Contents

List of Tables

List of Figures

A mi familia

Acknowledgements

I am grateful to my supervisor Charles Bellemare for his guidance and dedication. I look forward to collaborate with him in the years to come. I would also like to thank my co-supervisor Bruce Shearer, Arthur van Soest, who welcomed me at Tilburg University, and the members of my thesis committee: Marion Goussé, Sabine Kröger and Kim Lehrer.

Introduction

This thesis contains three essays in labour economics and applied econometrics presented as independent chapters. Collectively, they investigate how individuals adjust their choices to different environmental factors such as fatigue, working conditions or social interactions. Understanding how individuals integrate the surrounding environment into their own choices can help applied researchers to select more adequate models, design better incentive mechanisms, and improve their experimental designs.

The first two chapters, coauthored with Charles Bellemare and Bruce Shearer are closely related to personnel economics. They investigate how fatigue at work and productivity shocks may interact with workers' choice of effort and their observed outcome. Our empirical analysis is embedded in a principal-agent context in which a tree-planting firm (the principal) observes productivity but not the effort exerted by a worker (the agent).

Chapter 1 measures the causal effects of fatigue and rest on tree-planters productivity. The regular working week may not be optimal in terms of the spells of work and rest that maximize productivity. Firms and workers could benefit from alternative work schedules that improve productivity and earnings. The problem is that fatigue and rest are to some extent endogenous, therefore measuring their true causal impact is not straightforward. In our firm, planters who expect low productivity in a given day tend to take time off to recover. We do not observe the output of planters who experience physical discomfort or face other exogenous events affecting their productivity. Instead, we observe workers who are motivated enough for an exhausting day of tree-planting and who probably expect high earnings from their work. This self-selection makes the observed sample more productive, had workers not been allowed to take days off. Ignoring this endogeneity would lead to underestimation of the effects of fatigue and rest on productivity. To overcome this difficulty we use an instrumental variable strategy, we exploit national public holidays and the relocation of workers to planting sites as natural instruments. We found that an additional day of work significantly reduces productivity by around 9%, while an extra day of rest significantly increases average productivity. We use the estimates of a linear panel model to predict the productivity of tree planters under different work schedules. We find that the five consecutive work day schedule is not optimal for the firm. In particular, workers' productivity can be increased by up to 6.5% when days of rest are interspersed between shorter work spells. Moreover, our results suggest that workers' fatigue may be a source of bias in field experiments. Strong fatigue effects can potentially offset a treatment effect.

Chapter 2 studies how agents incorporate productivity shocks into their effort choice. Tree-planters often face idiosyncratic productivity shocks such as hard soil, rocky terrain or other working conditions that affect their outcomes. Investigating how productivity shocks affect planters' effort is relevant for designing optimal compensation systems. The structural form of the production function determines how workers may react to different economic incentives. If the productivity shock is additive, workers effort is exclusively determined by the incentives offered by the firm. If instead the production function has a multiplicative form, productivity shocks are taken into account in the optimization process and affect effort. In this case the responsiveness of the agents to incentives may vary depending on the shock they experience. Firms should take this behaviour into account when selecting their compensation system. In practice, testing additivity of the production function boils down to testing for heteroskedasticity in a linear production model. A major difficulty is that workers effort and the incentives offered by the firm are simultaneously determined. We overcome this identification problem by using field experiments, which induce variation on effort by exogenously changing workers' piece rate. We find that productivity shocks are not separable from other productivity determinants in the case of tree-planters. This means that productivity is better modeled by a multiplicative function of effort and a random productivity shock.

