NC STATE UNIVERSITY
Center for Environmental and Resource Economic Policy
College of Agriculture and Life Sciences
https://cenrep.ncsu.edu

Productivity in Piece-Rate Labor Markets: Evidence from Rural Malawi

Raymond P. Guiteras, B. Kelsey Jack

Center for Environmental and Resource Economic Policy
Working Paper Series: No. 17-019
September 2017

Suggested citation: Guiteras, Raymond P. and B. Kelsey Jack (2017). Productivity in Piece-Rate Labor Markets: Evidence from Rural Malawi (CEnREP Working Paper No. 17-019). Raleigh, NC: Center for Environmental and Resource Economic Policy.

Productivity in Piece-Rate Labor Markets: Evidence from Rural Malawi*

Raymond P. Guiteras
North Carolina State University†

B. Kelsey Jack
Tufts University‡

September 2017

Abstract

Piece-rate compensation is a common feature of developing country labor markets, but little is known about how piece-rate workers respond to incentives, or the tradeoffs that an employer faces when setting the terms of the contract. In a field experiment in rural Malawi, we hired casual day laborers at piece rates and collected detailed data on the quantity and quality of their output. Specifically, we use a simplified Becker-DeGroot-Marschak mechanism, which provides random variation in piece rates conditional on revealed reservation rates, to separately identify the effects of worker selection and incentives on output. We find a positive relationship between output quantity and the piece rate, and show that this is solely the result of the incentive effect, not selection. In addition, we randomized whether workers were subject to stringent quality monitoring. Monitoring led to higher quality output, at some cost to the quantity produced. However, workers do not demand higher compensation when monitored, and monitoring has no measurable effect on the quality of workers willing to work under a given piece rate. Together, the set of worker responses that we document lead the employer to prefer a contract that offers little surplus to the worker, consistent with an equilibrium in which workers have little bargaining power.

JEL Codes: C93, J22, J24, J33, O12

*We received helpful comments from the editor (Jeremy Magruder), two anonymous referees, James Berry, Drusilla Brown, Brian Dillon, Constanca Esteves-Sorenson, Andrew Foster, Jessica Goldberg, John Ham, Robert Hammond, John List, Hyuncheol Bryant Kim, Ellen McCullough, A. Mushfiq Mobarak, Stephanie Rennane and seminar participants at the University of Chicago, University College London / L.S.E., Duke University, Case Western Reserve, RAND and NEUDC. John Anderson and Stanley Mvula were invaluable in leading the field team. Portions of this paper were previously circulated under the title "Incentives, Selection and Productivity in Labor Markets: Evidence from Rural Malawi." We gratefully acknowledge financial support from ICRAF and the University of Maryland Department of Economics. The Supplementary Materials referenced in the text are posted at `http://go.ncsu.edu/rpguiter.malawi-labor.suppl`.

†4326 Nelson Hall, Campus Box 8109, Raleigh, NC 27695. Email: rpguiter@ncsu.edu.

‡Corresponding author. 314 Braker Hall, Medford, MA 02155. Email: Kelsey.Jack@tufts.edu.

1 Introduction

Piece rates are a common feature of developing-country labor markets (Jayaraman and Lanjouw 1999; Ortiz 2002). While there is a long-standing theoretical literature that acknowledges the importance of piece rates (e.g., Stiglitz (1975), Baker (1992) and Lazear (2000)), less is known about how such contracts operate in practice in developing countries (Baland et al. 1999; Newman and Jarvis 2000).[1] In fact, large-scale representative household and labor income surveys typically ignore contract structure, making even the most basic descriptive research difficult.[2] The small number of household surveys that separate earnings by contractual arrangement show a substantial share of rural labor is through piece rate contracts. For example, according to India's National Sample Survey (2011), 25.4 and 22.3 percent of person-days were paid by piece rate for rural men and women, respectively. Similarly, the Indonesia Family Life Survey (2007) shows that 12.9 percent of rural households had at least one member working for a piece rate. Piece rate contracts are also often observed in agricultural labor markets in developed countries (e.g., Paarsch and Shearer 1999; Bandiera et al. 2005; Chang and Gross 2014), and factory workers in both developed and developing countries are often paid with piece rates (e.g., Hamilton et al. 2003; Schoar 2014; Atkin et al. 2017).

Research on piece rate contracts in rural developing country labor markets is of policy interest given the substantial literature documenting frictions in developing country labor markets and persistent inequality between workers and employers (for overviews, see Rosenzweig 1988; Behrman 1999; Roumasset and Lee 2007). Employers face decisions over the level and type of compensation to offer, which may affect production both through a direct incentive effect and an effect on worker selection into the job. Piece rates provide direct performance incentives, but present tradeoffs if the employer values quality as well as quantity of output. Monitoring of output quality can be expensive and may increase the reservation

[1] A related literature examines the relationship between contractual form, worker and task characteristics and productivity. For example, using panel data, Foster and Rosenzweig (1994) show that piece rate contracts result in higher productivity and less moral hazard than time rate or share tenancy contracts. Eswaran and Kotwal (1985) develop theory to explain the variety of standard labor contracts in rural agricultural employment.

[2] Household surveys typically ask for earnings and hours worked and report the ratio of earnings to hours as a "wage," but do not investigate whether the work arrangement compensates on the basis of time worked or output. For example, Malawi's Labor Force Survey (2013) describes its methodology as follows: "wages received against actual hours worked to receive those wages was used to calculate mean wage per hour which was extrapolated to monthly gross." Malawi's Integrated Household Survey (2011) does distinguish between "Wage Employment" and "*Ganyu* labor," with 52 percent of rural households reporting at least some *ganyu*. In Malawi, the Chichewa word for casual labor, *ganyu*, translates as *piece work*, though other pay arrangements exist (Whiteside 2000), and the survey does not investigate further. Surveys that do make the distinction include India's National Sample Survey, the Indonesia Family Life Survey, and South Africa's National Income Dynamics Survey.

rate that workers require. The employer's preferred contract will depend on the relative importance of selection and incentive channels, and on how workers respond to incentives for both quality and quantity, which has implications for the division of surplus between the worker and the employer.

We investigate how workers respond to different contractual arrangements in the context of informal day labor markets in rural Malawi. We conduct a field experiment which mimics many features of a naturally occurring casual labor contract, including the nature of the compensation (piece rate) and the type of work (a menial agricultural task). In a simplified Becker-DeGroot-Marschak (BDM, Becker et al. 1964) exercise, 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.[3] The resulting dataset allows us to isolate a number of determinants of worker productivity that are typically confounded in observational datasets.[4]

First, we decompose the relationship between the piece rate and output into selection (the ability of the workers recruited) and incentives. 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. Incentives for quality do not, however, affect reservation rates. The selection and incentive effects of piece rates are of opposite signs. While higher piece rates encourage more effort, they also – surprisingly – attract workers that are slightly less productive, on average, though the negative relationship

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

[4] A number of recent field experiments in development economics have relied, as we do, on a two stage randomization to isolate the effect of self selection on outcomes. Hoffman (2009) uses the Becker-DeGroot-Marschak mechanism to study how the intra-household allocation of bednets varies both with willingness to pay and with price paid. Karlan and Zinman (2009) randomize interest rates before and after take up in a consumer credit experiment in South Africa to distinguish the effect of adverse selection from that of moral hazard on loan default rates. Ashraf, Berry, and Shapiro (2010) and Cohen and Dupas (2010) use similar two-stage pricing designs to isolate the screening effect of prices for health products. Kim et al. (2017) study selection and incentive effects of static incentives (performance bonuses) and dynamic incentives (opportunity for career advancement) among skilled workers (survey enumerators) in Malawi.

is economically small and not very robust.[5]

Second, the design is stratified by gender, which is an important determinant of labor market outcomes in many developing country contexts, including rural Malawi. We find substantial differences in behavior between men and women on both the extensive and intensive margins. In our setting, women accept lower piece rates, produce more and higher quality output, and earn more per day, both unconditionally and controlling for minimum WTA and the piece rate received. 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. The observed gender differences are consistent with an outside option for men that rewards different skills than those required by the bean sorting task and a greater likelihood that men are pursuing casual labor to meet immediate cash needs.

We combine our estimates to calibrate the optimal combination of piece rates and quality monitoring in our setting and find that low piece rates combined with quality monitoring maximizes employer profits. This is implied by four of our main empirical findings: (1) willingness to accept even low piece rates is high, (2) reservation rates do not increase when quality monitoring is explicit, (3) without monitoring, higher piece rates do not attract higher-ability workers, and (4) higher piece rates lead to only modest increases in output. Our more surprising findings – that workers do not demand higher compensation in exchange for the costs imposed by quality monitoring and that higher piece rates fail to attract more productive workers – help make contracts that pay little and penalize low quality output preferred by the employer. Consistent with qualitative evidence from our setting, these findings point to an equilibrium in which the bargaining power is squarely in the hands of the employer, who retains much of the surplus from the transaction.

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 performance pay in well-functioning labor markets.[6] While these studies

[5] We observe the relationship between reservation piece rates and productivity (worker selection) within the sample that shows up for the experiment. The relationship between reservation rates and productivity may look different in the population as a whole, however, the relevant sample in which to measure selection for this particular employment opportunity consists of those willing to work for the highest offered piece rate or less.

[6] 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 mo-

suggest that selection is an important determinant of worker ability, they tend to compare selection across types of compensation scheme rather than across different strengths of incentive within the same scheme.[7] Our experimental design varies both the level and type of incentive scheme – namely whether workers face explicit quality incentives – and separately measures the selection and incentive effects of the former and the combined effect of the latter. In contrast with much of the existing literature, we find no effect of worker selection on productivity – if anything, selection is negative among men. Features of the labor market, including gender differentiated tasks, appear to drive this result.

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).[8] Most related to our findings by gender is the work by Foster and Rosenzweig (1996), which shows that productivity differences by gender and task can explain specialization in rural labor markets, where piece rates mitigate some of the statistical discrimination present under time rate contracts. While previous studies have shown that men and women face different labor market opportunities (e.g., Beaman et al. 2015), our design allows us to characterize the margins on which these differences operate.

More broadly, we contribute to a large literature on rural developing country labor markets and offer 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.

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

---

tivated employees – and find that hiring more motivated workers is a substitute for monitoring the quality of output in a piece rate task.

[7] 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.

[8] On the other hand, in a study setting very similar to ours, Goldberg (2016) randomly varies daily wages in rural Malawi and finds similar supply elasticities for men and women during the low labor demand season. In a meta-analysis of 28 studies of performance pay – all but two of which use subjects from developed countries – Bandiera et al. (2016) find little evidence that men and women respond differently to performance pay relative to other compensation schemes. Our evidence on the incentive effective of performance pay is broadly consistent with the papers they review; our selection results are not. We conjecture that we find bigger gender differences on the selection margin because of the differences in labor market opportunities for men and women in our setting.

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 effort, participation and the effects of monitoring. We use the framework to discuss potential gender differences.

