0.708
0.31	7.222	0.123	0.877	0.31	7.222	0.292	0.708
0.32	7.398	0.169	0.831	0.32	7.398	0.181	0.819
0.33	7.579	0.227	0.773	0.33	7.579	0.261	0.739
0.34	7.762	0.291	0.709	0.34	7.762	0.322	0.678
0.35	7.950	0.178	0.822	0.35	7.950	0.139	0.861
0.36	8.142	0.217	0.783	0.36	8.142	0.097	0.903
0.37	8.338	0.220	0.780	0.37	8.338	0.105	0.895
0.38	8.537	0.279	0.721	0.38	8.537	0.068	0.932
0.40	8.740	0.272	0.728	0.40	8.740	0.068	0.932
0.41	8.946	0.253	0.747	0.41	8.946	0.073	0.927
0.42	9.155	0.200	0.800	0.42	9.155	0.103	0.897
0.43	9.365	0.138	0.862	0.43	9.365	0.130	0.870
0.44	9.578	0.128	0.872	0.44	9.578	0.129	0.871
0.45	9.792	0.100	0.900	0.45	9.792	0.163	0.837
0.46	10.007	0.136	0.864	0.46	10.007	0.211	0.789
0.47	10.223	0.136	0.864	0.47	10.223	0.210	0.790
0.48	10.439	0.135	0.865	0.48	10.439	0.209	0.791
0.49	10.656	0.288	0.712	0.49	10.656	0.153	0.847
0.51	10.872	0.346	0.654	0.51	10.872	0.132	0.868
0.52	11.088	0.411	0.589	0.52	11.088	0.167	0.833
0.53	11.304	0.336	0.664	0.53	11.304	0.213	0.787
0.54	11.519	0.395	0.605	0.54	11.519	0.249	0.751
0.55	11.734	0.410	0.590	0.55	11.734	0.210	0.790
0.56	11.947	0.585	0.415	0.56	11.947	0.221	0.779
0.57	12.160	0.654	0.346	0.57	12.160	0.264	0.736
0.58	12.372	0.564	0.436	0.58	12.372	0.280	0.720
0.59	12.582	0.602	0.398	0.59	12.582	0.326	0.674
0.60	12.792	0.528	0.472	0.60	12.792	0.374	0.626
0.61	13.002	0.528	0.472	0.61	13.002	0.381	0.619
0.63	13.212	0.531	0.469	0.63	13.212	0.379	0.621
0.64	13.424	0.444	0.556	0.64	13.424	0.425	0.575
0.65	13.637	0.521	0.479	0.65	13.637	0.462	0.538
0.66	13.852	0.595	0.405	0.66	13.852	0.494	0.506
0.67	14.070	0.658	0.342	0.67	14.070	0.451	0.549
0.68	14.292	0.567	0.433	0.68	14.292	0.442	0.558
0.69	14.517	0.537	0.463	0.69	14.517	0.441	0.559
0.70	14.747	0.474	0.526	0.70	14.747	0.403	0.597
0.71	14.983	0.584	0.416	0.71	14.983	0.300	0.700
0.72	15.225	0.636	0.364	0.72	15.225	0.187	0.813
0.74	15.474	0.717	0.283	0.74	15.474	0.244	0.756
0.75	15.731	0.661	0.339	0.75	15.731	0.255	0.745
0.76	15.999	0.404	0.596	0.76	15.999	0.282	0.718
0.77	16.277	0.534	0.466	0.77	16.277	0.273	0.727
0.78	16.569	0.590	0.410	0.78	16.569	0.235	0.765
0.79	16.876	0.548	0.452	0.79	16.876	0.229	0.771
0.80	17.203	0.508	0.492	0.80	17.203	0.205	0.795
0.81	17.551	0.406	0.594	0.81	17.551	0.167	0.833
0.82	17.925	0.436	0.564	0.82	17.925	0.140	0.860
0.83	18.328	0.530	0.470	0.83	18.328	0.133	0.867
0.85	18.764	0.763	0.237	0.85	18.764	0.135	0.865
0.86	19.238	0.783	0.217	0.86	19.238	0.124	0.876
0.87	19.757	0.839	0.161	0.87	19.757	0.099	0.901
0.88	20.330	0.904	0.096	0.88	20.330	0.075	0.925