Chapter 3 explores an entirely different subject related to behavioural economics. I use a public good game conducted in rural Mali to investigate how individual choices react to two experimental treatments: the presence of a local leader and the possibility of communicating. I use expectations about total public goods provision to estimate individual preferences for conditional and unconditional cooperation. I find that both experimental treatments incentivize public goods provision, but they do it through different channels. Participation of local leaders effectively changes individual choices by increasing unconditional cooperation, while allowing participants to converse before they decide on contributions fosters conditional cooperation. This means that group communication ameliorate public good provision only when participants expect others to cooperate. In fact, communication may even worsen the outcome when expectations are low. I use the structural model to predict individual choices if expectations were different. I find that even in the most pessimistic scenario in which all participants expect zero public good provision, 60% would still choose to cooperate. Overall, expectations are responsible for around 24% of the observed contributions.

The three essays of this thesis highlight some of the strengths and weaknesses of the experimental approach in economics. On the one hand well-designed experiments can neatly solve observability problems. In this thesis experiments are used to induce variation in unobserved workers' effort and to elicit unobserved expectations about the actions of other individuals. On the other hand experiment convey potential risks. Their realism component, which makes them so appealing, can become a threat to their own validity. For instance, this thesis bring to attention workers' fatigue as an environmental factor that may bias experimental results when ignored in the field.

Even when making individual choices we are all influenced by our environment, our social context, our expectations, and a wide variety of external components. The true relevance of these factors and how to incorporate them without overcomplexifying the economic models is an empirical question left to applied researchers. This thesis is a small contribution to that task.

Chapter 1

Fatigue, Rest, Productivity, and Work Schedule: An empirical analysis using personnel records

Charles Bellemare María Adelaida Lopera Bruce Shearer

1.1 Introduction

Many firms and organizations offer jobs with fixed working schedules. Perhaps the most common schedule in western countries requires working five consecutive days before taking two days of rest. The five consecutive day schedule is believed to have been introduced in 1908 by a New England spinning mill in order to allow its Jewish workers to observe the Sabbath (see Rybczyński, 1991). Over time, other firms and nations have adopted similar work schedules primarily as a way to harmonize work practices, and not necessarily because it maximizes worker productivity.

Worker productivity may decrease because of accumulated fatigue and insufficient rest. When strong enough, the negative effects of fatigue and the positive effects of resting may require that firms adjust their work schedules in order to increase productivity (see Saez, 2011, for a recent theoretical analysis). This can be especially important for physically demanding jobs.

Empirical evidence on the relationship between worker productivity, fatigue, and rest is rather limited in economics. Hamermesh (1990) estimates the marginal effect of on-the-job rest on wages using panel data on self-reported time allocation. He concludes that the first few minutes of rest increase subsequent productivity (wages). However, this increase is just enough to compensate for the non-working period, and longer break times are predicted to reduce productivity. Biddle and Hamermesh (1990) model sleep as a choice variable jointly determined with wages and leisure. Their results suggest that the relationship between sleep and wages has an inverted-U shape.

In this paper we focus on measuring the relationship between productivity, fatigue, and rest. We do so by analyzing the payroll records of a Canadian tree-planting firm operating in British Columbia. Tree-planting is a simple but physically demanding job. As a result, accumulated fatigue and rest are two potentially important determinants of worker productivity. Estimating the effects of rest and fatigue is in principle straightforward when using payroll records of the firm. These records contain detailed information on worker productivity and can be used to construct episodes of work and rest for each worker throughout the season. However, workers in our firm can decide to take days off during the working week, thus partially determining their own fatigue and rest at any given point in time. If the decision to take a day off depends on unobservable shocks correlated with productivity, measured fatigue and rest are both potentially endogenous explanatory variables determining productivity. Omitting to take account of this endogeneity may result in biased estimates of the impact of fatigue and rest on productivity. We address this issue by exploiting two natural instruments. First, public holidays occur during the planting season and provide workers with compulsory rest during the workweek. Second, workers are assigned to crews, with each crew assigned to plant on a given block. Each crew typically completes work on its assigned block before being relocated to the following scheduled block. In some cases, relocation of a crew and its equipment requires several days, forcing workers to take longer resting periods. We show that both instruments are significantly correlated with our measures of fatigue and rest.