2.1 Setup

A firm values output quantity, Y, and loses revenue when output quality, Q, falls below a threshold $\bar{Q}$. It offers a piece rate r to workers for production of Y 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 of 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

$$\begin{aligned} \text{Income}\,(Y, Q; r, M) &= \begin{cases} rYQ & \text{if } M = 1 \\ rY & \text{if } M = 0 \end{cases} \\ &= rYQM + rY\,(1 - M) \end{aligned}$$

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

The worker chooses to allocate effort toward production of Y and Q, which together determine the cost of effort $c(Y, Q)$, 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:[9]

$$c\,(Y, Q; \gamma) = c\,(Y, Q)\,/\gamma.$$

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

The worker's utility is her income minus her effort cost:

$$\begin{aligned} U(Y,Q;r,\gamma,M) &= \begin{cases} rYQ - c(Y,Q)/\gamma & \text{if } M=1 \\ rY - c(Y,Q)/\gamma & \text{if } M=0 \end{cases} \\ &= rYQM + rY(1-M) - c(Y,Q)/\gamma \end{aligned}$$

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

2.2 Worker's optimal response

If the firm does not monitor ($M=0$), the worker's optimal response (conditional on her participation constraint, given by equation 6, below) is to set quality $Y_{NM}^{*}=0$ and quantity Y_{NM}^{*} determined by the first-order condition

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

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

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

or an interior solution with (Y_{M}^{*},Q_{M}^{*}) solving the system of first-order conditions

$$\mathrm{FOC}_{Y_M} : rQ_{M}^{*} = \frac{1}{\gamma}\frac{\partial c}{\partial Y}\bigg|_{\left(Y_{M}^{*},Q_{M}^{*}\right)} \tag{3}$$

$$\mathrm{FOC}_{Q_M} : rY_{M}^{*} = \frac{1}{\gamma}\frac{\partial c}{\partial Q}\bigg|_{\left(Y_{M}^{*},Q_{M}^{*}\right)}. \tag{4}$$

Intuitively, in (3) the worker sets the marginal revenue from a unit of output[10] 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. $\partial Y^{*}/\partial\gamma > 0$ 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

[10]Given the optimal quality level Q_{M}^{*}, the quality-adjusted piece rate is rQ_{M}^{*}.

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 the sign of the cross-partial $\partial^2 c(Y,Q)/\partial Y \partial Q$. For the task we study, this cross-partial is likely to be positive (i.e., at a given level of effort, quantity and quality are likely to be substitutes), in which case an increase in the piece rate increases output quantity at a cost of a reduction in output quality.

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, which we denote $V(\gamma, r; M)$,[11] and their outside option, which we denote $\underline{V}(\gamma)$. 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[12]

$$\text{PC-M}: \; V(\gamma, r; M=1) = rY^*_M Q^*_M - c(Y^*_M, Q^*_M)/\gamma \geq \underline{V}(\gamma) \tag{5}$$

$$\text{PC-NM}: \; V(\gamma, r; M=0) = rY^*_{NM} - c(Y^*_{NM}, 0)/\gamma \geq \underline{V}(\gamma) \tag{6}$$

which lead to reservation rates

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

We are interested in comparative statics with respect to monitoring (the relationship

[11] $V(\gamma, r; M)$ is a value function, i.e., the net utility (income minus effort cost) to a worker of productivity γ at her optimal response (Y^*, Q^*) to a contract offer of (r, M).

[12] If the participation constraints do not hold, the worker supplies $Y = 0, Q = 0$.

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 should increase to compensate her. The second, whether the reservation piece rate is positively or negatively related to productivity (i.e. the sign of $\partial \underline{r}/\partial\gamma$), is ambiguous. More productive workers will require a higher piece rate, i.e., $\partial\underline{r}/\partial\gamma > 0$, if an increase in γ makes the participation constraint more difficult to satisfy. Consider the case $M = 1$.[13] The left-hand-side of (5) has derivative[14]

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

The right-hand side of (5) has derivative $\underline{V}'(\gamma)$. If $\underline{V}'(\gamma) < 0$, i.e. if workers with 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 $c(Y_M^*, Q_M^*)/\gamma^2$ 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 Gender

In the context of our model, worker gender is primarily relevant through the joint distribution of productivity, γ, and the outside option, $\underline{V}$ and through the cost of effort. Men and women can have different distributions of productivity, of the outside option, or the relationship between these two, i.e. the function $\underline{V}(\gamma)$. Effort costs are also important: if the correlation between productivity and effort costs differs between men and women, it may lead to gender differences in both reservation rates and effort.

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*.[15] In Malawi, like in many rural agricultural settings in developing countries, labor markets are highly

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

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

[15] See Whiteside (2000), Dimowa et al. (2010) and Sitienei et al. (2016) for enlightening discussions of *ganyu* in Malawi.

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 dried beans into eight categories.[16] Sorted beans receive a price premium of roughly 50 percent. This task is well-suited to our 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[17] 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 into which the mixed beans must be sorted.[18] 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 Y_i as the number of units (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., $PR_i \geq \underline{PR}_i$,

[16] 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.

[17] 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, and Section S1 of the Supplementary Materials for a participant flow diagram.

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

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.[19] Second, it creates random variation in the actual piece rate 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.[20]

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.[21] The participant indicates which of the rates she will accept, the lowest of which we record as her reservation piece rate.[22] 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.[23]

Table 1 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

[19] 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 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 earnings are not large enough to plausibly affect willingness to work in a neoclassical model.

[20] Berry, Fischer, and Guiteras (2015) 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."

[21] 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. We calibrate employer profits in Section 5.

[22] 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, especially since our goal is to compare participants with different WTA rather than to measure WTA with great precision.

[23] 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 Section S2 of 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.

observed for participants who draw a piece rate at least as high as their reservation piece rate.

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 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). This limits 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 sorting the mixed beans into eight categories. 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. Consequently, we cannot completely separate the effect of the monitoring treatment on worker selection (WTA) from the effect on incentives.

The monitoring procedure (both as implemented and as described to the subjects) was that two categories of beans (out of the eight sorted categories) 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.[24] 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. Rather than returning to the same villages across seasons, we chose to work in different villages in different seasons. This makes the worker pools not directly comparable across seasons, but even if we had returned to the same villages, the pools would not have been the same since – in addition to likely differences in selection into the study – learning, differences in trust, etc., would have persisted across seasons. 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 of their minimum WTA was their best strategy. The distribution of possible piece rates

[24]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.

was made clear during the description of the task and the performance of the skit. Subjects were also informed that they would receive a participation fee of 50 MWK for each day they participated, plus their earnings from the day's work.[25] The participation fee was intended to offset the time costs of participants who showed up but were not awarded a contract. The information provided during the orientation session presumably resulted in some selection out of the study by workers with a reservation piece rate above the highest piece rate offered. We argue that this selected sample is the relevant one for understanding the determinants of productivity for the labor market activity that we study.

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 used 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.

Conditional on attending 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.[26] 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 and the quality for every unit sorted. At the end of the work day, partially sorted units were paid according to the fraction of the unit sorted.[27]

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.[28] The participation

[25] We note a tradeoff associated with offering a participation fee that is high relative to the average earnings in the experiment. On the one hand, the participation fee may help offset selection into the experiment. On the other hand, if individuals have an income goal for the day (as in Farber (2005); Dupas and Robinson (2013)) or utility that is very concave in income, then the participation fee may dampen worker response to the piece rate. The target earnings model will dampen the response to the piece rate regardless of the participation fee, and reasonable utility functions are unlikely to generate curvature sufficient for the participation fee to make a substantial difference.

[26] 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.

[27] Fifteen minutes before the end of the work day, enumerators stopped handing out units for sorting, so most workers completed their final unit. Any proportional payments were estimated by the enumerators.

[28] Survey data were collected in two parts. The first, more comprehensive part, covering basic demographics

fee was contingent on the participant completing the survey.

We gathered additional qualitative data to provide context on the *ganyu* market in Malawi. Specifically, we interviewed 53 workers and 8 employers at trading posts and agri-processing centers in and around Lilongwe.[29] We targeted locations known to employ *ganyu* workers and asked questions pertaining both to the work conducted at the location and to past *ganyu* contracts. The most commonly observed *ganyu* activities among workers on the day of the interview was sorting groundnuts, maize, soya and other beans, and offloading bags from trucks arriving at the trading posts and processing centers. About half of these were paid a time rate (55 percent) and half a piece rate (45 percent). Most (74 percent) said that their employer monitored the quality of the work, though the likelihood of monitoring was considerably lower for workers on piece rate than on time rate contracts (62.5 versus 82.8 percent). Most often, workers report that monitoring is done by supervisors who observe work in process, though some monitoring appears more systematic, with random spot checks or an inspection stage to the process. Penalties for unsatisfactory performance are severe: workers report having to do the entire task over again or losing out on the day's pay.

3.3 Descriptive statistics

3.3.1 Characteristics and participation

Key characteristics of participants are described in Table 2, 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.[30]

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. Because the study samples from different villages in the low and high labor demand seasons,

---

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 and participants' expected output that day. In both cases, the survey was conducted regardless 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.

[29] This qualitative data collection was completed in June 2017, as part of a revision to this paper.

[30] Covariate balance for the piece rate draws and the monitoring treatment is shown in Table S1. Because we see some imbalance, we include controls in selected specifications in all of our main analyses.

we report means and standard deviations separately for each season. Some differences are statistically significant. 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.[31] 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.[32]

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 in each season, explaining differences in average participant age or other income sources. 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.[33]

Table 3 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 (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, due both to fewer repeat workers during this period and to use of the lottery to limit the number of participants in all weeks. 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.

3.3.2 Willingness to accept

Panel B of Table 3 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

[31] 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.

[32] Summary statistics on a broader set of survey measures are reported in Table S2 of the Supplementary Materials.

[33] 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.

2 MWK lower than for men.[34] 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.[35] 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, plus the 50 MWK show up fee.[36] This is consistent with the high rate of labor force participation even at very low daily wages observed by Goldberg (2016). Conditional on working, mean daily earnings were over 170 MWK, which exceeds average daily wages reported in the 2004 IHS and induced over 90 percent of the adult population in Goldberg's study to agree to work.[37]

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

[34] For correlations between WTA and other characteristics, see Table S3, which reports the pairwise correlation between outcomes (WTA, quantity and quality) and survey measures.

[35] The acceptance rates plotted in Figure 1 are provided in Table S4.

