NBER WORKING PAPER SERIES

INCENTIVES, SELECTION AND PRODUCTIVITY IN LABOR MARKETS: EVIDENCE FROM RURAL MALAWI

Raymond P. Guiteras
B. Kelsey Jack

Working Paper 19825
http://www.nber.org/papers/w19825

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2014

We thank seminar and workshop participants at the University of Chicago, University College London/ L.S.E., Duke University, Case Western Reserve, RAND and NEUDC. James Berry, Constanca Esteves-Sorenson, Andrew Foster, Jessica Goldberg, John Ham, John List, Jeremy Magruder and A. Mushfiq Mobarak provided useful comments. John Anderson and Stanley Mvula were invaluable in leading the field team. We gratefully acknowledge financial support from ICRAF and the University of Maryland Department of Economics. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Incentives, Selection and Productivity in Labor Markets: Evidence from Rural Malawi
Raymond P. Guiteras and B. Kelsey Jack
NBER Working Paper No. 19825
January 2014
JEL No. C93,J22,J24,J33,O12

ABSTRACT

An observed positive relationship between compensation and productivity cannot distinguish between two channels: (1) an incentive effect and (2) worker selection. We use a simplified Becker-DeGroot-Marschak mechanism, which provides random variation in piece rates conditional on revealed reservation rates, to separately identify the two channels in the context of casual labor markets in rural Malawi. A higher piece rate increases output in our setting, but does not attract more productive workers. Among men, the average worker recruited at higher piece rates is actually less productive. Local labor market imperfections appear to undermine the worker sorting observed in well-functioning labor markets.

Raymond P. Guiteras
Department of Economics
3114 Tydings Hall
University of Maryland
College Park, MD 20742
guiteras@econ.umd.edu

B. Kelsey Jack
Department of Economics
Tufts University
314 Braker Hall
Medford, MA 02155
and NBER
kelsey.jack@tufts.edu

An online appendix is available at:
http://www.nber.org/data-appendix/w19825

1 Introduction

An observed positive relationship between compensation and productivity cannot distinguish between two channels: (1) an incentive effect: a higher piece rate motivates productivity; (2) selection: if higher-productivity workers have better outside options, then increasing compensation attracts a higher-productivity pool of workers. These channels have different implications for efficiency: the first effect represents a causal effect and increased output, whereas the second could be simply a zero-sum reallocation of labor. In the absence of information on reservation piece rates, even randomly assigned piece rates cannot solve this problem unless worker selection is also exogenously determined.

A long-standing literature describes the role of sorting in labor markets, including the effect of market frictions on efficiency (Sattinger 1993; Abowd et al. 1999; Shimer and Smith 2000). The empirical literature from developed countries finds that worker selection constitutes a meaningful share of the relationship between compensation and productivity (e.g. Lazear 2000; Dohmen and Falk 2011). More recently, in a developing country context, Fafchamps, Söderbom, and Benhassine (2009) examine the relationship between wages, education and gender in the manufacturing sector. While they find evidence of sorting, they also emphasize that various market failures can interfere with efficient sorting. To the extent that higher productivity employers do not match with more productive workers, labor markets may allocate workers inefficiently across jobs. However, few quasi-experimental studies on the determinants of worker productivity are available for developing countries, perhaps due to the scarcity of detailed micro-data. This paper joins a small but growing number of studies that use experimental variation to examine the performance of existing markets in developing countries,[1] and is the first to separate the effects of worker selection from worker effort.

We use a simplified Becker-DeGroot-Marschak mechanism (BDM, Becker et al. 1964) to identify the incentive and selection channels in the context of informal day labor markets in rural Malawi. Workers choose the minimum piece rate at which they are willing to accept a one-day contract to perform a simple task: sorting beans by type and quality. A piece rate offer is then generated randomly, determining whether the worker is given a contract and, if so, the piece rate. Random assignment to a quality monitoring treatment provides exogenous variation in the workers' incentives to trade off quantity of output for quality. The experiment is conducted over four consecutive days in each of twelve villages, spanning both the low and high labor demand seasons.[2]

[1]For example, Goldberg (2011); Beaman and Magruder (2012); Beaman et al. (2013); Bloom et al. (2013); Dupas and Robinson (2013) and others.

[2]Agriculture in Malawi is rainfed with a single cropping cycle per year. Peak labor demand occurs

We use the resulting data to examine a number of determinants of worker productivity. First, we separate the effects of a higher piece rate on the average productivity of contracted workers from its direct effect on productivity. We also compare effort allocation toward quantity versus quality with and without explicit incentives for quality. We observe that workers are responsive to the piece rate in terms of the quantity of output produced, and that output quantity and quality are substitutes. Consequently, the introduction of explicit quality monitoring improves the average quality of production but at a quantity cost: workers are slower but more precise when errors are penalized. The selection and incentive effects of piece rates are of opposite signs. While higher piece rates encourage more effort, they also attract workers that are less productive, on average.

Second, the design is stratified by gender, which is an important determinant of labor market outcomes in rural Malawi. Both extensive and intensive margin effects are sensitive to gender. In our setting, women both accept lower piece rates and produce more and higher quality output. Furthermore, the overall negative selection effect is driven exclusively by men, for whom a higher minimum piece rate is associated with lower output quantity. We do not observe significant selection among women. These observed gender differences are consistent with an outside option for men that rewards different skills than those required by the bean sorting task.

Third, the experiment is implemented during both the low and the high agricultural seasons, across which the opportunity cost of time and the marginal value of money vary. Specifically, during the high season, the opportunity cost of time is relatively high since on- and off-farm labor demand is high, but the marginal value of income is also relatively high, since this is a cash-poor season and a time of year in which households report frequent food shortages. In spite of better self-reported outside options during the high season, participants have lower stated willingness to accept and produce more output, which may be due to a higher marginal value of income during these months. Men are more productive during the high season, while the productivity of women does not vary with the season. Though the study design does not hold the worker pool constant across seasons, the observed seasonality in the results is consistent with differences in the available outside option and fluctuations in liquidity constraints by season.

Fourth, our experimental design introduces variation in the piece rates received by workers seated together during the observed task. We use within-group variation in piece rates to test for peer effects. We find that the average exogenously determined piece rate for other

from November to May, during which crops are planted, tended and harvested. We refer to this as the "high" season. Food shortages and liquidity constraints are most acute in the months leading up to harvest, specifically January, February and March.

workers in the group has a positive effect on worker productivity in small groups, but the effect decreases as the size of the work group increases, and is negative in large groups.

This paper contributes to a number of different strands of literature in both labor and development economics. First, both observational and experimental studies have examined the relative importance of worker selection and worker effort in determining the total productivity effect of incentives in well functioning labor markets.[3] While these studies suggest that selection is an important determinant of worker ability, they tend to compare selection across types rather than levels of compensation. Selection may be less important for determining overall output in settings where effort can be measured and tied directly to compensation, which is more likely to be the case for output quantity than quality.[4] Our experimental design varies both the level and type of incentive scheme and separately measures the selection and incentive effects of the former and the combined effect of the latter.

Second, studies in development economics on gender differences in labor supply date back several decades, and consistently document differences in supply elasticities by gender (Bardhan 1979; Rosenzweig 1978). In a recent field experiment, Goldberg (2011) randomly varies daily wages in rural Malawi and finds similar supply elasticities for men and women during the low labor demand season. In another recent field experiment, Beaman, Keleher, and Magruder (2013) study the role of gender and social networks in job matching among educated workers in urban Malawi, finding that hiring via referrals exacerbates women's disadvantages in labor markets. While numerous previous studies have shown that men and women face different labor market opportunities, our design allows us to characterize the margins on which these differences operate.

Third, workers may be sensitive to the effort choices or incentives of their peers (Gaechter et al. 2010). Peer effects may be driven by the production technology (e.g., Mas and Moretti (2009)) or the incentive scheme (e.g., Bandiera et al. (2005)). Generally, studies of peer

[3]In the context of a U.S. factory producing windshield glass, Lazear (2000) concludes that approximately half of the productivity gains from a switch from wages to piece rates is due to changes in worker composition, i.e. selection. Dohmen and Falk (2011) document sorting on both productivity and other worker characteristics in a laboratory setting. Eriksson and Villeval (2008) also use a laboratory setting to generate exogenous variation in incentive schemes and observe both sorting and effort effects. In a month-long data entry task, Heywood et al. (2013) examine a different type of selection – the employer's recruitment of motivated employees – and find that hiring more motivated workers is a substitute for monitoring the quality of output in a piece rate task.

[4]Where effort can be measured, the optimal piece rate depends on the elasticity of effort with respect to the piece rate (Stiglitz 1975). For example, in a study of workers in a tree planting firm in British Columbia, Paarsch and Shearer (1999) estimate an elasticity of effort, as measured by the number of trees planted per day, with respect to the piece rate. A substantial literature also examines the effects of different levels and types of incentives on worker effort choice (e.g. Bandiera et al. (2005); Fehr and Goette (2007); Bandiera et al. (2010)), including exogenous variation in monitoring (Nagin et al. 2002) but cannot typically identify both worker effort and worker composition effects.

effects find a positive spillover from highly productive workers to less productive colleagues, particularly when incentives to free ride are small. However, evidence from Kenya suggests that in settings with strong redistribution norms, peer effects may instead lead to negative spillovers from less productive individuals (Jakiela and Ozier 2012). To date, field studies of peer effects predominantly examine developed country labor markets, and may not generalize well to developing countries. We show evidence of positive productivity spillovers in a setting where redistribution norms are likely to be important, though decrease and eventually become negative as free riding incentives increase.

The study design offers a novel approach to characterizing labor market supply and productivity parameters in an environment where data are typically scarce. While the point estimates are specific to our study context, the findings provide several pieces of unique evidence and offer a methodology for generating rich micro-data in a setting where data constraints typically interfere with clean empirical identification. Our design cleanly separates the selection margin from the incentive margin, and shows that higher piece rates are more important for generating high effort than for attracting high quality workers.[5] While previous work has shown meaningful selection effects in developed country settings, our setting is characterized by highly imperfect labor markets, which may undermine worker sorting on productivity. To the extent that the findings differ from previous work in labor economics, they are consistent with the imperfections in local labor and credit markets that are pervasive in developing countries.

The paper proceeds as follows. Section 2 provides a simple theoretical model to motivate the experiment and frame the empirical analysis. Section 3 describes the experimental design and implementation. Section 4 presents the empirical results. Section 5 concludes.

2 Model

To provide a framework for our analysis, we describe a simple model of effort choice under a piece rate scheme. The model generates predictions about selection, effort, and the effects of monitoring. We extend the framework to discuss gender differences and peer effects.

[5] A number of recent field experiments in development economics have relied on a two stage randomization to isolate the effect of self selection on outcomes. Karlan and Zinman (2009) randomized interest rates before and after take up in a consumer credit experiment in South Africa, which allowed them to distinguish the effect of adverse selection from that of moral hazard on loan default rates. Ashraf et al. (2010) and Cohen and Dupas (2010) used similar two-stage pricing designs to isolate the screening effect of prices for health products.

2.1 Setup

A firm values output quantity, q, and loses revenue when output quality, Q, falls below a threshold $\bar{Q}$. It offers a piece rate r to workers for production of q and may also choose to monitor Q using a costly monitoring technology M. The monitoring technology, M, is binary ($M \in \{0, 1\}$), and is perfectly able to detect Q when Q falls below the threshold $\bar{Q}$. We assume there is a lower bound on quality $\underline{Q}$ such that the firm can costlessly detect $Q < \underline{Q}$ even when $M = 0$. We normalize Q such that $\underline{Q} = 0$ and $\bar{Q} = 1$.