0.89	20.966	0.941	0.059	0.89	20.966	0.053	0.947
0.90	21.679	0.922	0.078	0.90	21.679	0.049	0.951
0.91	22.488	0.963	0.037	0.91	22.488	0.037	0.963
0.92	23.412	0.913	0.087	0.92	23.412	0.039	0.961
0.93	24.481	0.863	0.137	0.93	24.481	0.027	0.973
0.94	25.727	0.922	0.078	0.94	25.727	0.020	0.980
0.96	27.182	0.930	0.070	0.96	27.182	0.000	1.000
0.97	28.861	0.824	0.176	0.97	28.861	0.000	1.000
0.98	30.744	0.853	0.147	0.98	30.744	0.000	1.000
0.99	32.760	0.894	0.106	0.99	32.760	0.000	1.000

Table 18: Point-Specific P-Values for Θ_l (First-Order Dominance Test)

Percentile	Θ_l	P-val $H_2 : G_{l\|Ntr} < G_{l\|CSR+Trtmt}$	P-val $H_3 : G_{l\|Ntr} > G_{l\|CSR+Trtmt}$	Percentile	Θ_l	P-val $H_2 : G_{l\|Ntr,w=\$11} < G_{l\|CSR,w=\$15}$	P-val $H_3 : G_{l\|Ntr,w=\$11} > G_{l\|CSR,w=\$15}$
0.01	1.515	0.000	1.000	0.01	1.515	0.000	1.000
0.02	1.828	0.000	1.000	0.02	1.828	0.000	1.000
0.03	2.503	0.000	1.000	0.03	2.503	0.000	1.000
0.04	2.892	0.000	1.000	0.04	2.892	0.000	1.000
0.05	3.135	0.030	0.970	0.05	3.135	0.020	0.980
0.07	3.416	0.054	0.946	0.07	3.416	0.022	0.978
0.08	3.691	0.035	0.965	0.08	3.691	0.025	0.975
0.09	3.935	0.043	0.957	0.09	3.935	0.011	0.989
0.10	4.148	0.055	0.945	0.10	4.148	0.234	0.766
0.11	4.338	0.018	0.982	0.11	4.338	0.001	0.999
0.12	4.514	0.021	0.980	0.12	4.514	0.002	0.998
0.13	4.681	0.023	0.977	0.13	4.681	0.003	0.997
0.14	4.842	0.013	0.987	0.14	4.842	0.000	1.000
0.15	4.998	0.009	0.991	0.15	4.998	0.001	0.999
0.16	5.152	0.011	0.989	0.16	5.152	0.003	0.997
0.18	5.304	0.021	0.979	0.18	5.304	0.004	0.996
0.19	5.455	0.040	0.960	0.19	5.455	0.005	0.995
0.20	5.606	0.065	0.935	0.20	5.606	0.004	0.996
0.21	5.758	0.119	0.881	0.21	5.758	0.011	0.989
0.22	5.912	0.119	0.881	0.22	5.912	0.087	0.913
0.23	6.067	0.188	0.812	0.23	6.067	0.085	0.915
0.24	6.224	0.171	0.829	0.24	6.224	0.142	0.858
0.25	6.384	0.179	0.821	0.25	6.384	0.150	0.850
0.26	6.546	0.241	0.759	0.26	6.546	0.061	0.939
0.27	6.710	0.180	0.820	0.27	6.710	0.060	0.940
0.29	6.878	0.127	0.873	0.29	6.878	0.098	0.902
0.30	7.048	0.056	0.944	0.30	7.048	0.067	0.933
0.31	7.222	0.087	0.913	0.31	7.222	0.067	0.933
0.32	7.398	0.125	0.875	0.32	7.398	0.067	0.933
0.33	7.579	0.128	0.872	0.33	7.579	0.113	0.887
0.34	7.762	0.129	0.871	0.34	7.762	0.177	0.823
0.35	7.950	0.069	0.931	0.35	7.950	0.045	0.955
0.36	8.142	0.177	0.823	0.36	8.142	0.046	0.954
0.37	8.338	0.238	0.762	0.37	8.338	0.048	0.952
0.38	8.537	0.332	0.668	0.38	8.537	0.050	0.950
0.40	8.740	0.330	0.670	0.40	8.740	0.038	0.962
0.41	8.946	0.263	0.737	0.41	8.946	0.041	0.959
0.42	9.155	0.258	0.742	0.42	9.155	0.044	0.956
0.43	9.365	0.182	0.818	0.43	9.365	0.033	0.967