[36] These calculations consider only the acceptance of piece rates revealed by the BDM. Attrition or selective attendance might cause us to over-estimate acceptance rates if high minimum WTA individuals attended fewer work days. We note that, conditional on attending any work days, minimum WTA is correlated with the number of days attended. Specifically, a 5 MWK increase in minimum WTA is associated with 0.25 to 0.4 fewer days in attendance. However, acceptance rates of low piece rates remains high even if we take the extreme stance and interpret decisions not to attend as rejections all offered piece rates. In this scenario, 43 percent accept an offer of 10 MWK. Alternatively, if we assume that those who chose not to attend all days have a stable minimum WTA equal to their highest observed minimum WTA in the BDM, then 53 percent accept an offer of 10 MWK. Importantly, the attendance decision is unrelated to the previous day's randomly drawn piece rate or to the previous day's monitoring treatment. We further note that the decision of how to treat missing minimum WTA observations will only change our interpretation of selection effects on productivity if the workers with high minimum WTA who choose not to attend all days (or at all) also have high productivity in the bean sorting task.

[37] Workers sorted an average of 7.35 units per day (Table 3). Workers reported that they expected to sort an average of 6.74 units per day (Table S2).

3.3.3 Quantity and quality of worker output

The primary measures of productivity, number of units sorted per day (Y) and average number of errors per unit (Q), are summarized in Panel C of Table 3.[38] 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 (see Table S3), consistent with our model's single productivity parameter for quality and quantity.

4 Empirical Results

We present our empirical strategy and results together. 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 which covariates predict minimum WTA. Second, we estimate the determinants of both quantity and quality of output, separating selection – minimum WTA, conditional on the piece rate – from incentive effects, which are associated both with the randomly determined piece rate and with the quality monitoring treatment. Finally, we estimate differences in minimum WTA and productivity determinants between men and women.

4.1 Determinants of minimum WTA

In this subsection, we examine which covariates predict minimum WTA. To study the predictors of reservation piece rates, we regress minimum WTA on characteristics of the market, specifically, indicators for the labor season (High = 1 for high season), the monitoring treatment, M_{id}, whether the participant is female, F_i, and day of the week, DoW_d:

$$\text{MinWTA}_{id} = \phi\text{High}_i + \lambda M_{id} + \zeta F_i + \tau\text{DoW}_d + \delta\text{Dist}_i + c_i + \varepsilon_{id}, \tag{7}$$

where c_i is a participant random effect, fixed effect or omitted, depending on the specification.

Table 4 shows estimation results in cross section (columns 1-2), using random effects

[38] 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.

(columns 3-4) and individual fixed effects (column 5).[39,40] Columns (1)-(4) include district fixed effects (Dist_i). Minimum WTA is slightly lower in the high season. 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. It also may be due to the relatively greater cash constraints that households typically face during the high season. Contrary to the model's prediction, 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. As shown in Table S6, this insignificant and negative effect of monitoring on minimum WTA continues to hold even when we restrict the sample to individuals who worked under both types of contracts, i.e., the 299 observations from 204 subjects who participated in BDM on days 3 or 4 and who had worked under both no monitoring and monitoring contracts on earlier days. This may reflect a preference for more complete contracts, or inattention to the contractual details at the time of bidding.

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. Some of the variation in minimum WTA within-individual is correlated with the self-reported outside option, as shown in Table S7. Specifically, individuals who report that their alternative activity for the day was associated with cash income (other casual labor, working own business) have lower reservation rates, as do those who would have worked on their own farms. Importantly for identification in the next sections, minimum WTA is not affected by treatments on the previous day (for subjects who attend on multiple days), i.e., the piece rate draw and the monitoring treatment do not affect BDM behavior on the subsequent day.

[39]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, with similar results. We also include results from pooled OLS in our main specifications because of the less stringent assumptions for identification with repeated observations over time.

[40]In spite of the interval nature of our outcome data, mean regression appears not to suppress much relevant information; repeating the analysis with interval regression leads to very similar results (see Table S5).

4.2 Determinants of productivity

Next, we investigate the role of both selection into the contract and the incentives provided by the contract on worker productivity. The thought experiment, as described in Section 3.1.1, is to first compare workers with different minimum WTA who receive the same random piece rates, and second to compare workers with the same minimum WTA who receive different random piece rates. With a very large sample, we could do this completely nonparametrically, e.g. by comparing outcomes for each cell in Table 1. However, in our finite sample, some cell sizes are very small, requiring us to make some functional form assumptions. The basic intuition is unchanged: by controlling for the piece rate, we isolate the selection channel; by controlling for minimum WTA, we isolate the direct effect of incentives on productivity. These latter 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.[41]

We begin with a descriptive summary of the findings. As a middle ground between flexibility and precision, we estimate a parametric model with quadratics in minimum WTA and the piece rate, their interaction, all interacted with the monitoring treatment. Specifically, we estimate

$$\begin{aligned} y_{id} \;=\;& \beta_1 \text{minWTA}_{id} + \beta_2 \text{minWTA}_{id}^2 + \beta_3 \text{PR}_{id} + \beta_4 \text{PR}_{id}^2 + \beta_5 \left(\text{minWTA}_{id} \times \text{PR}_{id}\right) \\ +\;& \beta_6 \left(M_{id} \times \text{minWTA}_{id}\right) + \beta_7 \left(M_{id} \times \text{minWTA}_{id}^2\right) \\ +\;& \beta_8 \left(M_{id} \times \text{PR}_{id}\right) + \beta_9 \left(M_{id} \times \text{PR}_{id}^2\right) + \beta_{10} \left(M_{id} \times \text{minWTA}_{id} \times \text{PR}_{id}\right) \\ +\;& \lambda M_{id} + \phi \text{High}_i + \zeta F_i + \tau \text{DoW}_d + \delta \text{Dist}_i + c_i + \varepsilon_{id}, \end{aligned} \tag{8}$$

with worker random effects and calculate estimated differences relative to the omitted category of a worker in the no monitoring treatment, with a minimum WTA of 5 MWK per unit sorted, and facing a piece rate of 5 MWK.[42]

Figure 2 plots these estimated differences. The horizontal axis is the randomly varied piece rate, and each line corresponds to a category of minimum WTA. The incentive effect, i.e., differences in productivity for workers with the same minimum WTA who earn different piece rates, is seen by comparing horizontally along a given line. Worker selection, i.e.,

[41] 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 has a significant effect on stated minimum WTA.

[42] In the Supplementary Materials, we provide several robustness checks: the results presented here exclude the 13 worker-day observations with minimum WTA of 25, little changes when these are included (Figure S1); results are similar with a linear specification (Figure S2); allowing a completely nonparametric estimation (estimating each minimum WTA-by-piece rate-by-monitoring cell separately) results in noisy estimates that are generally consistent with what we see from this flexible parametric model (Figure S3).

differences in productivity for workers with different minimum WTA but facing the same piece rate, is shown by comparing vertically across lines at a given piece rate. Figure 2a, which shows effects on output quantity in the absence of monitoring, suggests a slight negative selection effect on the number of units per day at piece rates above 15 MWK, and incentive effects that increase output most effectively in the lowest WTA group. Figure 2b suggests that monitoring eliminates most of the negative selection effect and regularizes the effect of incentives on output. In other words, we observe upward sloping lines in all minimum WTA categories in Figure 2b but not Figure 2a. Furthermore, monitoring seems to have reduced differences in output across minimum WTA categories overall, although some negative selection persists among those with the highest reservation piece rates. A similar story emerges on the quality dimension, as seen by comparing Figures 2c and 2d. Figure 2c shows that, in the absence of monitoring, there exists heterogeneity across worker type both in quality of output at a given piece rate and in the response of quality to the piece rate. In comparison, Figure 2d shows that the first-order effect of monitoring is a reduction in the error rate across all worker types and piece rates, and additionally that monitoring reduces heterogeneity across workers – slopes are similar across all worker types and, with the exception of the highest reservation piece rate types, there is not much difference in the level of quality across types, holding the piece rate fixed.

4.2.1 Worker selection

To quantify and summarize the descriptive results of the previous subsection, we estimate regression models that are linear in the main variables of interest, minimum WTA and the piece rate. 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:

$$\begin{aligned} y_{id} &= \beta \text{minWTA}_{id} + \alpha \text{PR}_{id} + \lambda \text{M}_{id} \\ &+ \phi \text{High}_i + \zeta F_i + \tau \text{DoW}_d + \delta \text{Dist}_i + c_i + \varepsilon_{id}. \end{aligned} \tag{9}$$

We interpret the coefficient β as selection within our sample of participants: the relationship between the reservation piece rate and output, holding the actual piece rate constant. Tables 5 and 6 show the relationship between WTA and the number of units sorted per day and the number of errors per unit, respectively.[43] Column (1) is estimated by pooled OLS,

[43] 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

and column (3) by worker random effects.[44] With respect to quantity, we observe slightly *negative* selection in some specifications: 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 in the random effects model (columns 4-6, Table 5), 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) and in the pooled OLS specification (columns 4-6, Table 5) is close to zero and statistically insignificant. The same specification with number of errors per unit as the dependent variable shows no significant relationship between minimum WTA and quality of output, though the coefficient on minimum WTA is consistently negative (Table 6).

4.2.2 Incentives

BDM randomly assigns the piece rate paid among those who work, so it is straightforward to test the causal effect of incentives on productivity using Equation (9). 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.

Table 5 shows the effect of the piece rate on quantity of output, controlling for the worker's reservation piece rate. Our basic specification is estimated by pooled OLS in column (1) and with worker random effects in column (3). Increasing the piece rate by 10 MWK increases the number of units sorted per day by between 0.18 and 0.26 units, relative to a mean of 7.35 units. Going from the lowest piece rate (5 MWK) to the highest piece rate (25 MWK) increases output by between 0.4 to 0.6 units per day. 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 (Table S8). In the high season, workers sort almost half a unit more per day.

Table 6 shows the effect of the piece rate on quality of output, measured by the number of errors per unit sorted. 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 unit is decreasing in the day of the week, consistent with individuals gaining experience with the task.[45] This effect is robust to the inclusion

operating with quality held fixed, nor vice versa.

[44] Fixed effects estimates are provided in Tables S8 and S9. We control for covariates as listed in Table 4; specifications without controls are provided in Tables S10 and S11. Finally, we provide estimates using data only from the first day, to limit potential bias from attrition correlated with treatment (Tables S12 and S13). The results are qualitatively similar, although with less precision – for example, the effect of the price draw is no longer statistically significant.

[45] Given that workers in our sample come from agricultural households and that households do their own post-harvest processing, including sorting maize, groundnuts and beans, we conjecture that most workers would be familiar with similar sorting tasks, though unfamiliar with the particulars of our work arrangements.

of individual fixed effects (Table S9), 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.

4.2.3 Monitoring

The level effect of stricter monitoring on quantity and quality of output is measured by λ in estimating equation (9) 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 5.[46] The coefficient on monitoring is negative and significant, lowering output by between -0.440 and -0.689 units 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 unit, as shown in Table 6. The coefficient on monitoring is between -0.561 and -0.744 units per day, or about three-quarters of a standard deviation. Monitoring does appear to divert effort toward output quality at a cost of some quantity.