Workers are offered a piece rate r for each unit of output. If the firm is monitoring ($M = 1$) and quality falls below the threshold $\bar{Q} = 1$, then the worker receives a quality-adjusted piece rate rQ. If the firm is not monitoring ($M = 0$), the worker receives r per unit output regardless of quality as long as $Q \geq 0$. In either regime (monitoring or not), the worker is not paid for output with $Q < 0$. The worker's income, therefore, is

$$\begin{aligned} y\left(q, Q; r, M\right) &= \begin{cases} rqQ & \text{if } M = 1 \\ rq & \text{if } M = 0 \end{cases} \\ &= rqQM + rq\left(1 - M\right) \end{aligned}$$

for all $Q \in [0, 1]$.[6]

The worker chooses to allocate effort toward production of q and Q, which together determine the cost of effort $c(q, Q)$, which is increasing and convex in each argument. Workers are indexed by their productivity, $\gamma \geq 1$, which for simplicity we model as entering multiplicatively and symmetrically between quantity and quality:[7]

$$c\left(q, Q; \gamma\right) = c\left(q, Q\right)/\gamma.$$

Worker utility is their income minus their effort cost:

$$\begin{aligned} U\left(q, Q; r, \gamma, M\right) &= \begin{cases} rqQ - c\left(q, Q\right)/\gamma & \text{if } M = 1 \\ rq - c\left(q, Q\right)/\gamma & \text{if } M = 0 \end{cases} \\ &= rqQM + rq\left(1 - M\right) - c\left(q, Q\right)/\gamma \end{aligned}$$

for all $Q \in [0, 1]$.

[6] $Q > 1$ cannot be optimal for the worker, since she is not paid for quality above the threshold. Similarly, the worker will never produce $Q < 0$, since in either regime he knows that he will not be paid.

[7] In the data, quality and quantity do seem to move together, in that their correlations with key covariates generally have the same sign. See discussion in Section 3.3.3.

2.2 Worker's optimal response

If the firm does not monitor ($M = 0$), the worker's optimal response is to set quality $Q^*_{NM} = 0$ and quantity q^*_{NM} determined by the first-order condition

$$q^*_{NM} \quad : \quad \frac{1}{\gamma}\left.\frac{\partial c}{\partial q}\right|_{\left(q^*_{NM},0\right)} = r. \tag{1}$$

If the firm does monitor ($M = 1$), the worker's optimum is either a corner solution, with $Q^*_M = 1$ and quantity q^*_M determined by the first-order condition

$$q^*_{M} \quad : \quad \frac{1}{\gamma}\left.\frac{\partial c}{\partial q}\right|_{\left(q^*_{M},1\right)} = r, \tag{2}$$

or an interior solution with (q^*_M, Q^*_M) solving the system of first-order conditions

$$\mathrm{FOC}_{q_M} : rQ^*_M \quad = \quad \frac{1}{\gamma}\left.\frac{\partial c}{\partial q}\right|_{\left(q^*_{M},Q^*_M\right)} \tag{3}$$

$$\mathrm{FOC}_{Q_M} : rq^*_M \quad = \quad \frac{1}{\gamma}\left.\frac{\partial c}{\partial Q}\right|_{\left(q^*_{M},Q^*_M\right)}. \tag{4}$$

Intuitively, in (3) the worker sets the marginal revenue from a unit of output[8] equal to the marginal effort cost in the quantity dimension, while in (4) the worker sets the marginal revenue from an improvement in quality equal to the marginal effort cost in the quality dimension. Since c is convex in both arguments, the first order conditions imply that higher-productivity workers produce more output and weakly higher quality output, i.e. $\partial q^*/\partial\gamma > 0$ and $\partial Q^*/\partial\gamma \geq 0$, with $\partial Q^*/\partial\gamma = 0$ when $M = 0$ or at the corner solution with $Q^*_M = 1$.

In the absence of monitoring ($M = 0$), a higher piece rate unambiguously increases effort in the quantity dimension, but quality will not improve. Similarly, a worker under monitoring ($M = 1$) optimizing at the corner ($Q^*_M = 1$), with first-order condition given by Equation (2), will unambiguously increase quantity as the piece rate increases, holding quality constant until she is moved to an interior solution, which will only occur if quantity and quality are substitutes. For a worker under monitoring ($M = 1$) at an interior solution given by Equations (3) and (4), optimal quantity will increase in response to an increase in the piece rate. Whether quality increases or decreases depends on whether quantity and quality are complements or substitutes in the worker's production function. For the task we study, they are likely to be substitutes ($\partial^2 c\,(q, Q)\,/\partial q\partial Q > 0$), in which case an increase in

[8]Given the optimal quality level Q^*_M, the quality-adjusted piece rate is rQ^*_M.

the piece rate increases output quality at a cost of a reduction in output quantity.

2.3 Selection

As the piece rate and monitoring technology are varied, workers will choose whether or not to accept a contract according to their utility under the contract and their outside option, $\underline{V}(\gamma)$. We index the outside option by the productivity parameter to emphasize that a worker's outside option will depend on her overall productivity, which may be reflected in γ, her productivity in this task. While we cannot sign this relationship unambiguously, $\underline{V}'(\gamma) > 0$ if workers who are more productive in this task have better outside options. This is likely to be the case for workers with outside options that reward similar skills.

The worker's participation constraints with and without monitoring are

$$V(\gamma, r; M = 1) = rq_M^* Q_M^* - c(q_M^*, Q_M^*)/\gamma \geq \underline{V}(\gamma) \tag{PC-M}$$

$$V(\gamma, r; M = 0) = rq_{NM}^* - c(q_{NM}^*, 0)/\gamma \geq \underline{V}(\gamma) \tag{PC-NM}$$

which lead to reservation rates

$$\begin{aligned} \underline{r}_M &= \frac{\underline{V}(\gamma) + c(q_M^*, Q_M^*)/\gamma}{q_M^* Q_M^*} \\ \underline{r}_{NM} &= \frac{\underline{V}(\gamma) + c(q_{NM}^*, 0)/\gamma}{q_{NM}^*}. \end{aligned}$$

We are interested in comparative statics with respect to monitoring (the relationship between $\underline{r}$ and M) and selection (the relationship between $\underline{r}$ and γ). The first is relatively simple: $\underline{r}_M > \underline{r}_{NM}$. This follows from the fact that $V(\gamma, r; M = 1) < V(\gamma, r; M = 0)$: monitoring imposes a constraint on the worker, so r must increase to compensate her. The second, whether selection is positively or negatively related to productivity (i.e. the sign of $\partial\underline{r}/\partial\gamma$) is ambiguous. Selection improves productivity, i.e. $\partial\underline{r}/\partial\gamma > 0$, if an increase in γ makes the participation constraint more difficult to satisfy. We consider the case $M = 1$.[9] The left-hand-side of (PC-M) has derivative[10]

$$\frac{dV(\gamma, r; M = 1)}{d\gamma} = \frac{\partial V(\gamma, r; M = 1)}{\partial\gamma} = \frac{c(q_M^*, Q_M^*)}{\gamma^2} > 0.$$

The right-hand side of (PC-M) has derivative $\underline{V}'(\gamma)$. If $\underline{V}'(\gamma) < 0$, i.e. if workers with

[9] The derivation when $M = 0$ is the same.

[10] Because $V(\gamma, r; M = 1)$ is a value function, by the envelope theorem it is sufficient to consider the partial derivative.

higher productivity in this task have lower-value outside options, then clearly $\partial \underline{r}/\partial\gamma < 0$. If $\underline{V}'(\gamma) > 0$, then the sign of $\partial \underline{r}/\partial\gamma$ depends on the relative magnitudes of $c(q_M^*, Q_M^*)/\gamma^2$ and $\underline{V}'(\gamma)$. Intuitively, as a worker's productivity increases, whether the minimum piece rate required for her to participate increases or decreases depends on how rapidly her effort cost decreases relative to the improvement in her outside option.

2.4 Direct and Indirect Effects of Monitoring and Piece Rates

In this section, we consider the overall effect of firm's choice variables, M and r. The simple point is that the firm must consider both direct effects on worker incentives for a fixed workforce, and the indirect effect via the selection mechanism.

For a given worker, monitoring increases quality and reduces quantity, as long as they are substitutes in the worker's production function. The general equilibrium effect may be larger or smaller than this partial equilibrium effect, depending on how monitoring affects the composition of the work force. From the participation constraints above, workers for whom (PC-NM) is satisfied but (PC-M) is not satisfied at the current piece rate will exit the work force when monitoring is introduced. If, however, these are relatively low-productivity workers, the effect on average quantity may be muted.

Similarly, for a given worker, an increase in the piece rate will increase the quantity produced, while the effect on quality is ambiguous, as shown in Section 2.2. Additionally, a change in the piece rate may change the composition of the labor force, which is clear from the participation constraints and the discussion in Section 2.3. As with monitoring, whether this selection effect from a change in the piece rate reinforces or counteracts the direct effect is ambiguous – here, it depends on the degree to which higher piece rates attract more or less productive workers.

2.5 Gender

In the context of our model, worker gender is primarily relevant through the joint distribution of productivity, γ, and the outside option, $\underline{V}$. Men and women can have different distributions of productivity, of the outside option, or the relationship between these two, i.e. the function $\underline{V}(\gamma)$. If the skills that enhance productivity in this task are more similar to the skills that enhance productivity in the opportunities available to women than to men, then we are more likely to observe positive selection in our experiment for women than we are for men.[11]

[11] Selection into different incentive schemes may be affected by factors other than the reservation wage. For example, Niederle and Vesterlund (2007) find that, in the laboratory, women choose competitive com-

2.6 Peer effects

In a neoclassical productivity model, workers respond to their own incentives to exert effort. However, worker productivity may also be affected by other factors such as the effort choices of their peers. Two possible channels are highlighted in the existing literature.[12] First, greater effort by peers may decrease a worker's own effort cost. In the context of our model, the marginal cost of output decreases as the output of peers increases. Second, workers may respond to relative incentives because of fairness norms, which affect the cost of effort. The predictions are ambiguous. On the one hand, workers who receive piece rates below those of their peers may be discouraged from exerting effort. Alternatively, workers who receive piece rates above those of their peers may lower effort to avoid social pressure to redistribute earnings.[13] Since output quantity is easier to observe than quality, we expect peer effects to be more likely to affect quantity.

3 Experimental design and implementation

To study productivity in the casual labor market, we create new demand for casual labor under controlled conditions that generate random variation in worker incentives. The context is informal day labor markets in rural Malawi, where such work is called *ganyu*. In Malawi, like in many rural agricultural settings in developing countries, labor markets are highly seasonal. Households both buy and sell labor, both for daily wages and in piece-rate-based jobs. In our study, workers are hired to sort harvested, dried beans into eight categories.[14] Sorted beans receive a price premium of roughly 50 percent. This task is well-suited to our

pensation schemes less often than men, even when their potential earnings are higher under competitive incentives. Gneezy et al. (2009) conduct a similar study across matriarchal and patriarchal societies and find that the lab results are reversed (women are more competitive) in the matriarchal society. Together these and numerous other studies suggest that men and women may sort into different incentive schemes based on underlying preferences. Women may also be differentially responsive to other aspects of the employment relationship, including the relationship with the employer (reciprocity) or relationships with colleagues (peer effects).