We find that the estimated effects of fatigue and rest are weak and in some cases insignificant when both variables are assumed exogenous and their effects are estimated by ordinary least squares. In particular, we find that an extra day of rest has no significant impact on average daily productivity, while an additional day of work significantly reduces productivity by 1%. Our instrumental variable approach yields substantially stronger estimates. We find that an extra day of rest significantly increases average productivity by 4.2% or 5.8% depending on the specification estimated. An additional day of work is predicted to significantly reduce productivity by 9.8% or 9.1%, depending again on the specification estimated. Our results are consistent with workers taking days off when faced with a negative productivity shock. This implies that the observed sample is more productive than it would otherwise be if workers were not allowed to take days off during the week. As a result, ordinary least squares estimates using the observed sample underestimate the effects of rest and fatigue.

We use our estimated model to predict worker productivity under alternative work schedules, varying the length (in days) of the work and rest spells. We find that shorter work spells can increase worker productivity by up to 6.5% relative to the baseline schedule of working five consecutive days. These results highlight the potential gains that can be achieved by using more flexible work schedules. This paper is organized as follows. Section 1.2 presents an overview of the tree-planting and the firm we analyze. Section 1.3 describes our data. Section 1.4 presents the models we estimate. Section 1.5 presents our estimation results. Section 1.6 discusses the predictions of our model. Section 1.7 summarizes and concludes.

1.2 Tree-planting

Our data come from a mid-sized tree-planting firm operating in British Columbia, Canada. This province is the largest producer of timber in North America; therefore, extensive reforestation is central to a steady supply of the market. Typically, tree-planting firms are chosen to plant seedlings on harvested tracts through a process of competitive bidding. Depending on the land tenure arrangement, either a timber-harvesting firm or the Ministry of Forests and Range call for sealed bids concerning the cost per tree planted in a number of areas. Forestry firms estimate the cost at which they can complete each contract and submit offers. The lowest bidders are selected to perform the work. Bidding for contracts takes place in the late autumn. After this process, the selected firms commit to reforest their corresponding areas dispersed across the province. The following year, from early spring to late summer, the firms fulfill their planting contracts.

Our particular firm divides each area into homogenous *blocks* previous to the planting. After reviewing conditions on a particular block, the manager assigns a piece rate to be paid to all workers planting on that block. This rate takes into account the expected number of trees that a worker can plant. Steep or rocky terrain slows planters, rendering planting more difficult than in flat terrain or smooth soil. To compensate for the effort needed to plant, the piece rate is higher in difficult terrain. Blocks are in their turn divided into *plots*, each of which is allocated to a planter during a field-day. Workers are hired on seasonal contracts; they live near the planting area in accommodations provided by the firm. There are no penalties for occasional absenteeism aside from the forgone earnings of the day. Apart from the weeks that include statutory holidays, on which the firm cannot operate by law, most planters work five days a week.

At the start of a field-day, the manager assigns each planter that is present for work to a single plot. There is no systematic matching of workers to planting conditions. Once everyone is assigned, each worker receives a box full of seedlings and a shovel. A truck transports planters to their individual sites, where they spend the day. The manager evaluates the plots afterwards to ensure quality. Poorly planted trees must be replanted at the planter's expense. However, incidents involving poor quality are rare. The task of planting a tree consists of digging a hole, pacing a seedling and covering its roots with soil. The simplicity and homogeneity of the activity facilitates measuring workers productivity. Since planters are paid piece rate, daily earnings are strictly proportional to the number of trees planted during a given day. Upon completion of planting on the block, workers are relocated to new blocks.

1.3 Data

Our data consist of an unbalanced panel containing 5,102 observations on 155 workers who planted during the 2005 and 2006 planting seasons. Table 1.1 summarizes the structure of our dataset and the main variables used in our analysis.