To consider how monitoring affects selection, we add an interaction between minWTA_{id} and M_{id} to Equation (9). The results are reported in columns (2) and (4) of Tables 5 and 6. First, among workers who reveal a low minimum WTA, output quantity is lower in the monitoring condition than in the no-monitoring condition. However, there is no difference in output quantity by monitoring among workers who reveal a high minimum WTA.[47] Second, without quality monitoring, higher minimum WTA workers produce slightly more errors; with quality monitoring, higher minimum WTA workers produce slightly fewer errors. Both of these are imprecisely estimated, i.e., the difference between high and low minimum WTA workers is statistically insignificant in both the monitoring and no monitoring conditions. To consider how monitoring affects incentives, we add an interaction between PR_{id} and M_{id} to Equation (9). The results are reported in columns (3) and (5) of Tables 5 and 6. There does not appear to be a strong interaction with the piece rate, neither in the quantity nor quality dimension.

To summarize, these results suggest that (a) the quantity-quality tradeoff associated with monitoring is greater for low minimum WTA workers, (b) monitoring corrects for the

Thus, we might expect more learning about the work flow and the details of our sorting requirements than about bean sorting in general.

[46] 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.

[47] Recall that minimum WTA was elicited after the worker is informed about the monitoring treatment; the effect of monitoring via minimum WTA should therefore be interpreted as suggestive.

slight negative selection effect on output quality and (c) monitoring shifts the allocation of effort between quantity and quality but does not affect the causal effect of the piece rate on productivity. Introducing quality monitoring leads to level shifts in output quality (errors fall) and quantity (output decreases), which suggest that quality and quantity are substitutes in production in our setting. At the same time, we see no difference in the slope on the effect of higher piece rates on either dimension of output between the monitoring and no monitoring conditions. Notably, higher piece rates have little effect on output quality whether quality is directly incentivized or not. This is consistent with workers targeting $\underline{Q}$ in the absence of monitoring and $\bar{Q}$ under monitoring, at all piece rates. However, the tradeoff between quality and quantity suggests that workers may reciprocate higher piece rates in the absence of monitoring by sharing some surplus on the quality dimension.[48]

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 (7) with interactions of the female variable with the labor season (High) and monitoring treatment (M). Table 7 shows the results for reservation rates by gender.[49] Overall, women are willing to work for less: their WTA is 2.3 to 3.0 Kwacha lower than that of men. Quality monitoring has an imprecisely negative effect on minimum WTA for both men and women, with an insignificantly larger effect for men. 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 strong selection for women on covariates negatively associated with minimum WTA.

In discussing gender differences in productivity, as in Section 4.2, we begin with descriptive results from a flexible parametric specification, specifically by estimating a variant of Equation (8), but here interacting the terms of interest with an indicator for whether

[48] In an Akerlof-type model (Akerlof 1982) of gift exchange, workers take their stated reservation rates as the reference wage and reward higher piece rates with higher effort. Given the piece rate nature of the work contract, the relevant dimension of effort on which to share surplus is output quality. As the incentive to produce more output increases, in the absence of quality incentives, workers may test $\underline{Q}$ if it is not made explicit (it is not in our setting), such that quality should deteriorate at higher piece rates in the absence of reciprocity. Some prior work suggests monitoring crowds out worker reciprocity (Dickinson and Villeval 2008). Our findings are consistent with such a model, or with other factors that may lead workers to deliver quality even when their incentive to do so is not explicit, such as career concerns or internalized norms around shirking.

[49] Table S14 replicates the analysis of Table 7 but using interval regression as the estimator, with similar results.

the worker is female, instead of an indicator for the monitoring treatment. Specifically, we estimate

$$
\begin{aligned}
y_{id} &= \beta_1 \text{minWTA}_{id} + \beta_2 \text{minWTA}_{id}^2 + \beta_3 \text{PR}_{id} + \beta_4 \text{PR}_{id}^2 + \beta_5 \left(\text{minWTA}_{id} \times \text{PR}_{id}\right) \quad (10)\\
&+ \beta_6 \left(F_{id} \times \text{minWTA}_{id}\right) + \beta_7 \left(F_{id} \times \text{minWTA}_{id}^2\right)\\
&+ \beta_8 \left(F_{id} \times \text{PR}_{id}\right) + \beta_9 \left(F_{id} \times \text{PR}_{id}^2\right) + \beta_{10} \left(F_{id} \times \text{minWTA}_{id} \times \text{PR}_{id}\right)\\
&+ \lambda M_{id} + \phi \text{High}_i + \zeta F_i + \tau \text{DoW}_d + \delta \text{Dist}_i + c_i + \varepsilon_{id},
\end{aligned}
$$

with worker random effects and calculate estimated differences relative to the omitted category of a worker in the no monitoring treatment, with a minimum WTA of 5 MWK per unit sorted, and facing a piece rate of 5 MWK. These estimates, plotted in Figure 3, shows that male workers are more heterogeneous, and that the negative selection effect noted above is largely concentrated among men. This interpretation becomes even more clear when we further interact with monitoring, as in Figures 4 and 5 – women are somewhat less heterogeneous than men without monitoring (Figure 4), and monitoring largely eliminates heterogeneity among women, but not men (Figure 5).

Differences in productivity, controlling for reservation piece rates, are obtained by re-estimating Equation (9) allowing the effects of monitoring, minimum WTA, the piece rate and the labor season to vary by gender. Tables 8 and 9 show the effects on the quantity of output and quality of output respectively.

Recall that, on average, females sort more units of beans with fewer errors than do men. Monitoring reduces the quantity of output for both genders, but more so for females (Table 8). 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 fewer units of beans (a 10 MWK increase in the reservation piece rate is associated with sorting 1/5 to 1/2 fewer units per day, significant only in the random effects specifications); among women, the relationship is inconsistently signed and insignificant. Thus, the negative selection described in Section 4.2 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), or if men who produce higher output also have a lower minimum willingness to accept, consistent with cash needs driving both margins.

Survey evidence helps address alternative interpretations of the negative selection observed among men. First, expected output is positively correlated with actual output for both men and women, indicating that men are not systematically more likely to mis-estimate their own productivity. In fact, men on average make more accurate predictions of their output. Second, reporting a cash-generating outside option is negatively correlated with reservation piece rates, even within participant. Men are significantly more likely to report an outside option that generates cash earnings, such as other casual labor, while women are significantly more likely to report housework as their outside option. Third, high productivity men who state low reservation rates are not able to make up for lower piece rates with higher output: for a given piece rate, daily earnings do not increase with minimum WTA. Factors not measured in our survey may also matter. If, for example, women place extra value on the earnings in the experiment because they can be kept private, then they may both be willing to work for less and may work harder conditional on receiving a contract, consistent with the gender differences we observe.

In spite of the differences in minimum WTA across season among women, output by women is the same in the high and low season. Men, on the other hand, produce more output in the high season than in the low season, approximately one unit per day. This difference may also be related to the differences in men and women's outside options and need for cash income, and how they vary with the labor season, or to differences in the subject pool in the low and high labor demand seasons.

Both men and women are similarly responsive to monitoring in their allocation of effort toward output quality (Table 9, with additional results in Tables S15 and S16). 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.

5 Discussion

We implement a unique experimental design in casual labor markets in rural Malawi to understand the workings of a common form of piece-rate labor, to observe gender differences in these markets and to test for other behavioral determinants of productivity. Raising the piece rate has little effect on the quality of worker interested in the job; if anything, selection effects are associated with lower productivity. On the other hand, a higher piece rate significantly increases the quantity of output, controlling for workers' reservation rates, but does not reduce quality. Explicit incentives for output quality reduce the error rate in production at some cost to output quantity, but do not affect worker willingness to accept the task. Furthermore, the more complete contracts that include quality incentives appear to regularize output and reduce some of the differences in response to the piece rate that arise through worker self-selection when quality is unmonitored.

Comparing the extensive (participation) versus intensive (effort) margins, conditional on showing up to the experiment, participation is more responsive to the piece rate than is effort. Over the range of piece rates offered, the arc elasticity of participation with respect to the piece rate is 0.58 (Table S17), while the elasticity of output quantity with respect to the piece rate (controlling for the reservation piece rate) is 0.06 (Table S18). The reasonably high participation elasticity suggests that the variation in piece rates in our study has a substantial effect on hiring outcomes, conditional on showing up for the experiment. See Tables S17 and S18 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.

We perform a simple calibration of employer profits at different piece rates, with and without quality monitoring, and by gender. (See Section S3 of the Supplementary Materials for details.) Assuming a penalty for mis-sorted beans, and costs per worker-day associated with hiring and administering the work and with quality monitoring, the employer prefers to monitor output quality at all piece rates. Under our main assumptions about the quality premium and supervision costs, the highest overall profit per bag comes from the lowest piece rate combined with worker penalties for low quality output, i.e., the contract that is least profitable to the worker. However, the profit-maximizing contract depends on the gender of the worker. The profit implications of quality monitoring are modest when the worker is female, and the profit-maximizing contract for a female worker is a low piece rate without monitoring. For male workers, on the other hand, profits vary considerably more with the contractual arrangements: at low piece rates, profits are nearly 50 percent higher if quality is monitored, yet higher piece rates narrow the profitability gap between monitored and unmonitored male workers. Differences in the profit-maximizing contract by gender arise exclusively from the worker responses that we document, and give employers an incentive to discriminate based on this easily observed trait. Overall, the high acceptance rates for the contractual arrangements that are least favorable to the worker, together with selection effects from higher piece rates that are, if anything, negative in the sample that shows up to our experiment, make high piece rates even less attractive from the perspective of the employer. In our study setting, rural households have few well-paying outside options for cash earnings, which places the bargaining power squarely in the hands of the employer.

The context in which the study takes place appears to shape many of the findings. For example, qualitative survey evidence (see Section 3.2) suggests that quality monitoring is common in piece rate work arrangements, and accompanied by high penalties for violation of only loosely defined quality standards. Participants in our study may have therefore preferred the more complete terms offered by the monitoring contract over the potentially ambiguous contract without explicit quality monitoring. As another example, 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). This may explain some of the observed negative selection by men in our study. 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. We also observe that minimum WTA is lower among participants who report an outside option associated with other income-generating activities, including those that would help meet immediate cash needs, such as other casual labor (see Table S7).[50]

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. In addition, additional data collection that offers a more nuanced picture of the contractual forms common for different tasks and in different markets would facilitate further study of the contracting frictions in rural labor markets. Finally, our results suggest that further data collection on the prevalence of piece rate contracts in developing country rural wages would stimulate additional research into the structure and distributional effects of rural labor markets, between men and women, and between employers and workers.

[50]This is also consistent with Goldberg (2016), who reports that 70 percent of participants in her study report immediate cash needs as their reason for working. Fink et al. (2014) provide further evidence from a similar setting in neighboring Zambia that cash needs drive casual labor supply, particularly during the high season.

References

George A. Akerlof. Labor contracts as partial gift exchange. *The Quarterly Journal of Economics*, 97(4):543–569, 1982.

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.