[12] Bandiera et al. (2010) estimate the effects of social ties on those working near each other, in a piece rate setting where any externalities between workers are purely social. They find that workers adjust their effort upward or downward to more closely match the productivity of those with whom they have close social ties. In a more laboratory-like setting, Falk and Ichino (2006) find evidence of peer effects even in the absence of social ties. Though the incentives in their set up offer no rewards for cooperation, peer effects increase overall output by raising the productivity of the least productive workers.

[13] Gaechter et al. (2010) show that, in a laboratory experiment, peers' wages and effort choices affect one's own effort choice, but only when these are observable. Jakiela and Ozier (2012) provide evidence that redistribution norms deter profitable investments in a rural developing country setting.

[14] Specifically: *nanyati* (light brown or red with stripes), *zoyara* (small white), *khaki* (beige), *zofira* (small red), *phalombe* (large red), *napilira* (red with white stripes), *zosakaniza* (mixed / other) and discards (e.g. rotten, soybeans, stones, etc.). The categories are derived from discussions with purveyors of sorted beans in the Lilongwe market.

study for several reasons: it is a familiar, common task for *ganyu*, typically compensated by piece rates; output has clear quantity and quality dimensions; it is a task where output can respond strongly to effort (in this case, focus and concentration) but effort is not physically taxing.

3.1 Experimental design

Subjects[15] are first invited to a "day zero" training session at which the task is explained and they are shown examples of the categories of beans.[16] Then, on each of the next four days, we obtain each participant's reservation piece rate $\underline{PR}_i$ (truthful revelation is incentive-compatible in our design, as discussed in Section 3.1.1 below) and make a randomized piece rate offer PR_i, which determines whether the participant is hired ($PR_i \geq \underline{PR}_i$) and the piece rate, if hired, per unit (PR_i). Workers who are hired work for the remainder of the day, about six hours on average. We measure output q_i as the number of units ("scoops", approximately 800g) sorted in a six-hour day. We also record a quality measure Q_i, the number of errors in a random sample of beans from a category. A randomized monitoring treatment, described below, explores workers' multitasking problem (quantity vs. quality) and the impact of rewarding output quality on the tradeoff between quantity and quality.

3.1.1 Randomization and the Becker-DeGroot-Marschak Mechanism

We use the Becker-DeGroot-Marschak mechanism (BDM) to uncover reservation piece rates, determine who works and set the piece rate. In BDM, the participants first states her reservation piece rate, $\underline{PR}_i$. A piece rate PR_i is then drawn at random from a jug. If the random draw is less than the reservation piece rate, i.e. $PR_i < \underline{PR}_i$, the participant is not hired. If the random draw is at least as high as the reservation piece rate, i.e. $PR_i \geq \underline{PR}_i$, then the participant is hired at a piece rate of PR_i. Using BDM provides two key advantages. First, by breaking the link between the stated reservation piece rate and the actual piece rate paid, it makes truthful revelation of minimum willingness to accept (WTA) the dominant strategy for the participant.[17] Second, it creates random variation in the actual piece rate

[15] Throughout, we refer to those with whom we interact at any stage as *subjects*, those who are present at the beginning of the work day and wish to participate as *attendees,* those who participate in BDM as *participants*, and those hired to work as *workers.* Not all attendees are participants because participation was capacity constrained. When this constraint was binding, participation was decided by lottery. See Section 3.2 for details.

[16] We also provide subjects with visual aids during the sorting process, including examples of each of the sorted bean categories.

[17] The work activity was conducted on four consecutive days in each village, giving subjects the opportunity to participate in the BDM exercise on multiple days. This could present a problem for the incentive-compatibility of BDM. In its traditional use to measure willingness to pay for products, the option to play

paid to workers with identical reservation piece rates. That is, two participants with the same reservation piece rate, $\underline{PR}_i = \underline{PR}_j$, can face different actual piece rates, $PR_i \neq PR_j$, and this difference will be determined purely by chance. This random variation allows us to isolate the causal effect of the piece rate on productivity.[18]

We implement a simplified version of BDM, in which a surveyor presents an individual participant with a menu of 5 piece rates: $5, 10, 15, 20, 25$ MWK per unit sorted.[19] The participant indicates which of the rates she will accept, the lowest of which we record as her reservation piece rate.[20] She then draws the actual piece rate offer from a uniform distribution with the same support as the reservation piece rates. Her draw determines whether she will work, and if so at what rate.[21]

The table below shows the possible outcomes of the game, with reservation piece rates in rows and piece rate offers in columns. The matrix is upper triangular because outcomes are only observed for participants who draw a piece rate at least as high as their reservation piece rate.

BDM multiple times could lead the subject to bid below her true WTP in early rounds. However, BDM is still incentive-compatible if decisions are independent across days. This would not be the case if, for example, the work was very physically demanding and effort on one day affected one's disutility of effort the next day. Another violation would occur if there were income effects, i.e. working one day increased NPV lifetime earnings appreciably and led to more consumption of leisure. We do not believe either of these are present in our current context: the work was by design not physically taxing, and these earnings are not large enough to plausibly affect willingness to work in a neoclassical model.

[18]Berry, Fischer, and Guiteras (2011) emphasize a third benefit of BDM: the ability to estimate heterogeneous treatment effects. Chassang et al. (2012) provide theoretical foundations, placing BDM in the class of "selective trials."

[19]All figures are in Malawi Kwacha. At the time of the study, the official exchange rate was roughly 150 MWK per US dollar. Given the price premium for sorted beans on the market and abstracting from the costs of hiring and monitoring workers, an employer would find it profitable to hire workers to sort beans at piece rates up to 40 MWK per unit sorted.

[20]To be precise, she reveals a range on her reservation piece rate. For example, if she indicates that 15 MWK is the lowest rate she will accept, the her true reservation piece rate is in the interval $(10, 15]$. We believe this loss in resolution is more than outweighed by the gain in simplicity.

[21]This description of the implementation of BDM is simplified. In practice, the surveyor leads the subject through a series of checks designed to confirm that the subject is indeed willing to work at the rates she says she will accept, and indeed prefers not working to working at the rates she declines. Our complete script in English is provided in the Supplementary Materials. All subjects attend a training session prior to BDM implementation in which the surveyors perform a skit with several examples designed to communicate the incentive-compatibility of BDM. The BDM decisions are elicited in private, so only the participant and the interviewer know her piece rate, unless she chooses to reveal it. Of course, whether or not she works is observed by everyone.

$\underline{PR}_i$	PR_i				
	5	10	15	20	25
5	$(5,5)$	$(5,10)$	$(5,15)$	$(5,20)$	$(5,25)$
10		$(10,10)$	$(10,15)$	$(10,20)$	$(10,25)$
15			$(15,15)$	$(15,20)$	$(15,25)$
20				$(20,20)$	$(20,25)$
25					$(25,25)$
> 25					

Without knowledge of the reservation piece rate, differences in productivity across piece rates (columns) are confounded with differences in productivity across workers with different reservation piece rates (rows). The benefit of BDM is the ability to make comparisons of outcomes across rows and down columns. A comparison across a row shows the causal effect of the piece rate, holding the reservation piece rate constant. A comparison down a column shows the association between the reservation piece rate and output, holding the actual piece rate fixed. Since we can only observe individuals working at or above their reservation piece rates, the number of comparisons that can be made varies. For example, we will have a lot of information in the relationship between the piece rate and output for those with very low reservation piece rates (row 1), but none for those with very high reservation piece rates (row 5). Conversely, we will obtain no information on the association between the reservation piece rate and productivity when the actual piece rate is very low (column 1), but we will have a lot of information on this association when the actual piece rate is very high (column 5). This will limit our ability to conduct fully nonparametric, cell-by-cell analysis – without a very large sample, some functional form assumptions will be necessary.

3.1.2 Output quality versus output quantity

A higher piece rate gives a worker a clear incentive to work faster. However, sheer quantity is not the only desired outcome: incorrect sorting of beans lowers the value of the final product. To investigate this tradeoff between quantity and quality, we randomize a monitoring treatment that increases workers' incentives to produce quality output.

Quality is measured by recording the error rate. In both the monitoring and no monitoring treatments, two randomly determined categories of beans were checked for errors each time a worker presented a sorted unit. Possible errors include mis-categorized beans, flawed beans (with holes or rotten areas), or other foreign materials. The number of errors for each of the checked categories was recorded for each unit sorted, and the categories for evaluation were re-randomized (with replacement) for each unit.

In addition to measuring this quantity-quality relationship, we are interested in learning how this relationship changes when we make the workers' pay dependent on quality. We randomly assigned half of the subjects each day, stratified by gender, to a monitoring treatment. Subjects assigned to monitoring were told before stating their minimum WTA that each unit of sorted beans would be checked for quality. The procedure (both as implemented and as described to the subjects) was that two categories of beans would be randomly selected and then a quantity equal to the size of a small handful from each category would be checked for errors. A unit was accepted if two or fewer errors were detected in each sample, and rejected if three or more errors were detected in either sample. Workers were not told and could not observe which category was being evaluated, and the category was randomly assigned for each unit. If either sample failed, the workers were required to return to their workstation to correct errors. Upon resubmission, two categories were randomly selected again (with replacement of the original categories) and the procedure repeated. This acted as a time tax on carelessness, since they were not given a new unit of beans until the unit under consideration was approved. The monitoring and and no monitoring groups were physically separated to the extent possible during the day to reduce the salience of monitoring to the non-monitoring group. To reduce Hawthorne effects, the checks for workers not assigned to quality monitoring were performed after the worker received her next unit of beans and returned to her workstation to continue sorting.

3.2 Implementation

The experiment was implemented in 12 villages in six districts in Central Malawi over a period of six weeks in the low labor demand season (July-August) and a second six week period during the high labor demand season (January-February). In each of the six districts, a list of 12 or more suitable villages was obtained from a District Agriculture Extension Officer.[22] We then randomly selected 2 villages from each district, one for implementation during the low labor demand season and a second during the high labor demand season. The village was informed of the activities approximately one week in advance and an open invitation was issued to attend the orientation and training session on a Monday afternoon. Subjects who participated in the orientation session were registered and became eligible to participate in the subsequent days' activities.

During the orientation session, the bean sorting task was explained and surveyors performed a skit to illustrate the BDM mechanism and show subjects that truthful revelation

[22]The villages were identified as locations where the collaborating NGO was not working. They were also selected on a number of characteristics, including distance from the district capital and distance from the road since these factors are likely to affect the functioning of labor markets in these villages.

of their minimum WTA was their best strategy. Subjects were informed that they would receive a participation fee of 50 MWK for each day they participated, plus their earnings from the day's work. The participation fee was emphasized to minimize self-selection into the experiment on subsequent days: we wanted to draw as representative as possible a sample of the village population.[23] Because of field capacity constraints, we limited the number of BDM participants on each day to 50. After the first three weeks of the first data collection period, the number was reduced to 40 to address implementation challenges caused by the high acceptance rates of even low piece rate offers. On a given work day, if more than 40 (50) of subjects arrived by the pre-specified start time, a lottery was conducted to select 40 (50) participants. Those who were not selected were compensated for their time with a bar of soap. This constraint was often binding: on average, 52.9 (s.d. 20.9) potential subjects arrived on time and were eligible to participate in the lottery if there was one (48.5 (s.d. 10.7) in the low season and 57.3 (s.d. 27.1) in the high season). A lottery was required on 15 of 24 days of the experiment in the low labor demand season, and on all 24 days in the high labor demand season.