Table 1.1 – Summary statistics

	mean (a)	std. dev. (b)	min. (c)	max. (d)
productivity: ln(wage)	5.249	0.362	3.401	6.305
fatigue: days of consecutive work	1.171	0.532	1	5
rest: days off on last leave	2.265	1.188	1	5
average working "week"	3.641	1.372	1	9
number of workers	155			
number of periods per worker	5 - 166			
observations	5,102			

Daily productivity is measured by the natural logarithm of daily individual earnings. This variable controls for the working conditions because piece rates are based on the required effort to plant. Our two main explanatory variables are *fatigue* and *rest*. The variable *fatigue* reflects the number of days of consecutive work since the last day off. It takes discrete values from 1 to 5 in our sample, with an average of 1.17. A more informative variable is the *average working "week"*, which shows the duration of working periods. The average working "week" is 3.6 days, which is lower than a regular working week of 5 days. This difference reflects the fact that planters regularly take one or two days off, presumably when they feel tired or when they foresee a day if low productivity. The variable *rest* represents the length (in days) of the last resting period and ranges from 1 to 5 in the sample with an average rest of 2.2 days, which roughly corresponds to the length of a normal weekend. Our coding implies that a planter who rested for two days during the weekend and worked from Monday to Friday is coded as $rest = 2$ and $fatigue = 3$ for his work on Wednesday, and $rest = 2$ and $fatigue = 5$ for his work on Friday.

The observed values of *fatigue* and *rest* depend partially on workers' decision to extend the weekend or to take a day off in between workdays. This decision may in turn depend on workers' anticipated productivity. In particular, a worker may wake up on a given day with physical pain (i.e. a negative productivity shock) and decide to rest instead of working. As a result, *fatigue* and *rest* are both potentially endogenous explanatory variables. To address this issue, we use two instrumental variables labelled *holiday* and *contract*, both constructed from operational constraints of the firm.

The instrument *holiday* is a dummy variable that takes the value of 1 when the current week counts a national holiday and 0 otherwise. By law, the firm closes on statutory holidays, forcing all workers to take a day off. There are 7 public holidays during the planting season, they affect the working week of 42% of the planters in our dataset and are all effective either on Friday or Monday. For a given weekday (say Wednesday), workers *fatigue* will be lower when a holiday took place the previous Monday. Ceteris paribus, holidays increase workers rest as they extend the duration of weekends.

The instrumental variable *contract* is a dummy variable that takes the value of 1 during the first week of planting on a new block and 0 otherwise. As explained in section 1.2, firms obtain tree-planting contracts across the province through a process of competitive bidding. The planting areas are not necessarily located near each other, forcing the firm to relocate workers and equipment at the beginning of a new contract. Relocation can take several days, thus increasing workers *rest* and independently of other productivity determinants. The firm tends to schedule planting for a longer time period during the first week of planting on a new block (starting work on a Sunday for example). This is done to compensate for the additional rest caused by the relocation to the new site. In our dataset, all workers are observed at the beginning of a contract at some point during the season. A formal exogeneity test is not possible because our model is just-identified, there are two endogenous variables and two available instruments.

1.4 Model

We are interested in the causal effects of *fatigue* and *rest* on the *productivity* of worker $i = 1, \dots, N$ in period $t = 1, \dots, T_i$:

$$\mathit{productivity}_{it} = \gamma'^p X_{it} + \alpha \, \mathit{fatigue}_{it} + \beta \, \mathit{rest}_{it} + \eta_i^p + \varepsilon_{it}^p. \tag{1.1}$$

The matrix X_{it} contains exogenous determinants of $\mathit{productivity}_{it}$ such as age or gender. Nevertheless, workers with the same observed characteristics could exhibit different outcomes due to other unobserved factors. The individual-specific component η_i^p controls for the unobserved heterogeneity of workers characteristics. The term ε_{it}^p, accounts for all other factors and is assumed to be an independent and identically distributed random variable.