David Atkin, Azam Chaudhry, Shamyla Chaudry, Amit K Khandelwal, and Eric Verhoogen. Organizational barriers to technology adoption: Evidence from soccer-ball producers in pakistan. *The Quarterly Journal of Economics*, 132(3):1101–1164, 2017.

George P. Baker. Incentive contracts and performance measurement. *Journal of Political Economy*, 100(3):598–614, 1992.

Jean-Marie Baland, Jean Drèze, and Luc Leruth. Daily wages and piece rates in agrarian economies. *Journal of Development Economics*, 59(2):445 – 461, 1999.

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.

Oriana Bandiera, Greg Fischer, Andrea Prat, and Erina Ytsma. Do women respond less to performance pay? building evidence from multiple experiments. 2016.

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

Lori Beaman, Niall Keleher, and Jeremy Magruder. Do job networks disadvantage women? Evidence from a recruitment experiment in Malawi. Mimeo, August 2015.

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

Jere R. Behrman. Labor markets in developing countries. In *Handbook of Labor Economics*, volume 3, Part B, pages 2859 – 2939. Elsevier, 1999.

James Berry, Greg Fischer, and Raymond Guiteras. Eliciting and utilizing willingness to pay: Evidence from field trials in Northern Ghana. MPRC Working Paper PWP-MPRC-2015-017, December 2015.

Tom Chang and Tal Gross. How many pears would a pear packer pack if a pear packer could pack pears at quasi-exogenously varying piece rates? *Journal of Economic Behavior & Organization*, 99:1–17, 2014.

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.

David Dickinson and Marie-Claire Villeval. Does monitoring decrease work effort?: The complementarity between agency and crowding-out theories. *Games and Economic Behavior*, 63(1):56 – 76, 2008.

Ralitza Dimowa, Katharina Michaelowa, and Anke Weber. Ganyu labour in Malawi: Understanding rural households' labour supply strategies. Number 29 in Proceedings of the German Development Economics Conference, 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.

Mukesh Eswaran and Ashok Kotwal. A theory of contractual structure in agriculture. *The American Economic Review*, 75(3):352–367, 1985.

Henry S. Farber. Is tomorrow another day? The labor supply of New York City cabdrivers. *Journal of Political Economy*, 113(1):46–82, 2005.

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.

Günther Fink, B. Kelsey Jack, and Felix Masiye. Seasonal credit constraints and agricultural labor supply: Evidence from Zambia. Mimeo, 2014.

Andrew D. Foster and Mark R. Rosenzweig. A test for moral hazard in the labor market: Contractual arrangements, effort, and health. *Review of Economics and Statistics*, 72(2): 213–227, 1994.

Andrew D Foster and Mark R Rosenzweig. Comparative advantage, information and the allocation of workers to tasks: Evidence from an agricultural labour market. *The Review of Economic Studies*, 63(3):347–374, 1996.

Jessica Goldberg. Kwacha gonna do? Experimental evidence about labor supply in rural Malawi. *American Economic Journal: Applied Economics*, 8(1):129–149, January 2016.

Barton H. Hamilton, Jack A. Nickerson, and Hideo Owan. Team incentives and worker heterogeneity: An empirical analysis of the impact of teams on productivity and participation. *Journal of Political Economy*, 111(3):465–497, 2003.

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

Vivian Hoffman. Intrahousehold allocation of free and purchased mosquito nets. *American Economic Review*, 99(2):236–241, 2009.

Rajshri Jayaraman and Peter Lanjouw. The evolution of poverty and inequality in Indian villages. *The World Bank Research Observer*, 14(1):1–30, 1999.

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

Hyuncheol Bryant Kim, Seonghoon Kim, and Thomas T. Kim. The selection and causal effects of work incentives on labor productivity: Evidence from a two-stage randomized controlled trial in malawi. IZA Discussion Paper 10644, March 2017. URL `http://ftp.iza.org/dp10644.pdf`.

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

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.

Constance Newman and Lovell Jarvis. Worker and firm determinants of piece rate variation in an agricultural labor market. *Economic Development and Cultural Change*, 49(1):137–169, October 2000.

Sutti Ortiz. Laboring in the factories and in the fields. *Annual Review of Anthropology*, 31 (1):395–417, October 2002.

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.

Mark R Rosenzweig. *Handbook of development economics*, volume 1, chapter Labor markets in low-income countries, pages 713–762. Elsevier, 1988.

James Roumasset and Sang-Hyop Lee. *Handbook of Agricultural Economics*, volume 3, chapter Labor: Decisions, contracts and organization, pages 2705–2740. Elsevier, 2007.

Antoinette Schoar. The importance of being nice: Supervisory skill training in the cambodian garment industry. 2014.

Isaac Sitienei, Ashok K. Mishra, and Aditya R. Khanal. *Informal Ganyu Labor Supply, and Food Security: The Case of Malawi*, chapter 11, pages 159–175. 2016.

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

Martin Whiteside. Ganyu labor in Malawi and its implications for livelihood security interventions. Agricultural Research and Extension Network Paper No. 99, January 2000.

Table 1: BDM Outcomes

Minimum WTA ($\underline{PR}_i$)	Piece rate paid (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					

Table 2: Descriptive Statistics for Participants

	All (1)	Low Season (2)	High Season (3)	Diff. (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)	[0.018]
Age	34.9	34.6	35.2	0.6
	(13.6)	(13.2)	(14.1)	[0.6]
Number of adults in household	3.10	3.16	3.04	-0.12**
	(1.68)	(1.59)	(1.78)	[0.06]
Years of education	4.23	3.91	4.57	0.66***
	(3.27)	(3.35)	(3.16)	[0.13]
Female headed household	0.252	0.201	0.305	0.104***
	(0.434)	(0.401)	(0.461)	[0.017]
Participated in ganyu in last week	0.38	0.33	0.42	0.09***
	(0.48)	(0.47)	(0.49)	[0.02]
Days of ganyu last week, conditional on positive	3.77	4.22	3.39	-0.83***
	(2.07)	(2.15)	(1.93)	[0.13]
Daily wage from recent ganyu (MKW)	302.6	257.7	344.6	86.9***
	(329.9)	(178.7)	(420.9)	[13.1]
Household produces maize	0.999	1.000	0.997	-0.003
	(0.038)	(0.000)	(0.055)	[0.002]
Household produces beans	0.657	0.686	0.627	-0.059***
	(0.475)	(0.464)	(0.484)	[0.018]
Typical per year months without adequate food	3.35	3.56	3.13	-0.43***
	(2.27)	(2.34)	(2.16)	[0.09]
Alternative activity: housework	0.152	0.216	0.079	-0.137***
	(0.360)	(0.411)	(0.270)	[0.016]
Alternative activity: other ganyu	0.200	0.257	0.135	-0.122***
	(0.400)	(0.437)	(0.342)	[0.018]
Alternative activity: work own land	0.408	0.269	0.569	0.300***
	(0.492)	(0.444)	(0.495)	[0.022]
Alternative activity: work own business	0.084	0.049	0.124	0.075***
	(0.277)	(0.216)	(0.330)	[0.013]

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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: Descriptive Statistics on Participation, BDM and Work Outcomes

		Day of Week				Season		Gender		Monitoring	
	All Days (1)	Day 1 (2)	Day 2 (3)	Day 3 (4)	Day 4 (5)	Low (6)	High (7)	Male (8)	Female (9)	No (10)	Yes (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)	(22.0)	(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		
	(7.0)	(8.0)	(6.1)	(8.1)	(6.1)	(5.9)	(6.9)	(5.9)	(6.0)		
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 subjects attending	.694	.621	.684	.695	.778	.727	.669	.657	.712		
	[.008]	[.016]	[.015]	[.015]	[.014]	[.011]	[.01]	[.014]	[.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)

Notes: 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. Statistics in Panel A are not computed separately by monitoring treatment because monitoring was not assigned until the BDM stage. 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 consists of all workers, i.e,. all participants in BDM who received contracts. Standard deviations in parentheses.

Table 4: Predictors of Minimum Willingness to Accept

	Pooled		Random Effects		Fixed Effects
	(1)	(2)	(3)	(4)	(5)
Monitoring	-0.431	-0.546**	-0.297	-0.333	-0.189
	(0.269)	(0.265)	(0.221)	(0.221)	(0.235)
High season	-0.780**	-0.223	-0.901**	-0.362	
	(0.377)	(0.399)	(0.386)	(0.408)	
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	No
Mean Dep. Var.	10.320	10.320	10.320	10.320	10.320
SD Dep. Var.	5.899	5.899	5.899	5.899	5.899
Num. of participants	682	682	682	682	682
Num. of observations	1,857	1,857	1,857	1,857	1,857

Notes: this table presents regressions of minimum willingness to accept (WTA) on season, monitoring, whether the participant was female, and day-of-week fixed effects, with the first day as the omitted category. Columns (1)-(2) pool data across participants and days. Columns (3)-(4) include random participant effects. Column (5) includes participant fixed effects. Columns (2) and (4) control for individual covariates (age, number of adults in household, number of other household members participating, years of education, whether the head of household is female, days of ganyu in the previous week and month, reported daily wage from recent ganyu, household type of agricultural output (beans, tobacco, other), typical number of months per year without adequate food, household sources of income (ganyu, selling food products, selling beer, selling crafts, small shop), alternative activity for that day (housework, other ganyu, work own land, work in own business), number of units participant expects to sort). All regressions include district fixed effects, although in column (5) these are absorbed by the individual fixed effects. Standard errors clustered by participant. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: Determinants of Quantity (Units Sorted per Day)

	Pooled OLS			Random Effects		
	(1)	(2)	(3)	(4)	(5)	(6)
Monitoring	-0.587***	-0.664***	-0.440*	-0.576***	-0.647***	-0.689***
	(0.082)	(0.172)	(0.239)	(0.064)	(0.140)	(0.198)
Minimum WTA	-0.004	-0.008	-0.004	-0.022**	-0.027**	-0.022**
	(0.011)	(0.015)	(0.011)	(0.010)	(0.012)	(0.010)
Monitoring × Minimum WTA		0.009			0.008	
		(0.018)			(0.014)	
Piece rate	0.018***	0.018***	0.022**	0.026***	0.026***	0.023***
	(0.006)	(0.006)	(0.009)	(0.005)	(0.005)	(0.007)
Monitoring × Piece rate			-0.008			0.007
			(0.013)			(0.010)
High Season	0.479***	0.479***	0.480***	0.456***	0.455***	0.455***
	(0.134)	(0.134)	(0.134)	(0.131)	(0.130)	(0.131)
Female	0.805***	0.803***	0.806***	0.830***	0.828***	0.830***
	(0.143)	(0.144)	(0.143)	(0.142)	(0.142)	(0.142)
Second day	1.148***	1.149***	1.147***	1.160***	1.160***	1.161***
	(0.097)	(0.097)	(0.098)	(0.087)	(0.087)	(0.087)
Third day	1.799***	1.800***	1.798***	1.819***	1.822***	1.820***
	(0.104)	(0.105)	(0.104)	(0.093)	(0.094)	(0.093)
Fourth day	1.937***	1.939***	1.936***	2.016***	2.018***	2.016***
	(0.093)	(0.094)	(0.093)	(0.081)	(0.082)	(0.081)
Mean Dep. Var.	7.350	7.350	7.350	7.350	7.350	7.350
SD Dep. Var.	1.975	1.975	1.975	1.975	1.975	1.975
Num. of workers	612	612	612	612	612	612
Num. observations	1461	1461	1461	1461	1461	1461