For each subject who attended the initial afternoon training session, we observe attendance decisions for every subsequent work day, for a total of four attendance observations per individual. Conditional on attending in a given day and being selected to participate in BDM and the survey, we also observe her reservation piece rate.[24] Participants whose BDM draw was greater than or equal to their stated reservation piece rate received a contract. For contracted workers, we observe the number of bean units that a worker sorts, the quality for every unit sorted, and her seating location relative to other workers.

A short survey was administered to every participant to collect basic covariates, in particular those likely to be associated with the opportunity cost of time.[25] The participation fee was contingent on the participant completing the survey.

[23]If participants were not selected to work, they were free to depart immediately. The majority of BDM selections occurred before 10:00 AM leaving the participant the rest of the day for alternative activities (e.g. home production, working on own farm, other casual labor).

[24]Individuals who participated in BDM in a previous session were given priority to maximize the balance within the panel of observations. This priority status did not depend on whether they received a contract.

[25]Survey data were collected in two parts. The first, more comprehensive part, covering basic demographics and other time-invariant variables, was conducted only once with each participant. That is, a subject who was selected to participate on a given day was not administered this part of the survey if she had participated (and therefore been surveyed) on a previous day. The second part was a brief set of questions on the subject's potential alternative activities for that day. In both cases, the survey was conducted independent of the outcome of the BDM experiment. However, for logistical reasons, both were administered *after* the BDM experiment was conducted and the results were known, so it is possible that the responses were affected by the result of the experiment.

3.3 Descriptive statistics

3.3.1 Characteristics and participation

Characteristics of participants are described in Table 1, which breaks the sample into the low and high seasons (six weeks per season). There were 689 total participants, 355 in the low labor season and 334 in the high season. Individuals could work multiple days of the week, which results in an unbalanced individual panel by day with 1875 observations, 1005 in the low season and 870 in the high season.

Selected individual survey measures gathered for participants in BDM bidding are shown in Table 1.[26] Over 60 percent of the sample is female and between 20 and 30 percent are from female-headed households. Effectively all participants work in agriculture, and approximately two-thirds of households grow beans. Close to 40 percent of the sample report performing some casual labor (*ganyu*) the previous week, and conditional on any *ganyu*, the mean is 3.8 days. A number of characteristics vary significantly with the season. Most notably, the daily wage reported for the most recent casual labor is significantly higher in the high labor demand season. Individuals who join during the high season report slightly fewer months per year of food shortage, suggesting that they are better off than participants in the low season.[27] In the high season, workers are less likely to list housework as one of their alternative activities for the day and more likely to list working their own land.

Several factors may contribute to the observed differences across labor seasons. First, the underlying characteristics of the villages visited may differ across seasons. Although our villages were randomly assigned to season, given our small number of villages (12) we cannot appeal to the law of large numbers to argue that the villages are likely to be well-balanced. Second, different types of individuals may have selected into the study, explaining differences in average participant age or other income sources. The fairly generous participation fee (50 MWK) was in part intended to mitigate this sort of selection. Finally, seasonal variation in labor demand and productive activities may explain differences in reported casual labor wages and outside options on the day of data collection.[28]

Table 2 provides descriptive statistics on participation, BDM outcomes, and work outcomes. In Panel A, we summarize attendance and participation rates overall (column 1) and by day (columns 2-5), by labor season (columns 6 and 7), and by participant gender

[26] Summary statistics on a broader set of survey measures are reported in Table S1 of the Supplementary Materials.

[27] Households are more likely to have run out of food in January (high season) than in July (low season), which suggests that this difference is not due to the salience of food shortages during food short months.

[28] These explanations are not mutually exclusive. For example, differences in income sources may be due both to self selection and underlying differences in the villages. Because of the difficulty distinguishing among them, we do not emphasize direct comparisons of results across labor season.

(columns 8 and 9). On average, the number of attendees is increasing through the week, with more attendees during the high season. The average share of registered subjects attending each day is lower for the high season, suggesting fewer repeat workers during this period. Individuals in the low labor season work an average of 2.8 days while individuals in the high season work an average of 2.6 days out of the possible 4 work days.

We also collected village rosters to determine the share of invited households in a village that attended the study session. In both the high and low labor demand season, the probability of receiving an invitation was around 85 percent. Conditional on receiving an invitation, around 25 percent of individuals attended the orientation session in the low labor demand season, versus around 48 percent in the high labor demand season ($p < 0.001$, after controlling for district fixed effects). The probability of attending, conditional on receiving an invitation, is about 5 percent higher for females than for males, though it does not differ significantly for males and females by labor season.

3.3.2 Willingness to accept

Panel B of Table 2 provides summary statistics on behavior in BDM. The first row shows the mean minimum WTA revealed in BDM, for the same categories as Panel A, and additionally by monitoring treatment (columns 10 and 11). The salient facts are that mean minimum WTA falls after the first day, and the mean minimum WTA for women is approximately 2 MWK lower than for men. We do not observe significant differences by season or by monitoring treatment. Figure 1 shows the share of participants accepting each of the 5 piece rates. The most striking fact is that most participants are willing to accept very low piece rates: over 60 percent of participants accept a piece rate of 10 MWK per unit, for which expected daily earnings would be approximately 70 MWK.[29] This is consistent with the high willingness to accept low daily wages observed by Goldberg (2011).

The bottom two rows of Panel B summarize "mistakes" in the BDM procedure. Very few participants ($<$ 3 percent) refused a drawn price that they had accepted in their BDM decisions. A larger share (13 percent) state ex-post that they would have been willing to accept a drawn rate that they had rejected in their BDM decisions. The ex-post refusal rate declines throughout the week, consistent with participants learning that stating one's true minimum WTA was their best strategy. It also declines across weeks (noisily, not reported), which suggests that surveyors improved at communicating the optimal strategy to participants. Of course, the participant's statement that she would have been willing to accept at a previously rejected rate is purely hypothetical and individuals may have wished

[29]Workers sorted an average of 7.35 units per day (Table 2). Workers reported that they expected to sort an average of 6.74 units per day (Table S1).

to express a willingness to work in their responses to this non-binding question.

3.3.3 Quantity and quality of worker output

The primary measures of productivity, number of units sorted per day (q) and average number of errors per unit (Q), are summarized in Panel C of Table 2.[30] The mean number of units sorted per day across all days is 7.35 (s.d. 1.97), which is increasing throughout the week, and the mean number of errors per unit is 1.88 (s.d. 1.01). The quantity of output is lower (0.59 fewer units per day) and the quality of output is higher (0.66 fewer errors per unit) in the monitoring treatment, suggested that workers sorted more carefully and therefore more slowly in the monitoring treatment. Females sort 0.76 more units per day than men, and commit slightly fewer errors per unit (0.16). This co-movement of quantity and quality is observed for several covariates, consistent with our model's single productivity parameter for quality and quantity.[31]

4 Empirical Results

We present our empirical strategy and results together, by theme. Our three outcome measures are minimum WTA as measured by BDM, quantity of output measured by the number of units of beans sorted per day, and quality of output measured by the number of errors per unit. We first discuss selection, i.e. the relationship between minimum WTA and productivity. Second, we estimate incentive effects, i.e. the causal effect of piece rates on output. Third, we estimate differences between men and women in selection and incentive effects. Finally, we test for peer effects.

4.1 Selection and productivity

In this subsection, we explore the relationship between reservation piece rates and productivity. First, we examine how well covariates predict minimum WTA. Second, we estimate the relationship between minimum WTA and output, which we interpret as the selection channel of of the relationship between piece rates and productivity.

To study the predictors of reservation piece rates, we regress minimum WTA on characteristics of the market, specifically, indicators for the labor season (Peak = 1 for high

[30] Q is recorded the first time the workers bring a unit of sorted beans to the enumerator, before they have been instructed to correct any errors above the threshold.

[31] See Table S3 in the Supplementary Materials, which reports the pairwise correlation between outcomes (WTA, quantity and quality) and survey measures.

season), the monitoring treatment, M_{id}, whether the participant is female, F_i, and day of the week, DoW_d:

$$\text{MinWTA}_{id} = \phi\text{Peak}_i + \lambda M_{id} + \zeta F_i + \tau\text{DoW}_d + \delta\text{Dist}_i + \varepsilon_{id}. \tag{5}$$

Table 3 shows estimation results in cross section (columns 1 and 2), with random effects (columns 3 and 4) and with individual fixed effects (column 5).[32] All columns include district fixed effects (Dist_i). Minimum WTA is slightly lower in the high season (columns 1 and 3). This appears to be driven in part by selection into the study: the relationship weakens when we include individual controls (columns 2 and 4), suggesting that participants in the high season had covariates associated with lower minimum WTA. The monitoring treatment does not increase minimum WTA – in fact, the coefficient is negative across specifications, although significant only in one. This is somewhat surprising, in that subjects do not demand greater compensation for the more stringent standards imposed by monitoring. Minimum WTA falls over the course of the week, which cannot be explained solely by selection given the robustness to individual fixed effects (column 5). Minimum WTA is about 1 MWK higher on the first day than on later days in the week, relative to a mean of 10 MWK in the sample.

Next, we investigate selection: the association between output (quality and quantity) and minimum WTA. To isolate the selection channel, we estimate the relationship between minimum WTA and our two outcome measures, controlling for the actual piece rate received by the worker:

$$\begin{aligned} y_{id} \quad = \quad & \phi\text{Peak}_i + \lambda\text{M}_{id} + \alpha\text{PR}_{id} + \psi\left(\text{PR}_{id} \times \text{M}_{id}\right) + \beta\text{minWTA}_{id} \\ & + \zeta F_i + \tau\text{DoW}_d + \delta\text{Dist}_j + \varepsilon_{id}. \end{aligned} \tag{6}$$

We interpret the coefficient β as selection: the relationship between the reservation piece rate and productivity, holding the actual piece rate constant. Tables 4 and 5 show the effects of relationship between WTA and the number of scoops sorted per day and the number of

[32]Throughout the paper, we view random effects estimation as preferred, since we only make causal claims about variables that are randomized (monitoring and actual piece rate), and therefore are orthogonal to any time-invariant unobservables. Using random effects allows us to estimate non-causal relationships between outcomes of interest and time-invariant observables (e.g. worker gender), which would otherwise be absorbed by fixed effects. Furthermore, when we estimate relationships between productivity and minimum WTA, fixed effects models discard any cross-worker variation in minimum WTA, and are instead estimating the relationship between productivity and within-worker day-to-day fluctuations in minimum WTA. While this relationship may be of interest in some contexts, it is not of primary interest here. Nevertheless, as a robustness check, we also report estimates from fixed-effects models for time-varying observables of interest, and the results are generally similar.

errors per scoop, respectively.[33] With respect to quantity, we observe *negative* selection: after controlling for the worker incentive provided by the piece rate (PR_{id}), the monitoring treatment and the day of the week, minimum WTA is negatively related to quantity of output (Table 4), though the size of the coefficient is small (a 10 MWK increase in minimum WTA lowers the number of units sorted per day by 0.20-0.30, relative to a mean of 7.4 units). The same specification with number of errors per scoop as the dependent variable shows no significant relationship between minimum WTA and quality of output, and the coefficient on minimum WTA is inconsistently signed (Table 5).[34]

4.2 Incentives and productivity

By controlling for minimum WTA, we can isolate the direct effect of incentives on productivity. These effects are due solely to changes in the worker's effort choice in response to a change in the piece rate or monitoring of output quality, both of which are randomized.[35]

4.2.1 Piece rate

Since BDM reveals the reservation piece rate for the worker, as well as randomly assigning an actual piece rate, it is straightforward to test the effect of incentives on productivity. The estimate comes from Equation (6) above: since we can control for minimum WTA, and the piece rate is (conditionally) random, we can interpret the coefficient α causally as the incentive effect of the piece rate in the absence of incentives for quality. Below, we examine how the response to the piece rate changes in the presence of quality monitoring (ψ).