The true worker's *fatigue* and *rest* are difficult to measure and usually unknown to the researcher. Instead, it is easy to observe their discrete counterparts. We could think for example of the number of days worked in a row as the observed $\mathit{fatigue}_{it}$, and the number of days off during the last leave as rest_{it}. A common solution for modeling variables similar to the latent $\mathit{fatigue}_{it}^*$ and rest_{it}^*, is to assume that individuals make an ordered discrete choice

$$\begin{aligned} fatigue_{it} &= \begin{cases} 1 & \text{if} \quad fatigue_{it}^* \leq f_1 \\ 2 & \text{if} \quad f_1 < fatigue_{it}^* \leq f_2 \\ 3 & \text{if} \quad f_2 < fatigue_{it}^* \leq f_3 \\ 4 & \text{if} \quad f_3 < fatigue_{it}^* \leq f_4 \\ 5 & \text{otherwise,} \end{cases} \\ rest_{it} &= \begin{cases} 1 & \text{if} \quad rest_{it}^* \leq r_1 \\ 2 & \text{if} \quad r_1 < rest_{it}^* \leq r_2 \\ 3 & \text{if} \quad r_2 < rest_{it}^* \leq r_3 \\ 4 & \text{if} \quad r_3 < rest_{it}^* \leq r_4 \\ 5 & \text{otherwise.} \end{cases} \end{aligned} \tag{1.2}$$

The observed values correspond to the events of the underlying continuous variables crossing thresholds. For example, when $\mathit{fatigue}_{it}^*$ is larger than f_2 and lower than f_3, the planter works for their third day in a row and we observe $\mathit{fatigue}_{it} = 3$. The usual normalization to fix the scale in this type of model is to set $f_1 = r_1 = 0$.

In order to estimate their true causal effects, the variables $\mathit{fatigue}_{it}$ and rest_{it} must be uncorrelated with the error term as well as with the individual effects in equation (1.1). In practice, this is a strong assumption (see Section 1.3 for a discussion). To account for this potential endogeneity we build a rich model with three interrelated equations

$$\begin{aligned} \mathit{productivity}_{it} &= \gamma'^p X_{it} + \alpha \, \mathit{fatigue}_{it} + \beta \, \mathit{rest}_{it} + \eta_i^p + \varepsilon_{it}^p \\ \mathit{fatigue}_{it}^* &= \gamma'^f X_{it} + \delta'^f Z_{it} + \eta_i^f + \varepsilon_{it}^f \\ \mathit{rest}_{it}^* &= \gamma'^r X_{it} + \delta'^r Z_{it} + \eta_i^r + \varepsilon_{it,}^r \end{aligned} \tag{1.3}$$

where the matrix Z_{it} contains at least two valid instruments that influence worker *fatigue* and *rest* but not their productivity. This model is a system of equations that allows two forms of unobserved correlation: through the error term $\varepsilon_{it} = [\varepsilon^p_{it}, \varepsilon^f_{it}, \varepsilon^r_{it}]'$ and through the individual-specific component $\eta_i = [\eta^p_i, \eta^f_i, \eta^r_i]'$. The latter accounts for the unobserved correlation within individuals. If these individual characteristics are independent of the rest of regressors (*strong exogeneity*), we can model η_{it} as a random vector that distributes multivariate normal

$$\begin{pmatrix} \eta^p_i \\ \eta^f_i \\ \eta^r_i \end{pmatrix} \sim \mathbf{M}\mathbb{N}\left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_p & \rho^\eta_{pf}\sigma_p\sigma_f & \rho^\eta_{pr}\sigma_p\sigma_r \\ & \sigma^2_f & \rho^\eta_{fr}\sigma_f\sigma_r \\ & & \sigma^2_r \end{pmatrix}\right]. \tag{1.4}$$

Here, σ^2_p, σ^2_f and σ^2_r denote the variances of the unobserved heterogeneity components in the system of equations (1.3), and the parameters ρ^η_{pf}, ρ^η_{pr}, and ρ^η_{fr} represent their correlations. These correlations are indicative of whether or not productive workers tend to choose different work schedules due to their unobservable individual characteristics. A significant and positive ρ^η_{pf} indicates that more productive planters prefer to work more days in a row and accumulate more *fatigue*. A significant and positive ρ^η_{pr} means that more productive planters take longer periods of *rest*. Similarly, a positive and significant ρ^η_{fr} indicates that planters who accumulate more fatigue also tend to accumulate more days of consecutive rest. In practice, these correlations can take any sign.