Notes: this table presents regressions of quantity of output (number of units sorted per day) on whether the participant was assigned to the monitoring treatment, the piece rate the participant received (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. Columns (2) and (5) interact minimum WTA with monitoring; columns (3) and (6) interact the piece rate with monitoring. Columns (1)-(3) are estimated by pooled OLS, columns (4)-(6) with worker random effects. All regressions include controls for individual covariates as in Table 4, season fixed effects and district fixed effects. Standard errors robust to clustering at the worker level in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: Determinants of Quality (Errors per Unit Sorted)

	Pooled OLS			Random Effects		
	(1)	(2)	(3)	(4)	(5)	(6)
Monitoring	-0.647***	-0.575***	-0.744***	-0.614***	-0.561***	-0.722***
	(0.045)	(0.097)	(0.144)	(0.043)	(0.092)	(0.137)
Minimum WTA	-0.005	-0.001	-0.005	-0.003	-0.000	-0.003
	(0.006)	(0.008)	(0.006)	(0.006)	(0.008)	(0.006)
Monitoring × Minimum WTA		-0.008			-0.006	
		(0.010)			(0.009)	
Piece rate	0.009**	0.009**	0.006	0.009**	0.009**	0.006
	(0.004)	(0.004)	(0.006)	(0.004)	(0.004)	(0.006)
Monitoring × Piece rate			0.006			0.006
			(0.008)			(0.008)
High Season	-0.039	-0.038	-0.039	-0.050	-0.050	-0.051
	(0.063)	(0.063)	(0.063)	(0.065)	(0.065)	(0.066)
Female	-0.258***	-0.257***	-0.259***	-0.256***	-0.255***	-0.257***
	(0.072)	(0.072)	(0.072)	(0.075)	(0.075)	(0.075)
Second day	-0.275***	-0.275***	-0.274***	-0.293***	-0.293***	-0.292***
	(0.076)	(0.076)	(0.076)	(0.076)	(0.076)	(0.076)
Third day	-0.414***	-0.415***	-0.413***	-0.444***	-0.445***	-0.443***
	(0.073)	(0.074)	(0.073)	(0.072)	(0.072)	(0.072)
Fourth day	-0.464***	-0.466***	-0.464***	-0.491***	-0.493***	-0.491***
	(0.073)	(0.073)	(0.073)	(0.073)	(0.074)	(0.073)
Mean Dep. Var.	1.883	1.883	1.883	1.883	1.883	1.883
SD Dep. Var.	1.013	1.013	1.013	1.013	1.013	1.013
Num. of workers	612	612	612	612	612	612
Num. observations	1461	1461	1461	1461	1461	1461

Notes: this table presents regressions of quality of output (number of errors per unit) on whether the participant was assigned to the monitoring treatment, the piece rate the participant received (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. Columns (2) and (5) interact minimum WTA with monitoring; columns (3) and (6) interact the piece rate with monitoring. Columns (1)-(3) are estimated by pooled OLS, columns (4)-(6) with worker random effects. All regressions include controls for individual covariates as in Table 4, season fixed effects and district fixed effects. Standard errors robust to clustering at the worker level in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: Predictors of Willingness to Accept
Differential Effects by Gender

	Pooled		Random Effects	
	(1)	(2)	(3)	(4)
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)
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)
Female Level Effect	-2.259***	-2.836***	-2.514***	-3.031***
	(0.744)	(0.772)	(0.729)	(0.753)
Indiv. Controls	No	Yes	No	Yes
Mean Dep. Var.	10.320	10.320	10.320	10.320
SD Dep. Var.	5.899	5.899	5.899	5.899
Num. of participants	682	682	682	682
Num. of observations	1857	1857	1857	1857

Notes: this table presents differential effects by gender of season and monitoring on minimum WTA (MWK) per unit sorted. Additional regressors not reported: day-of-week fixed effects, district fixed effects, and, where noted, individual covariates as in Table 4. Columns (1)-(2) pool data across participants and days. Columns (3)-(4) include random participant effects. Standard errors are clustered by participant.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 8: Determinants of Quantity (Number of Units Sorted per Day)
Differential Effects by Gender

	Pooled OLS			Random Effects		
	(1)	(2)	(3)	(4)	(5)	(6)
Monitoring						
Among Men	-0.383***	-0.360	0.252	-0.376***	-0.196	-0.127
	(0.143)	(0.295)	(0.421)	(0.115)	(0.233)	(0.343)
Among Women	-0.662***	-0.792***	-0.729**	-0.654***	-0.811***	-0.901***
	(0.098)	(0.206)	(0.285)	(0.075)	(0.167)	(0.239)
Minimum WTA						
Among Men	-0.023	-0.022	-0.024	-0.048***	-0.039**	-0.049***
	(0.018)	(0.023)	(0.018)	(0.018)	(0.019)	(0.018)
Among Women	0.005	-0.002	0.005	-0.009	-0.019	-0.009
	(0.014)	(0.019)	(0.014)	(0.012)	(0.015)	(0.012)
Monitoring X Minimum WTA						
Among Men		-0.002			-0.019	
		(0.029)			(0.023)	
Among Women		0.015			0.019	
		(0.021)			(0.017)	
Piece rate						
Among Men	0.014	0.014	0.035**	0.027***	0.026***	0.035***
	(0.011)	(0.011)	(0.018)	(0.010)	(0.010)	(0.013)
Among Women	0.020***	0.020***	0.018	0.026***	0.026***	0.019**
	(0.008)	(0.008)	(0.011)	(0.006)	(0.006)	(0.008)
Monitoring X Piece rate						
Among Men			-0.036*			-0.014
			(0.022)			(0.017)
Among Women			0.004			0.014
			(0.015)			(0.013)
High Season						
Among Men	0.905***	0.904***	0.887***	0.962***	0.956***	0.956***
	(0.212)	(0.213)	(0.213)	(0.209)	(0.210)	(0.210)
Among Women	0.313*	0.311*	0.309*	0.260*	0.260*	0.253
	(0.160)	(0.159)	(0.160)	(0.154)	(0.154)	(0.155)
Female Level Effect	0.923**	0.997**	1.304***	0.985***	1.142***	1.241***
	(0.367)	(0.396)	(0.456)	(0.318)	(0.331)	(0.379)
Mean Dep. Var.	7.350	7.350	7.350	7.350	7.350	7.350
SD Dep. Var.	1.975	1.975	1.975	1.975	1.975	1.975
Num. of participants	612	612	612	612	612	612
Num. of observations	1461	1461	1461	1461	1461	1461

Notes: this table presents differential effects by gender of whether the participant was assigned to the monitoring treatment, the minimum piece rate the participant was willing to accept (Minimum WTA), the piece rate the participant received (Piece Rate), and season on quantity (number of units sorted per day). Columns (2) and (5) interact minimum WTA with monitoring; columns (3) and (6) interact the piece rate with monitoring. Columns (1)-(3) are estimated by pooled OLS, columns (4)-(6) with worker random effects. All regressions include controls for individual covariates as in Table 4, day-of-week fixed effects, and district fixed effects. Standard errors robust to clustering at the worker level in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

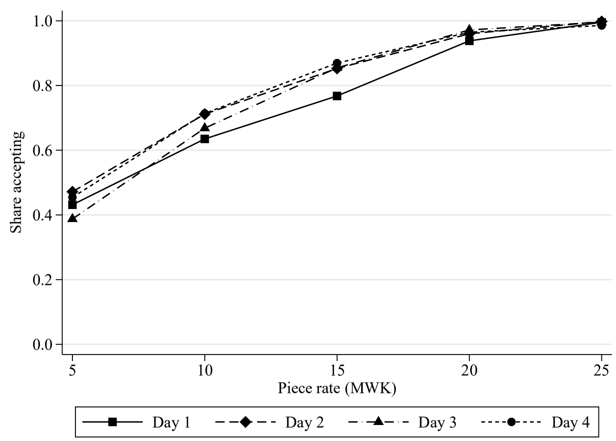
Table 9: Determinants of Quality (Number of Errors per Unit Sorted)
Differential Effects by Gender

	Pooled OLS			Random Effects		
	(1)	(2)	(3)	(4)	(5)	(6)
Monitoring						
Among Men	-0.648***	-0.446**	-0.672**	-0.589***	-0.402**	-0.761**
	(0.094)	(0.195)	(0.323)	(0.091)	(0.189)	(0.300)
Among Women	-0.653***	-0.628***	-0.725***	-0.631***	-0.627***	-0.674***
	(0.050)	(0.109)	(0.156)	(0.047)	(0.105)	(0.150)
Minimum WTA						
Among Men	-0.008	0.002	-0.008	-0.006	0.004	-0.006
	(0.011)	(0.017)	(0.011)	(0.011)	(0.016)	(0.011)
Among Women	-0.003	-0.001	-0.003	-0.002	-0.002	-0.002
	(0.007)	(0.009)	(0.007)	(0.007)	(0.009)	(0.007)
Monitoring X Minimum WTA						
Among Men		-0.021			-0.020	
		(0.019)			(0.018)	
Among Women		-0.003			-0.000	
		(0.011)			(0.011)	
Piece rate						
Among Men	0.022**	0.022**	0.022	0.021**	0.021**	0.015
	(0.009)	(0.009)	(0.014)	(0.009)	(0.009)	(0.014)
Among Women	0.004	0.004	0.002	0.005	0.005	0.003
	(0.004)	(0.004)	(0.007)	(0.004)	(0.004)	(0.007)
Monitoring X Piece rate						
Among Men			0.001			0.010
			(0.018)			(0.017)
Among Women			0.004			0.003
			(0.009)			(0.008)
High Season						
Among Men	-0.199	-0.205	-0.198	-0.199	-0.205	-0.195
	(0.126)	(0.125)	(0.125)	(0.129)	(0.128)	(0.128)
Among Women	0.036	0.038	0.034	0.021	0.023	0.020
	(0.065)	(0.065)	(0.065)	(0.068)	(0.068)	(0.069)
Female Level Effect	-0.107	-0.021	-0.085	-0.101	-0.015	-0.173
	(0.208)	(0.248)	(0.320)	(0.205)	(0.240)	(0.301)
Mean Dep. Var.	1.883	1.883	1.883	1.883	1.883	1.883
SD Dep. Var.	1.013	1.013	1.013	1.013	1.013	1.013
Num. of participants	612	612	612	612	612	612
Num. of observations	1461	1461	1461	1461	1461	1461