Table 4 shows the effect of the piece rate on quantity of output, controlling for the worker's reservation piece rate (columns 2 - 4) and individual characteristics (column 4). Increasing the piece rate by 10 MWK increases the number of scoops sorted per day by between 0.24 and 0.50 units, relative to a mean of 7.4 units. Going from the lowest piece

[33]While we present and discuss quantity and quality results separately, it is important to remember that they are jointly determined by the worker. That is, we should not think of determinants of quantity as operating with quality held fixed, nor vice versa.

[34]The preceding regressions impose a linear functional form. As an alternative, we estimate a more flexible model with indicators for each interval of minimum WTA and piece rate. The resulting semi-parametric relationship between minimum WTA and output is plotted in Figures S1 (quantity) and S2 (quality) in the Supplementary Materials. The results are generally similar to the linear specification: a slightly negative relationship between minimum WTA and quantity of output at low piece rates, which becomes insignificant at higher piece rates, while the relationship between minimum WTA and quality of output is insignificant at every piece rate category.

[35]Although monitoring is randomized, it is not random conditional on minimum WTA, since BDM participants announced their minimum WTA knowing whether or not they were assigned to the monitoring group. As noted in Section 4.1, we do not observe that being assigned to monitoring affects stated minimum WTA significantly.

rate (5 MWK) to the highest piece rate (25 MWK) increases output by between one-half to one unit per day. These effects are similar in magnitude to the selection effects described above, but with the opposite sign. The quantity of output is also increasing with the day of the week, an effect that is robust to the inclusion of individual fixed effects (column 5). In the high season, workers sort almost half a scoop more per day.[36]

Table 5 shows the effect of the piece rate on quality of output, measured by the number of errors per scoop of sorted beans. The piece rate appears to have little direct effect on quality of output, though the coefficient is consistently positive indicating that errors may be increasing in the piece rate. The number of errors per scoop is decreasing in the day of the week, consistent with individuals gaining experience with the task. This effect is robust to the inclusion of individual fixed effects (column 5), suggesting that it is not driven by changes in the composition of workers over the course of the week. The labor demand season does not appear to affect quality of output.

4.2.2 Monitoring

The effect of stricter monitoring on quantity and quality of output is measured by λ in Equation (6) above. If quantity and quality are substitutes, then workers must choose to allocate effort toward quantity or toward quality (reduce errors). If this is the case, λ will take on the same sign for the two output regressions (quantity and errors).

The direct effect of the monitoring treatment on quantity of output is shown in Table 4.[37] The coefficient on monitoring is negative and significant, lowering output by between -0.56 and -0.78 scoops per day, or about a third of a standard deviation. The loss in quantity of output is accompanied by a reduction in the number of errors per scoop, as shown in Table 5. The coefficient on monitoring is between -0.63 and -0.77 scoops per day, or around three-quarters of a standard deviation. Monitoring does appear to divert effort toward output quality at a cost of some quantity. As a more flexible alternative, Figure 2 plots estimates from a semi-parametric model in which the effect of each piece rate interval, interacted with the monitoring treatment, is estimated separately, controlling as in Equation (6) for WTA and indicators for female, day of week, high season, district and a level effect of

[36] As a robustness check, we also estimated a semi-parametric model with indicators for each piece rate, interacted with the monitoring treatment, controlling for minimum WTA and an indicator for female. The coefficients on each piece rate category are plotted in Figure S3 in the Supplementary Materials. The conclusions are generally similar to the simple linear specification of Equation (6), with output increasing with the piece rate draw, although the estimates are not very precise.

[37] By "direct," we mean holding selection constant by conditioning on minimum WTA. However, given that minimum WTA does not appear to respond to the monitoring treatment, this likely is a close approximation to the total effect.

monitoring.[38] The results are generally similar to the linear specification: monitoring clearly reduces errors per unit, and there does not appear to be a strong interaction with the piece rate.

4.3 Gender differences

To examine differential selection and effort choices by gender, we repeat the analyses and interact key regressors with a dummy variable indicating that the participant is female. Differences in reservation piece rates are obtained by re-estimating Equation (5) with interactions of the female variable with the labor season (Peak) and monitoring treatment (M). Table 6 shows the results for selection by gender. Women, but not men, display lower minimum WTA in the high season (columns 1 and 3). The strength of this relationship is reduced for both women and men when including individual controls (columns 2 and 4), but much more so for women, both in absolute and relative terms, suggesting very strong selection for women on covariates negatively associated with minimum WTA. Monitoring is significant in just one specification (for men), and with a perverse sign, since monitoring should not reduce minimum WTA.

Differences in productivity, controlling for reservation piece rates, are obtained by re-estimating Equation (6) allowing the effects of monitoring, minimum WTA, the piece rate and the labor season to vary by gender. Tables 7 and 8 show the effects on the quantity of output and quality of output respectively.

Monitoring reduces the quantity of output for both genders, but more so for females (Table 7). The pure incentive effect of the piece rate on quantity of output is similar across genders. However, the selection effect of minimum WTA on quantity of output does vary by gender. Men who exhibit a higher minimum WTA rate sort significantly fewer units of beans (a 10 MWK increase in the reservation piece rate is associated with sorting 1/3 to 1/2 fewer units per day); among women, the relationship is inconsistently signed and significant in only one specification. Thus, the negative selection described in Section 4.1 is driven entirely by the men in the sample. This is plausible if women's outside options are more similar to bean sorting, while men's outside options depend more on physical capacity, as in the returns to brawn in Pitt, Rosenzweig, and Hassan (2012). Men produce more output in the high season, approximately one-half unit per day, while women do not. This difference may also be related to the differences in men and women's outside options and how they vary with the labor season, or to differences in the subject pool in the low and high labor demand seasons.

[38]The full set of regression coefficients are provided in Table S4 of the Supplementary Materials.

Both men and women are similarly responsive to monitoring in their allocation of effort toward output quality (Table 8). The incentive effect of the piece rate has a small positive effect on the error rate for men and no effect for women. Minimum WTA is not associated with the error rate for either men or for women. Men do, however, reduce the number of errors per unit sorted in the high season, while error rates for women do not differ significantly across seasons.

4.4 Peer effects

In our design, workers self-select into work groups averaging 4 individuals. While these work groups are endogenously formed, workers in the same group experience different (random) piece rates.[39] That is, for worker i in group g, the piece rates of the other workers in her group, $\{\text{PR}_j\}_{i,j\in g,j\neq i}$, are random, after conditioning on the reservation piece rates of all workers in the group. We use this variation in piece rates to identify the effect of peers' compensation on a worker's own effort. There are potentially many ways that peer piece rates could affect a worker's output; we use a simple specification in which a worker's output depends on the mean piece rate of her peers, and this effect is allowed to vary flexibly with group size. More specifically, we regress y_{idg}, output from individual i in group g on day d, on a cubic interaction of N_{idg}, the number of workers in i's group on day d, and $\overline{\text{PR}}_{-i,g,d}$, the mean piece rate for other workers in the group, along with other controls:

$$\begin{aligned} y_{idg} &= \alpha\text{PR}_{id} + \beta\text{minWTA}_{id} + f\left(N_{idg}, \overline{\text{PR}}_{-i,g,d}\right) + \psi\overline{\text{minWTA}}_{-i,d,g} \\ &\quad +\nu N_{idg} + \gamma\text{M}_{id} + \delta\text{Dist}_{ig} + \tau\text{DoW}_d + \varepsilon_{idg}, \end{aligned} \tag{7}$$

where

$$\begin{aligned} f\left(N_{idg}, \overline{\text{PR}}_{-i,g,d}\right) &= \phi_0\overline{\text{PR}}_{-i,g,d} + \phi_1\left(\overline{\text{PR}}_{-i,g,d} \times N_{idg}\right) \\ &\quad +\phi_2\left(\overline{\text{PR}}_{-i,g,d} \times N_{idg}^2\right) + \phi_3\left(\overline{\text{PR}}_{-i,g,d} \times N_{idg}^3\right). \end{aligned}$$

To address the potential effect of reservation rates on the mean piece rate, we also control for the mean reservation rate of other workers in the group. Workers of different types may sort into groups of different sizes, so the N_{idg} variable should be interpreted as capturing the combined effect of group size and any unobserved differences in worker characteristics that vary with group size. Since the regressors of interest vary at the work group level, we compute standard errors robust to two-way clustering by work group g and individual i

[39] Workers are separated by monitoring treatment, and are not aware of others workers' wage piece rates when they choose where to sit.

(across days) (Cameron et al. 2011).

We find that workers in groups of two respond positively to the piece rate of their co-worker, but this effect decreases with group size.[40] Figure 3 plots marginal effects on output quantity from the model with a cubic interaction.[41] In a group of two workers, a one-MWK increase in the peer's piece rate results in an increase of one's own output of 0.036 (s.e. 0.016) units sorted per day. For every additional worker (beyond a minimum of one peer), in the work group, this effect decreases, and does so non-linearly. For groups of 3 to 5 workers, the peer effect is statistically indistinguishable from zero. For the largest relevant group size, the effect of a one-MWK increase in the average peer piece rate is negative, and lowers own productivity by 0.064 (s.e. 0.019) units per day. We do not find a similar effect on quality (Figure 4, Table S6), which suggests that workers are motivated to keep pace with their peers but do not sacrifice quality to do so. Overall, the results are inconsistent with either extreme fairness norms that depress the productivity of relatively less well paid group members or with redistribution norms that allow workers with lower compensation to expropriate the earnings of better paid peers.

5 Discussion

We implement a unique experimental design in casual labor markets in rural Malawi to measure selection and incentive effects, to observe gender differences in these markets and to test for other behavioral determinants of productivity. In our setting, production of quantity and quality are substitutes and workers allocate effort between these two types of output. A monitoring treatment shifts effort toward production of quality at a quantity cost.

The piece rate affects productivity through both the selection and incentive channels, but only the incentive channel improves productivity in our setting. Raising the piece rate significantly increases the quantity of output, controlling for workers' reservation rates, but does not reduce quality. Selection affects only quantity of output, not quality, and only for men. In fact, men display negative selection, with a higher piece rate associated with lower quantities. This relationship is stronger during the high season, when households are liquidity constrained and the opportunity cost of time is higher. Explicit incentives for output quality reduce the error rate in production, but do not affect worker selection into

[40] We observe 36 groups that include two workers, 83 with three, 211 with four, 41 with five and 10 with six workers. We only observe one group of seven workers form on one day and therefore omit it from our analysis of peer effects.

[41] Specifically, we plot $\hat{\phi}_0 + \hat{\phi}_1 N_{idg} + \hat{\phi}_2 N_{idg}^2 + \hat{\phi}_3 N_{idg}^3$, i.e. $\hat{\phi}_0 + 2\hat{\phi}_1 + 4\hat{\phi}_2 + 8\hat{\phi}_3$ for an individual in a group of size two (the smallest for which peer effects are relevant), $\hat{\phi}_0 + 3\hat{\phi}_1 + 9\hat{\phi}_2 + 27\hat{\phi}_3$ for a group of size three, etc. Table S5 in the Supplementary Materials provides detailed regression results.

the task. Finally, we see a nuanced picture of peer effects: peer compensation appears to encourage productivity in small groups, but discourage it in large groups.