0.44	9.578	0.134	0.866	0.44	9.578	0.036	0.964
0.45	9.792	0.106	0.894	0.45	9.792	0.039	0.961
0.46	10.007	0.141	0.859	0.46	10.007	0.063	0.937
0.47	10.223	0.142	0.858	0.47	10.223	0.063	0.937
0.48	10.439	0.185	0.815	0.48	10.439	0.062	0.938
0.49	10.656	0.235	0.765	0.49	10.656	0.093	0.907
0.51	10.872	0.236	0.764	0.51	10.872	0.092	0.908
0.52	11.088	0.292	0.708	0.52	11.088	0.136	0.864
0.53	11.304	0.331	0.669	0.53	11.304	0.135	0.865
0.54	11.519	0.391	0.609	0.54	11.519	0.180	0.820
0.55	11.734	0.380	0.620	0.55	11.734	0.182	0.818
0.56	11.947	0.351	0.649	0.56	11.947	0.281	0.719
0.57	12.160	0.411	0.589	0.57	12.160	0.363	0.637
0.58	12.372	0.420	0.580	0.58	12.372	0.321	0.679
0.59	12.582	0.352	0.648	0.59	12.582	0.389	0.611
0.60	12.792	0.421	0.579	0.60	12.792	0.389	0.611
0.61	13.002	0.482	0.518	0.61	13.002	0.391	0.609
0.63	13.212	0.454	0.546	0.63	13.212	0.353	0.647
0.64	13.424	0.439	0.561	0.64	13.424	0.354	0.646
0.65	13.637	0.439	0.561	0.65	13.637	0.424	0.576
0.66	13.852	0.514	0.486	0.66	13.852	0.489	0.511
0.67	14.070	0.552	0.448	0.67	14.070	0.488	0.512
0.68	14.292	0.540	0.460	0.68	14.292	0.421	0.579
0.69	14.517	0.581	0.419	0.69	14.517	0.420	0.580
0.70	14.747	0.527	0.473	0.70	14.747	0.338	0.662
0.71	14.983	0.527	0.473	0.71	14.983	0.277	0.723
0.72	15.225	0.455	0.545	0.72	15.225	0.191	0.809
0.74	15.474	0.483	0.517	0.74	15.474	0.285	0.715
0.75	15.731	0.413	0.587	0.75	15.731	0.276	0.724
0.76	15.999	0.423	0.577	0.76	15.999	0.204	0.796
0.77	16.277	0.495	0.505	0.77	16.277	0.235	0.765
0.78	16.569	0.524	0.476	0.78	16.569	0.203	0.797
0.79	16.876	0.437	0.563	0.79	16.876	0.191	0.809
0.80	17.203	0.458	0.542	0.80	17.203	0.146	0.854
0.81	17.551	0.406	0.594	0.81	17.551	0.092	0.908
0.82	17.925	0.377	0.623	0.82	17.925	0.071	0.929
0.83	18.328	0.323	0.677	0.83	18.328	0.077	0.923
0.85	18.764	0.388	0.612	0.85	18.764	0.124	0.876
0.86	19.238	0.597	0.403	0.86	19.238	0.108	0.892
0.87	19.757	0.762	0.238	0.87	19.757	0.096	0.904
0.88	20.330	0.850	0.150	0.88	20.330	0.082	0.918
0.89	20.966	0.890	0.110	0.89	20.966	0.071	0.929
0.90	21.679	0.908	0.092	0.90	21.679	0.059	0.941
0.91	22.488	0.906	0.094	0.91	22.488	0.061	0.939
0.92	23.412	0.737	0.263	0.92	23.412	0.047	0.953
0.93	24.481	0.867	0.133	0.93	24.481	0.025	0.975
0.94	25.727	0.866	0.134	0.94	25.727	0.000	1.000
0.96	27.182	0.916	0.084	0.96	27.182	0.000	1.000
0.97	28.861	0.801	0.199	0.97	28.861	0.000	1.000
0.98	30.744	0.133	0.867	0.98	30.744	0.000	1.000
0.99	32.760	0.855	0.145	0.99	32.760	0.000	1.000

Table 19: Point-Specific P-Values for Θ_l (First-Order Dominance Test): Combined Effects