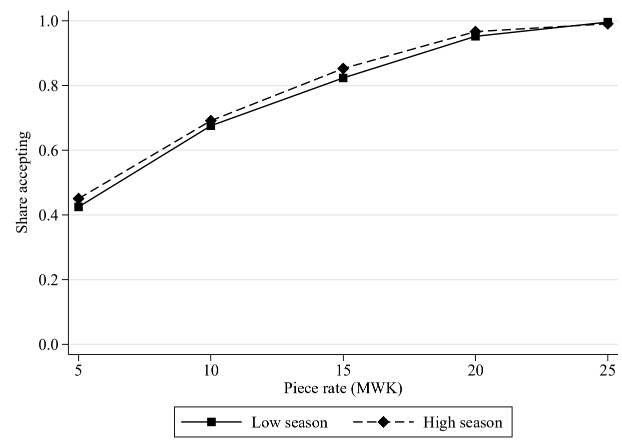
Notes: this table presents differential effects by gender of whether the participant was assigned to the monitoring treatment, the minimum piece rate the participant was willing to accept (Minimum WTA), the piece rate the participant received (Piece Rate), and season on quality (number of errors per unit sorted). Columns (2) and (5) interact minimum WTA with monitoring; columns (3) and (6) interact the piece rate with monitoring. Columns (1)-(3) are estimated by pooled OLS, columns (4)-(6) with worker random effects. All regressions include controls for individual covariates as in Table 4, day-of-week fixed effects, and district fixed effects. Standard errors robust to clustering at the worker level in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure 1: CDFs of minimum piece rate accepted

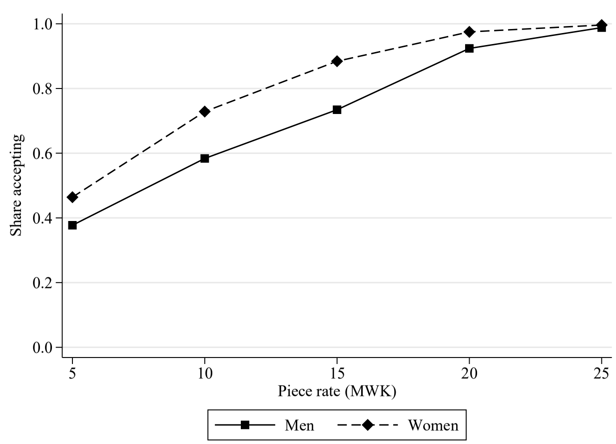
(a) By day of week



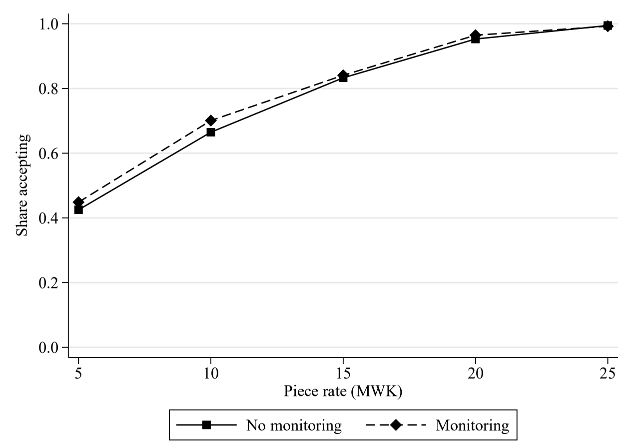
(b) By season



(c) By gender



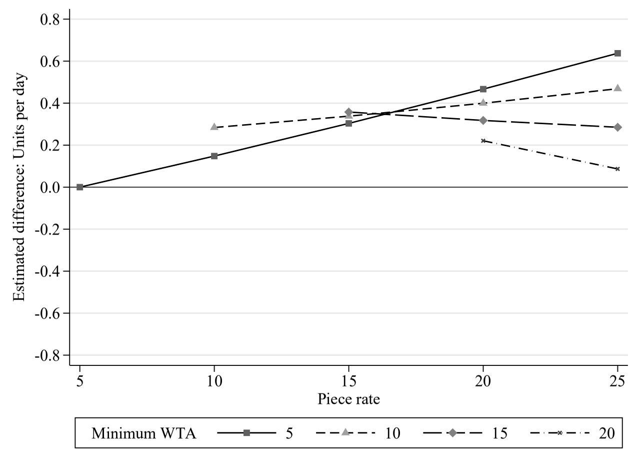
(d) By monitoring treatment



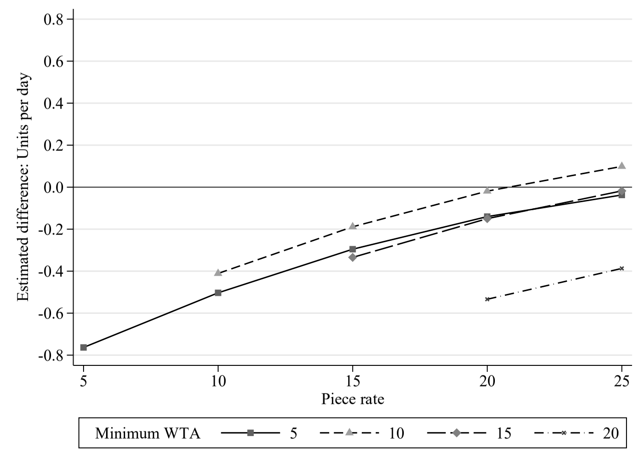
Notes: These figures plot the share of participants who agree to work at each of the given piece rates, i.e., those participants whose bids in BDM were as high or higher than that piece rate.

Figure 2: Selection and Incentives

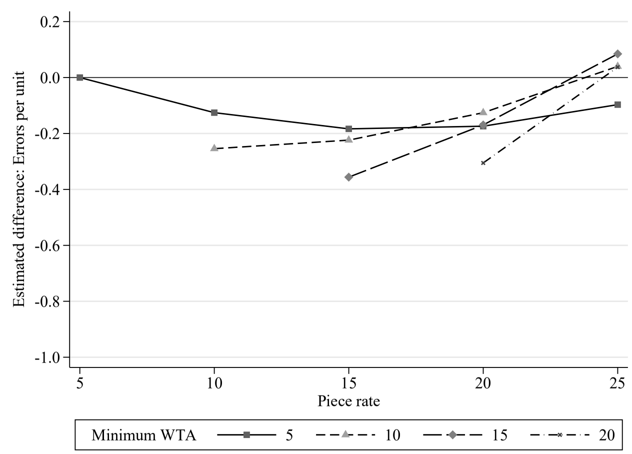
(a) Quantity — No monitoring



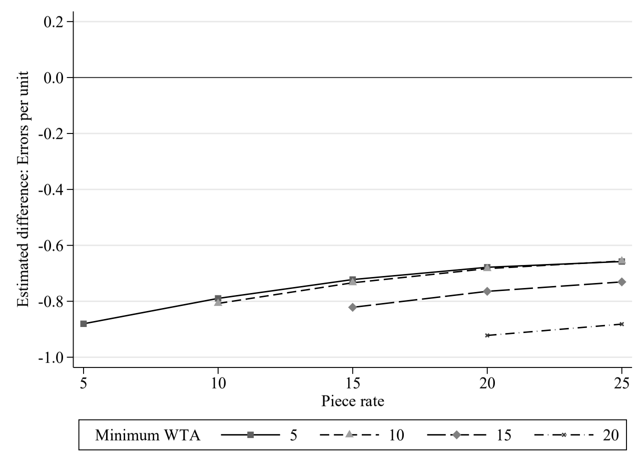
(b) Quantity — Monitoring



(c) Quality — No monitoring



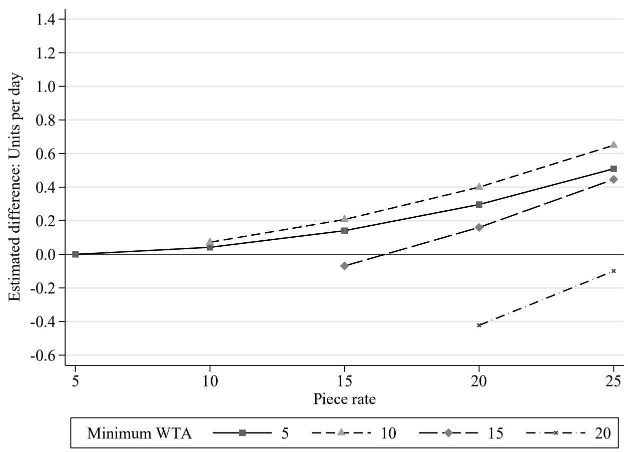
(d) Quality — Monitoring



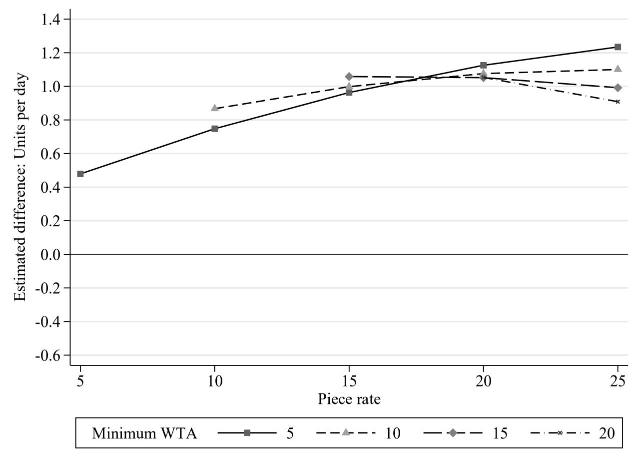
Notes: these figures plot estimated differences from regressions of either quantity (units sorted per day) or quality (errors per unit) on interacted quadratics in minimum willingness to pay and the piece rate (draw), interacted with monitoring, controlling nonparametrically for gender, season, district, day in the village and worker random effects. The estimated differences are relative to no monitoring, a minimum WTA of 5 MWK per unit sorted, and a piece rate of 5 MWK per unit sorted. The x-axis is the piece rate (draw), so each line compares workers who have the same minimum WTA but earn different piece rates.