Comparing the extensive (participation) versus intensive (effort) margins, it appears that participation is much more responsive to the piece rate than effort. Over the range of piece rates offered, the arc elasticity of participation with respect to the piece rate is 0.58, while the elasticity of output quantity with respect to the piece rate (controlling for the reservation piece rate) is 0.06. While monitoring lowers the level of output, it does not significantly affect the elasticity of output with respect to the piece rate, nor does it affect participation. (See Tables S7-S8 in the Supplementary Materials for details.) At the mean, introducing monitoring lowers output as much as a 30 MWK decrease in the piece rate. It takes workers in the monitoring treatment about 0.7 days longer to sort a 50 kg bag of beans than workers who are not being monitored. Well-sorted beans sell for up to 4,000 MWK more per 50 kg bag than unsorted beans, potentially justifying the time cost of monitoring.

The context in which the study takes place appears to shape several of the findings. Men and women in our setting face different outside options, which are more likely to involve hard manual labor for men and tasks like weeding or home activities for women, similar to the setting of Pitt et al. (2012). The negative selection in response to the piece rate demonstrated by men in our study is consistent with men who can make higher wages in manual labor being worse at detail oriented tasks, such as sorting beans. The available outside options also vary across the high and low labor demand seasons as do liquidity constraints. When the value of money is high, during the high labor demand season, workers have a lower WTA and are more responsive to incentives. Our study is limited in its ability to draw strong conclusions about the role of outside markets in shaping behavior within the experiment. Future work that generates exogenous variation in the value of the outside option would offer a more direct test of the hypothesis that labor market imperfections undermine sorting of workers based on productivity, with potential implications for efficiency and economic growth.

References

John M. Abowd, Francis Kramarz, and David N. Margolis. High wage workers and high wage firms. *Econometrica*, 67(2):251–333, 1999.

Nava Ashraf, James Berry, and Jesse M. Shapiro. Can higher prices stimulate product use? Evidence from a field experiment in Zambia. *American Economic Review*, 100(5): 2383–2413, 2010.

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

Oriana Bandiera, Iwan Barankay, and Imran Rasul. Social incentives in the workplace. *Review of Economic Studies*, 77(2):417–458, 2010.

Pranab K. Bardhan. Labor supply functions in a poor agrarian economy. *The American Economic Review*, 69(1):73–83, 1979.

Lori Beaman and Jeremy Magruder. Who gets the job referral? Evidence from a social networks experiment. *The American Economic Review*, 102(7):3574–3593, 2012.

Lori Beaman, Niall Keleher, and Jeremy Magruder. Do job networks disadvantage women? Evidence from a recruitment experiment in Malawi. Mimeo, 2013.

G. Becker, M. DeGroot, and J. Marschak. Measuring utility by a single-response sequential method. *Behavioral Science*, 9:226–236, 1964.

James Berry, Greg Fischer, and Raymond Guiteras. Eliciting and utilizing willingness to pay: Evidence from field trials in Northern Ghana. Mimeo, September 2011.

Nicholas Bloom, James Liang, John Roberts, and Zhichun Jenny Ying. Does working from home work? Evidence from a Chinese experiment. Mimeo, 2013.

A. Colin Cameron, Jonah B. Gelbach, and Douglas L. Miller. Robust inference with multiway clustering. *Journal of Business and Economic Statistics*, 29(2):238–249, 2011.

Sylvain Chassang, Gerard Padro i Miquel, and Erik Snowberg. Selective trials: A principal-agent approach to randomized controlled experiments. *American Economic Review*, 102 (4):1279–1309, 2012.

Jessica Cohen and Pascaline Dupas. Free distribution or cost-sharing? Evidence from a randomized malaria prevention experiment. *Quarterly Journal of Economics*, 125(1):1–45, 2010.

Thomas Dohmen and Armin Falk. Performance pay and multidimensional sorting: Productivity, preferences, and gender. *The American Economic Review*, 101(2):556–590, 2011.

P. Dupas and J. Robinson. The daily grind: Cash needs, labor supply and self-control. 2013.

Tor Eriksson and Marie Claire Villeval. Performance-pay, sorting and social motivation. *Journal of Economic Behavior and Organization*, 68(2):412 – 421, 2008.

Marcel Fafchamps, Måns Söderbom, and Najy Benhassine. Wage gaps and job sorting in African manufacturing. *Journal of African Economies*, 18(5):824–868, 2009.

Armin Falk and Andrea Ichino. Clean evidence on peer effects. *Journal of Labor Economics*, 24(1):39–57, 2006.

Ernst Fehr and Lorenz Goette. Do workers work more if wages are high? Evidence from a randomized field experiment. *The American Economic Review*, 97(1):298–317, 2007.

Simon Gaechter, Daniele Nosenzo, and Martin Sefton. The impact of social comparisons on reciprocity. Mimeo, May 2010.

Uri Gneezy, Kenneth L. Leonard, and John A. List. Gender differences in competition: Evidence from a matrilineal and a patriarchal society. *Econometrica*, 77(5):1637–1664, 2009.

Jessica Goldberg. Kwacha gonna do? Experimental evidence about labor supply in rural Malawi. Mimeo, September 2011.

J. Heywood, S. Siebert, and X. Wei. The consequences of a piece rate on quantity and quality: Evidence from a field experiment. 2013.

Pamela Jakiela and Owen Ozier. Does Africa need a rotten kin theorem? Experimental evidence from village economies. Mimeo, 2012.

Dean Karlan and Jonathan Zinman. Observing unobservables: Identifying information asymmetries with a consumer credit field experiment. *Econometrica*, 77(6):1993–2008, 2009.

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

Alexandre Mas and Enrico Moretti. Peers at work. *American Economic Review*, 99(1): 112–145, 2009.

Daniel S. Nagin, James B. Rebitzer, Seth Sanders, and Lowell J. Taylor. Monitoring, motivation, and management: The determinants of opportunistic behavior in a field experiment. *The American Economic Review*, 92(4):850–873, 2002.

Muriel Niederle and Lise Vesterlund. Do women shy away from competition? Do men compete too much? *The Quarterly Journal of Economics*, 122(3):1067–1101, 2007.

Harry J. Paarsch and Bruce S. Shearer. The response of worker effort to piece rates: Evidence from the British Columbia tree-planting industry. *The Journal of Human Resources*, 34 (4):643–667, 1999.

Mark M. Pitt, Mark R. Rosenzweig, and Mohammad Nazmul Hassan. Human capital investment and the gender division of labor in a brawn-based economy. *American Economic Review*, 102(7):3531–60, 2012.

Mark R. Rosenzweig. Rural wages, labor supply, and land reform: A theoretical and empirical analysis. *The American Economic Review*, 68(5):847–861, 1978.

Michael Sattinger. Assignment models of the distribution of earnings. *Journal of Economic Literature*, 31(2):831–80, June 1993.

Robert Shimer and Lones Smith. Assortative matching and search. *Econometrica*, 68(2): 343–369, 2000.

Joseph E. Stiglitz. Incentives, risk, and information: Notes towards a theory of hierarchy. *The Bell Journal of Economics*, 6(2):552–579, 1975.

Table 1: Descriptive Statistics for Participants

	All	Low Season	High Season	Diff.	
	(1)	(2)	(3)	(4)	
Number of participants	689	355	334		
Number of daily observations	1875	1005	870		
Female	0.665	0.690	0.638	-0.052	
	(0.472)	(0.463)	(0.481)	[0.036]	
Age	34.9	34.6	35.2	0.6	
	(13.6)	(13.2)	(14.1)	[1.1]	
Number of adults in household	3.11	3.16	3.06	-0.10	
	(1.68)	(1.59)	(1.77)	[0.13]	
Years of education	4.23	3.91	4.56	0.65	***
	(3.28)	(3.35)	(3.17)	[0.25]	
Female headed household	0.251	0.201	0.303	0.102	***
	(0.434)	(0.401)	(0.460)	[0.033]	
Participated in ganyu in last week	0.38	0.33	0.42	0.09	**
	(0.48)	(0.47)	(0.49)	[0.04]	
Days of ganyu last week, conditional on positive	3.76	4.22	3.38	-0.84	***
	(2.07)	(2.16)	(1.93)	[0.26]	
Daily wage from recent ganyu (MKW)	298.5	257.7	336.7	79.0	***
	(303.2)	(179.0)	(381.0)	[24.1]	
Household produces maize	0.999	1.000	0.997	-0.003	
	(0.038)	(0.000)	(0.055)	[0.003]	
Household produces beans	0.657	0.686	0.627	-0.059	
	(0.475)	(0.465)	(0.484)	[0.037]	
Typical per year months without adequate food	3.35	3.56	3.12	-0.44	***
	(2.27)	(2.34)	(2.16)	[0.17]	
Alternative activity: housework	0.180	0.267	0.074	-0.193	***
	(0.385)	(0.443)	(0.262)	[0.034]	
Alternative activity: other ganyu	0.206	0.235	0.172	-0.063	*
	(0.405)	(0.425)	(0.378)	[0.038]	

*, **, *** denote significance at 10%, 5% and 1%, respectively.

Notes: this table presents means of participants' characteristics during the low and high season, with standard deviations in parentheses, as well as differences in means, with the standard error of the estimated difference in brackets.

Table 2: Descriptive Statistics on Participation, BDM and Work Outcomes

	Day of Week					Season		Gender		Monitoring	
	All Days	Day 1	Day 2	Day 3	Day 4	Low	High	Male	Female	No	Yes
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Panel A: Attendance and Participation											
Number of attendees	52.9	47.3	52.1	52.9	59.3	48.5	57.3	15.5	37.4		
	(20.9)	(22)	(17.1)	(21.3)	(23.6)	(10.7)	(27.1)	(8.5)	(16.4)		
Number of participants	39.1	37.7	39.8	38.6	40.3	41.9	36.3	12.4	26.7		
	(7)	(8)	(6.1)	(8.1)	(6.1)	(5.9)	(6.9)	(5.9)	(6)		
Number of contracts awarded	30.5	27.7	32.3	30.3	31.8	31.8	29.3	8.9	21.7		
	(7.8)	(8.2)	(7.4)	(8.8)	(6.7)	(8.1)	(7.4)	(5.5)	(5.4)		
Proportion of registered workers attending	.474	.425	.467	.475	.531	.727	.367	.378	.53		
	[.007]	[.014]	[.014]	[.014]	[.014]	[.011]	[.008]	[.011]	[.009]		
Panel B: BDM											
Minimum WTA (MWK)	10.3	11.1	10.0	10.5	9.8	10.5	10.1	11.7	9.7	10.5	10.1
	(5.9)	(6.6)	(5.9)	(5.6)	(5.4)	(6.1)	(5.7)	(6.6)	(5.4)	(6.0)	(5.8)
Ex post refused contract	0.026	0.026	0.025	0.027	0.026	0.034	0.017	0.034	0.023	0.027	0.025
	(0.159)	(0.161)	(0.157)	(0.162)	(0.158)	(0.182)	(0.129)	(0.182)	(0.149)	(0.163)	(0.156)
Ex post would have accepted	0.126	0.233	0.112	0.061	0.072	0.176	0.054	0.109	0.138	0.128	0.124
	(0.333)	(0.425)	(0.318)	(0.241)	(0.260)	(0.382)	(0.227)	(0.313)	(0.346)	(0.335)	(0.330)
Panel C: Work outcomes											
Quantity: units sorted	7.35	6.02	7.21	7.90	8.12	7.15	7.57	6.81	7.57	7.65	7.06
	(1.97)	(1.61)	(1.74)	(1.97)	(1.87)	(1.81)	(2.12)	(1.97)	(1.93)	(2.06)	(1.85)
Quality: errors per unit	1.88	2.20	1.92	1.76	1.69	1.88	1.88	2.00	1.84	2.22	1.56
	(1.01)	(1.19)	(1.04)	(0.91)	(0.83)	(1.05)	(0.97)	(1.17)	(0.93)	(1.01)	(0.90)

Panel A: An attendee is defined as any subject who registers on the orientation day and is present at the beginning of a work day. A maximum of 40 attendees participate in BDM each day (50 in the first three weeks, see discussion in text). If participation is oversubscribed, 40 (50) of the attendees are selected by lottery for participation. Standard deviations in parentheses. Standard error of estimated proportion in brackets. Panel B: Sample is all participants in BDM. Minimum WTA is the participant's bid in BDM. Ex post refused contract indicates that the participant ultimately rejected a piece rate she had agreed to prior to the draw. Ex post would have accepted indicates that a participant who did not receive a contract, i.e. drew higher than her minimum WTA, stated in the exit survey that she would have accepted the piece rate drawn had she been given the opportunity. Standard deviations in parentheses. Panel C: Sample is all participants in BDM who received contracts. Standard deviations in parentheses.