Figure 3: Selection and Incentives by Gender

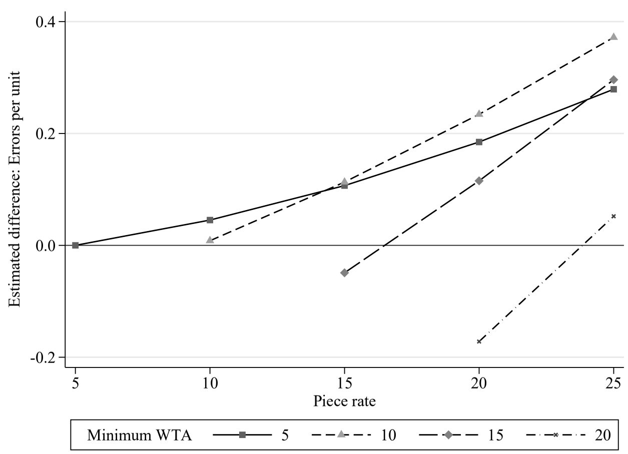
(a) Quantity — Male



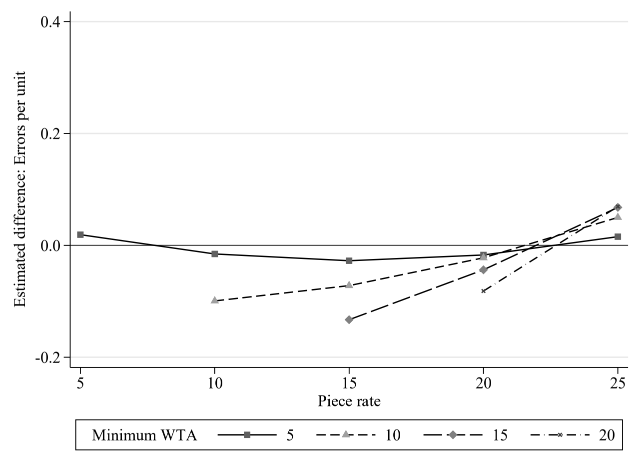
(b) Quantity — Female



(c) Quality — Male



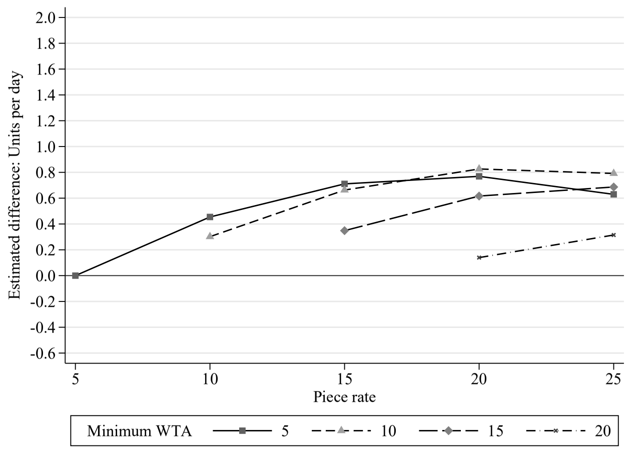
(d) Quality — Female



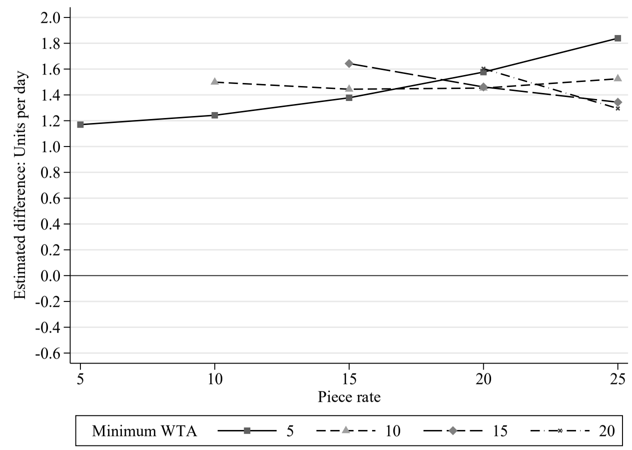
Notes: These figures plot estimated differences from regressions of either quantity (units sorted per day) or quality (errors per unit) on interacted quadratics in minimum willingness to pay and the piece rate (draw), interacted with gender, controlling nonparametrically for monitoring, season, district, day in the village and worker random effects. The estimated differences are relative to a male worker with a minimum WTA of 5 MWK per unit sorted and a piece rate of 5 MWK per unit sorted. The x-axis is the piece rate (draw), so each line compares workers who have the same minimum WTA but earn different piece rates.

Figure 4: Selection and Incentives by Gender
No monitoring

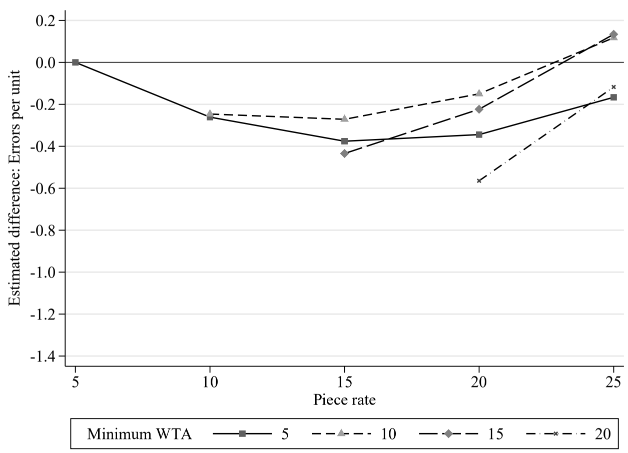
(a) Quantity — Male



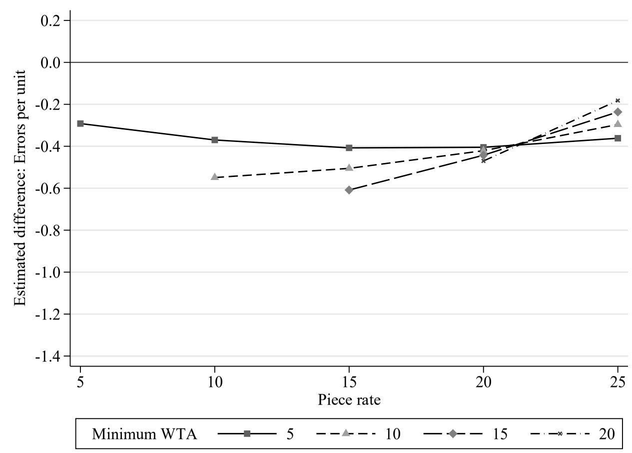
(b) Quantity — Female



(c) Quality — Male



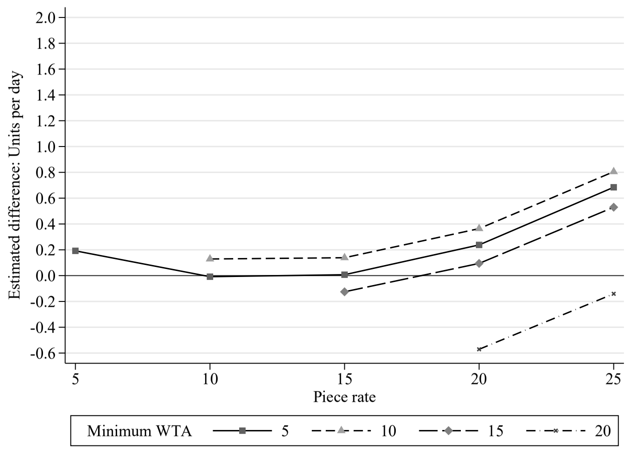
(d) Quality — Female



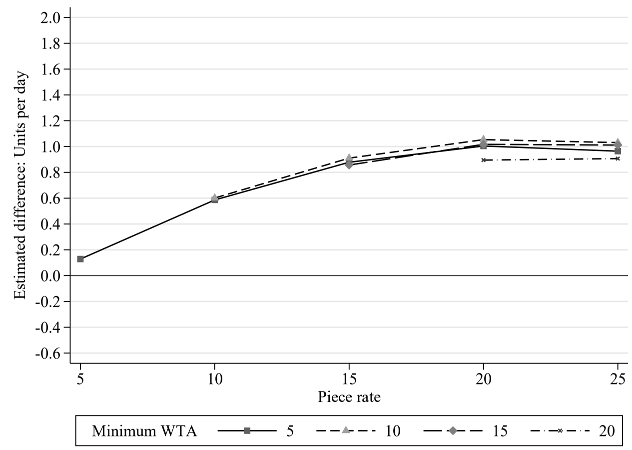
Notes: these figures plot estimated differences from regressions of either quantity (units sorted per day) or quality (errors per unit) on interacted quadratics in minimum willingness to pay and the piece rate (draw), interacted with gender and monitoring, controlling nonparametrically for season, district, day in the village and worker random effects. The estimated differences are relative to a male worker with a minimum WTA of 5 MWK per unit sorted and a piece rate of 5 MWK per unit sorted working in a no-monitoring contract. The x-axis is the piece rate (draw), so each line compares workers who have the same minimum WTA but earn different piece rates.

Figure 5: Selection and Incentives by Gender
Monitoring

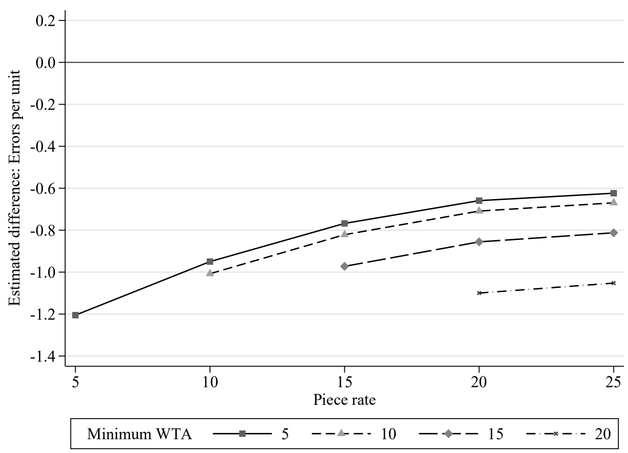
(a) Quantity — Male



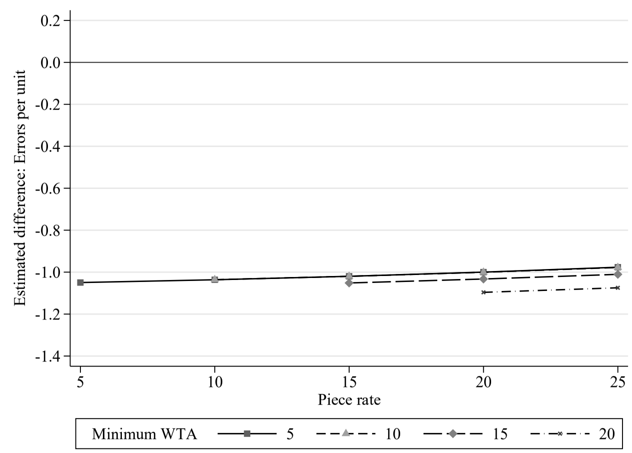
(b) Quantity — Female



(c) Quality — Male



(d) Quality — Female



Notes: these figures plot estimated differences from regressions of either quantity (units sorted per day) or quality (errors per unit) on interacted quadratics in minimum willingness to pay and the piece rate (draw), interacted with gender and monitoring, controlling nonparametrically for season, district, day in the village and worker random effects. The estimated differences are relative to a male worker with a minimum WTA of 5 MWK per unit sorted and a piece rate of 5 MWK per unit sorted working in a no-monitoring contract. The x-axis is the piece rate (draw), so each line compares workers who have the same minimum WTA but earn different piece rates.