Table 3: Determinants of Willingness to Accept

	Cross Section		Random Effects		Fixed Effects
	(1)	(2)	(3)	(4)	(5)
Peak Season	-0.780 **	-0.223	-0.901 **	-0.362	
	(0.377)	(0.399)	(0.386)	(0.408)	
Monitoring Treatment	-0.431	-0.546 **	-0.297	-0.333	-0.189
	(0.269)	(0.265)	(0.221)	(0.221)	(0.235)
Female	-2.321 ***	-2.662 ***	-2.635 ***	-2.875 ***	
	(0.449)	(0.484)	(0.453)	(0.490)	
Second day	-0.904 ***	-0.849 **	-0.998 ***	-0.914 ***	-1.040 ***
	(0.343)	(0.342)	(0.319)	(0.321)	(0.330)
Third day	-0.326	-0.303	-0.177	-0.149	-0.108
	(0.353)	(0.349)	(0.336)	(0.335)	(0.352)
Fourth day	-1.086 ***	-1.095 ***	-1.072 ***	-1.048 ***	-1.077 ***
	(0.351)	(0.352)	(0.339)	(0.340)	(0.353)
Indiv. Controls	No	Yes	No	Yes	
Mean Dep. Var.	10.32	10.32	10.32	10.32	10.32
SD Dep. Var.	5.899	5.899	5.899	5.899	5.899
Observations	1857	1857	1857	1857	1857

*, **, *** denote significance at 10%, 5% and 1%, respectively.

Notes: this table presents regressions of minimum willingness to accept (WTA) on season (peak labor demand), monitoring, whether the participant was female, and day-of-week fixed effects, with the first day as the omitted category. Columns (2) and (4) control for individual covariates. All regressions include district fixed effects, although in column (5) these are absorbed by the individual fixed effects. Standard errors robust to clustering at the participant level are in parentheses.

Table 4: Determinants of Quantity (Number of Units Sorted per Day)

	Random Effects								Fixed Effects	
	(1)		(2)		(3)		(4)		(5)	
Monitoring treatment	-0.564	***	-0.558	***	-0.782	***	-0.689	***	-0.769	***
	(0.066)		(0.064)		(0.195)		(0.198)		(0.195)	
Piece rate			0.031	***	0.024	***	0.023	***	0.025	***
			(0.005)		(0.007)		(0.007)		(0.007)	
Monitoring X Piece rate					0.013		0.007		0.012	
					(0.010)		(0.010)		(0.010)	
Minimum WTA	-0.006		-0.020	*	-0.020	*	-0.022	**	-0.026	**
	(0.010)		(0.011)		(0.011)		(0.010)		(0.011)	
High season	0.412	***	0.410	***	0.409	***	0.455	***		
	(0.130)		(0.131)		(0.131)		(0.131)			
Female	0.731	***	0.739	***	0.741	***	0.830	***		
	(0.139)		(0.139)		(0.139)		(0.142)			
Day 2	1.169	***	1.175	***	1.178	***	1.161	***	1.189	***
	(0.086)		(0.084)		(0.084)		(0.087)		(0.084)	
Day 3	1.847	***	1.847	***	1.848	***	1.820	***	1.871	***
	(0.091)		(0.091)		(0.091)		(0.093)		(0.091)	
Day 4	2.064	***	2.072	***	2.072	***	2.016	***	2.094	***
	(0.083)		(0.081)		(0.081)		(0.081)		(0.081)	
Indiv. Controls	No		No		No		Yes			
Mean Dep. Var.	7.350		7.350		7.350		7.350		7.350	
SD Dep. Var.	1.975		1.975		1.975		1.975		1.975	
Observations	1461		1461		1461		1461		1461	

*, **, *** denote significance at 10%, 5% and 1%, respectively.

Notes: this table presents regressions of quantity sorted (number of units) on whether the participant was assigned to the monitoring treatment, the piece rate the participant received, the interaction of monitoring and the piece rate, the minimum piece rate the participant was willing to accept (Minimum WTA), and day-of-week fixed effects, with the first day as the omitted category. All regressions include season and district fixed effects, although in column (5) these are absorbed by the individual fixed effects. Standard errors robust to clustering at the participant level are in parentheses.

Table 5: Determinants of Quality (Number of Errors per Unit)

	Random Effects				Fixed Effects
	(1)	(2)	(3)	(4)	(5)
Monitoring treatment	-0.628 ***	-0.627 ***	-0.670 ***	-0.722 ***	-0.696 ***
	(0.042)	(0.042)	(0.136)	(0.137)	(0.136)
Piece rate		0.008 **	0.007	0.006	0.005
		(0.004)	(0.006)	(0.006)	(0.006)
Monitoring X Piece Rate			0.002	0.006	0.004
			(0.008)	(0.008)	(0.008)
Minimum WTA	0.001	-0.003	-0.003	-0.003	0.000
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
High season	-0.064	-0.064	-0.064	-0.051	
	(0.057)	(0.057)	(0.057)	(0.066)	
Female	-0.171 **	-0.168 **	-0.168 **	-0.257 ***	
	(0.071)	(0.071)	(0.071)	(0.075)	
Day 2	-0.309 ***	-0.308 ***	-0.307 ***	-0.292 ***	-0.310 ***
	(0.076)	(0.076)	(0.076)	(0.076)	(0.076)
Day 3	-0.462 ***	-0.462 ***	-0.462 ***	-0.443 ***	-0.476 ***
	(0.073)	(0.073)	(0.073)	(0.072)	(0.072)
Day 4	-0.529 ***	-0.528 ***	-0.528 ***	-0.491 ***	-0.541 ***
	(0.073)	(0.073)	(0.073)	(0.073)	(0.074)
Indiv. Controls	No	No	No	Yes	
Mean Dep. Var.	1.883	1.883	1.883	1.883	1.883
SD Dep. Var.	1.013	1.013	1.013	1.013	1.013
Observations	1461	1461	1461	1461	1461

*, **, *** denote significance at 10%, 5% and 1%, respectively.

Notes: this table presents regressions of quality (number of errors per unit) on whether the participant was assigned to the monitoring treatment, the piece rate the participant received, the interaction of monitoring and the piece rate, the minimum piece rate the participant was willing to accept (Minimum WTA), and day-of-week fixed effects, with the first day as the omitted category. All regressions include season and district fixed effects, although in column (5) these are absorbed by the individual fixed effects. Standard errors robust to clustering at the participant level are in parentheses.

Table 6: Determinants of Willingness to Accept
Differential Effects by Gender

	Cross Section			Random Effects		
	(1)		(2)	(3)		(4)
High Season						
Among Men	-0.234		-0.113	-0.423		-0.268
	(0.790)		(0.764)	(0.793)		(0.772)
Among Women	-1.029	**	-0.282	-1.126	***	-0.411
	(0.415)		(0.449)	(0.427)		(0.463)
Monitoring						
Among Men	-0.887		-0.900	-0.622		-0.650
	(0.574)		(0.559)	(0.439)		(0.433)
Among Women	-0.223		-0.385	-0.155		-0.193
	(0.288)		(0.289)	(0.252)		(0.255)
Indiv. Controls	No		Yes	No		Yes
Mean Dep. Var.	10.32		10.32	10.32		10.32
SD Dep. Var.	5.899		5.899	5.899		5.899
Observations	1857		1857	1857		1857

*, **, *** denote significance at 10%, 5% and 1%, respectively.

Notes: this table presents differential effects by gender of season and monitoring. Additional regressors not reported are as in Table 3: whether the participant was female (i.e. the level effect), day-of-week fixed effects, district fixed effects, and, in columns (2) and (4), individual covariates. Standard errors robust to clustering at the participant level are in parentheses.

Table 7: Determinants of Quantity (Number of Units Sorted per Day)
Differential Effects by Gender

	(1)		(2)		(3)		(4)	
Monitoring								
Among Men	-0.325	***	-0.339	***	-0.452		-0.127	
	(0.125)		(0.121)		(0.364)		(0.343)	
Among Women	-0.659	***	-0.645	***	-0.914	***	-0.901	***
	(0.076)		(0.074)		(0.229)		(0.239)	
Piece rate								
Among Men			0.032	***	0.028	**	0.035	***
			(0.010)		(0.013)		(0.013)	
Among Women			0.030	***	0.022	***	0.019	**
			(0.006)		(0.008)		(0.008)	
Monitoring X Piece rate								
Among Men					0.006		-0.014	
					(0.018)		(0.017)	
Among Women					0.016		0.014	
					(0.012)		(0.013)	
Minimum WTA								
Among Men	-0.034	**	-0.049	***	-0.049	***	-0.049	***
	(0.017)		(0.018)		(0.018)		(0.018)	
Among Women	0.008		-0.006		-0.006		-0.009	
	(0.012)		(0.013)		(0.013)		(0.012)	
High season								
Among Men	1.013	***	1.020	***	1.024	***	0.956	***
	(0.225)		(0.226)		(0.226)		(0.210)	
Among Women	0.180		0.173		0.167		0.253	
	(0.156)		(0.157)		(0.157)		(0.155)	
Indiv. Controls	No		No		No		Yes	
Mean Dep. Var.	7.350		7.350		7.350		7.350	
SD Dep. Var.	1.975		1.975		1.975		1.975	
Observations	1461		1461		1461		1461	

*, **, *** denote significance at 10%, 5% and 1%, respectively.

Notes: this table presents differential effects by gender of whether the participant was assigned to the monitoring treatment, the piece rate the participant received, the interaction of monitoring and the piece rate, the minimum piece rate the participant was willing to accept (Minimum WTA), and season on the quality of output (number of errors per unit). Additional regressors not reported include whether the participant was female (i.e. the level effect), day-of-week fixed effects, district fixed effects, and, in column (4), individual covariates. Individual random effects in all specifications. Standard errors robust to clustering at the participant level are in parentheses.

Table 8: Determinants of Quality (Number of Errors per Unit)
Differential Effects by Gender

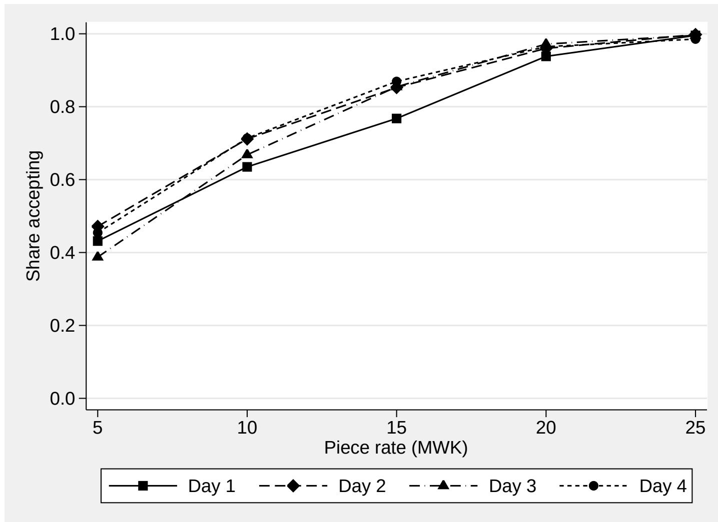
	(1)	(2)	(3)	(4)
Monitoring				
Among Men	-0.626 ***	-0.636 ***	-0.712 **	-0.761 **
	(0.089)	(0.090)	(0.298)	(0.300)
Among Women	-0.629 ***	-0.628 ***	-0.618 ***	-0.674 ***
	(0.047)	(0.047)	(0.148)	(0.150)
Piece rate				
Among Men		0.019 **	0.017	0.015
		(0.009)	(0.014)	(0.014)
Among Women		0.004	0.004	0.003
		(0.004)	(0.007)	(0.007)
Monitoring X Piece rate				
Among Men			0.004	0.010
			(0.017)	(0.017)
Among Women			-0.001	0.003
			(0.008)	(0.008)
Minimum WTA				
Among Men	0.004	-0.004	-0.004	-0.006
	(0.010)	(0.011)	(0.011)	(0.011)
Among Women	-0.001	-0.002	-0.002	-0.002
	(0.006)	(0.007)	(0.007)	(0.007)
Peak Labor				
Among Men	-0.237 *	-0.232 *	-0.230 *	-0.195
	(0.127)	(0.127)	(0.127)	(0.128)
Among Women	0.005	0.005	0.005	0.020
	(0.062)	(0.062)	(0.063)	(0.069)
Indiv. Controls	No	No	No	Yes
Mean Dep. Var.	1.883	1.883	1.883	1.883
SD Dep. Var.	1.013	1.013	1.013	1.013
Observations	1461	1461	1461	1461

*, **, *** denote significance at 10%, 5% and 1%, respectively.

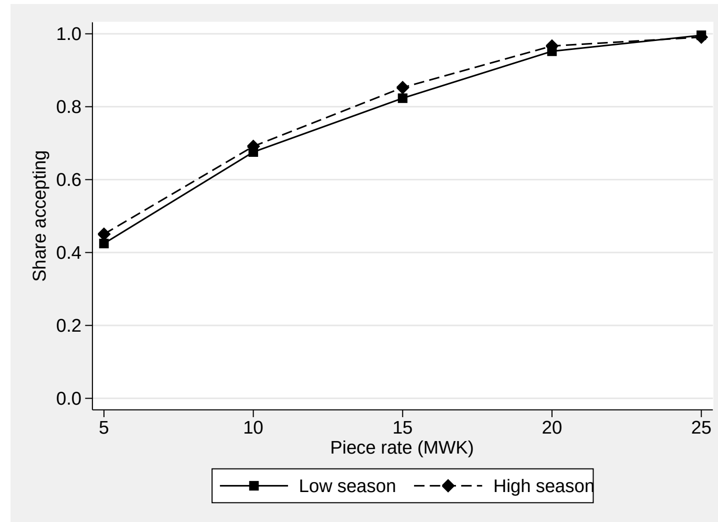
Notes: this table presents differential effects by gender of whether the participant was assigned to the monitoring treatment (Monitoring), the piece rate the participant received, the interaction of monitoring and the piece rate, the minimum piece rate the participant was willing to accept (Minimum WTA), and season on the quality of output (number of errors per unit). Additional regressors not reported include whether the participant was female (i.e. the level effect), day-of-week fixed effects, district fixed effects, and, in column (4), individual covariates. Individual random effects in all specifications. Standard errors robust to clustering at the participant level are in parentheses.

Figure 1: CDFs of minimum piece rate accepted

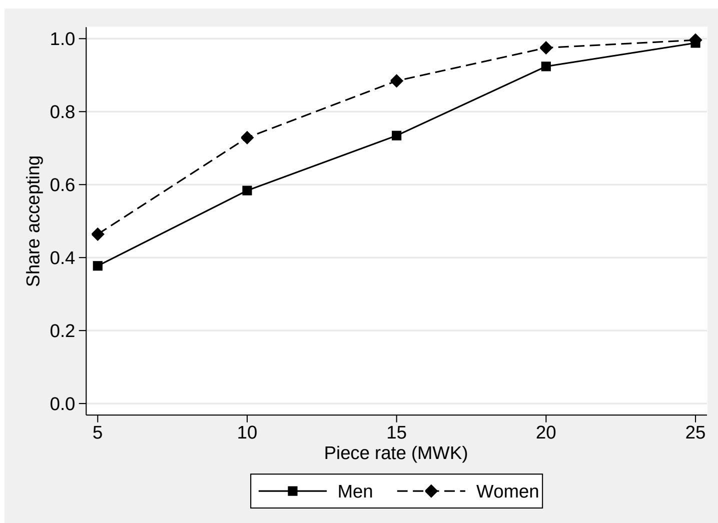
(a) By day of week



(b) By season



(c) By gender



(d) By monitoring treatment

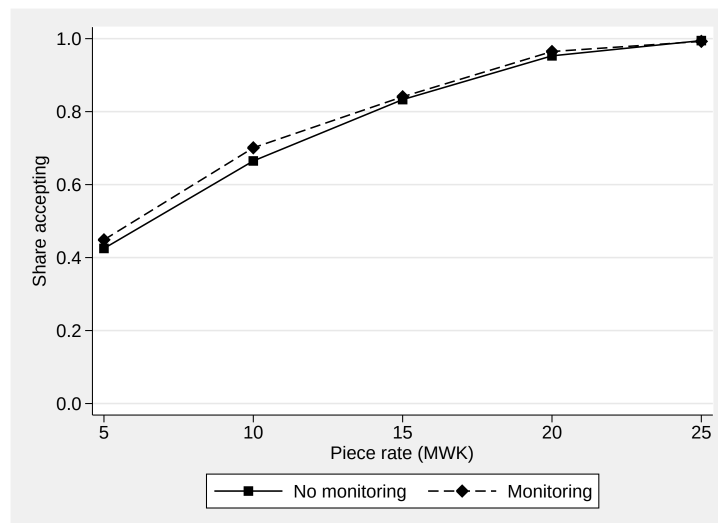


Figure 2: Errors per unit, by piece rate and monitoring treatment
Estimated coefficients

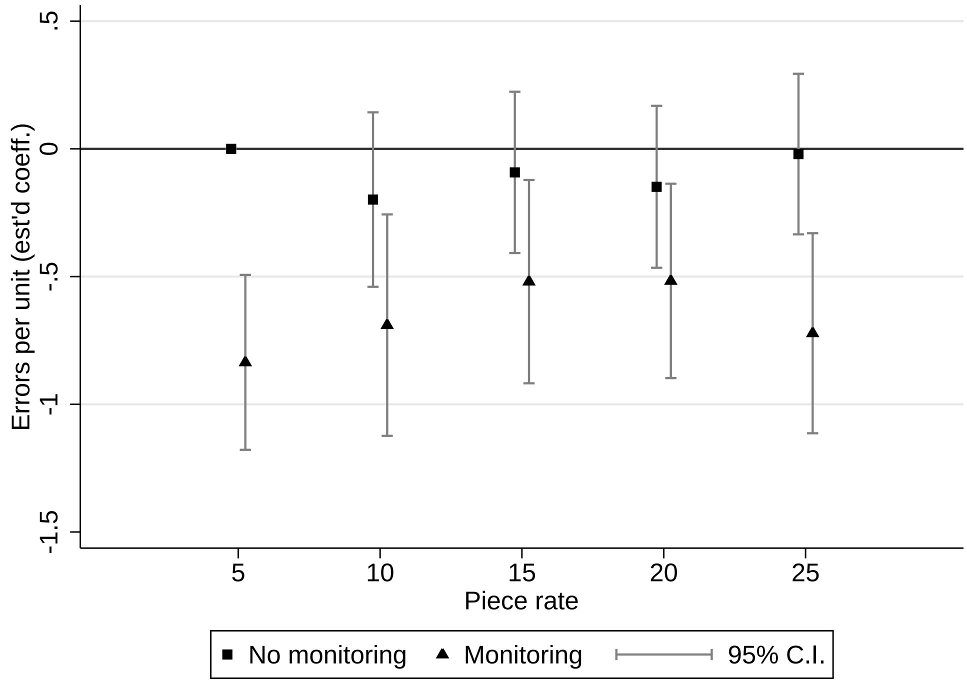


Figure 3: Peer effects: quantity
Marginal effect of mean piece rate, by group size

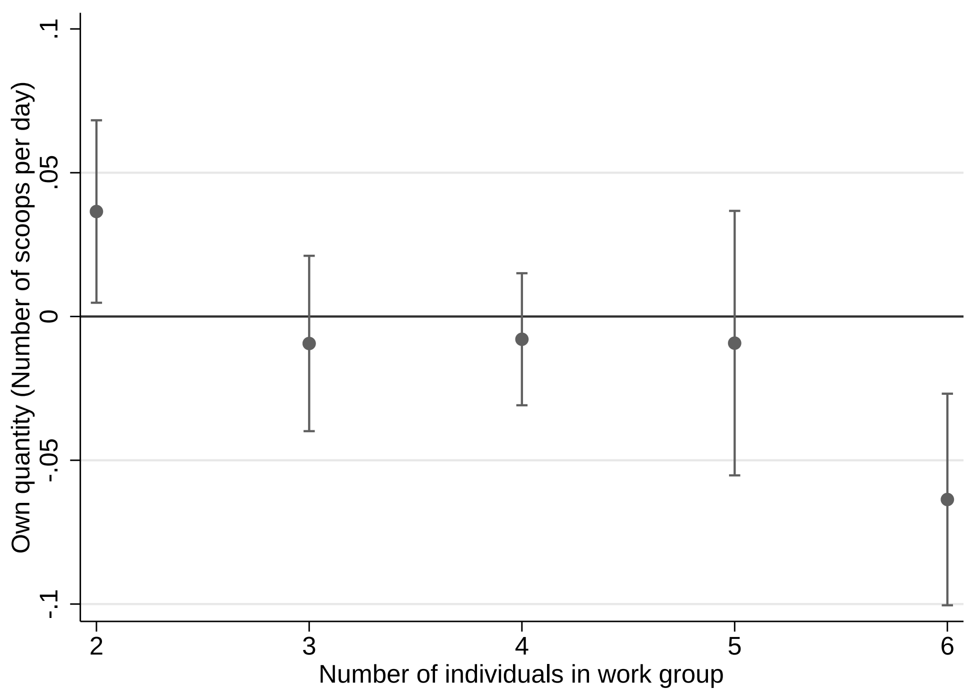


Figure 4: Peer effects: quality
Marginal effect of mean piece rate, by group size