Finally, we assume that the model in (1.3) is a mixed ordered probit in which the error terms jointly follow the multivariate normal

$$\begin{pmatrix} \varepsilon^p_{it} \\ \varepsilon^f_{it} \\ \varepsilon^r_{it} \end{pmatrix} \sim \mathbf{M}\mathbb{N}\left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & \sigma\rho_{pf} & \sigma\rho_{pr} \\ & 1 & \rho_{fr} \\ & & 1 \end{pmatrix}\right]. \tag{1.5}$$

On the diagonal of the covariance matrix, σ^2 represents the productivity variance. We set $\mathrm{Var}(\varepsilon^f_{it}) = \mathrm{Var}(\varepsilon^r_{it}) = 1$ for identification purposes. The plain covariances indicate that a single random shock in one of the three equations can "propagate" to the others through the correlation components ρ_{pf}, ρ_{pr}, and ρ_{fr}.

1.5 Estimation Results

Different approaches can be used to estimate the effect of *fatigue* and *rest* on productivity. A naïve method is to estimate equation (1.1) using panel data estimators such as fixed effects (FE) or random effects (RE). As we know, these estimators produce biased and inconsistent results due to the endogeneity of *fatigue* and *rest*. We compute FE and RE to measure the magnitude of association between regressors and productivity, but not causality.

To measure causal effects we estimate the system of equations (1.3). We combine Instrumental Variables (IV) and RE to obtain unbiased estimates. The variables *holiday* and *contract* serve as instruments to identify the causal effects of *fatigue* and *rest*. The RE account for individual-specific factors that may affect observations over several periods. This two-steps approach has at least two limitations. First, it does not take into account the discrete nature of *fatigue* and *rest*. Second, it requires the model equations to be correlated only through the endogenous variables. The error variances ε_{it} are assumed to uncorrelated between equations, while the individual effects η_i are simply ignored in the *fatigue* equation and the *rest* equation.

Simultaneous estimation of our system of equations is a more flexible approach. We compute Maximum Likelihood (ML) estimators over simulated RE to estimate our model in (1.3). See Appendix A.1 for estimation details. This estimator captures the discrete nature of the endogenous variables in equation (1.2) and permit our system to be correlated through its unobserved components. This correlation is captured by the full variance matrix of individual factors in equation (1.4) and the error variance in equation (1.5).

Table 1.2 – Estimation results

	endogeneity ignored		endogeneity of *fatigue* and *rest* taken into account			
	FE	RE	IV-RE	system of equations with RE		
	(a)	(b)	(c)	(d)	(e)	(f)
dependent variable	*prod.*	*prod.*	*prod.*	*prod.*	*fatigue*	*rest*
fatigue	-0.0125*** (0.005)	-0.0110** (0.005)	-0.0979** (0.041)	-0.0913*** (0.020)		
rest	-0.0007 (0.004)	-0.0017 (0.004)	0.0423** (0.021)	0.0578*** (0.009)		
holiday (IV)					-0.2414*** (0.079)	0.1641*** (0.051)
contract (IV)					0.3215*** (0.051)	0.7699*** (0.032)
Tuesday	0.0419*** (0.011)	0.0410*** (0.011)	0.0702*** (0.019)	0.0619*** (0.023)	0.6093*** (0.103)	0.1120 (0.125)
Wesnesday	-0.0038 (0.014)	-0.0063 (0.014)	0.0830* (0.044)	0.0712*** (0.026)	1.1916*** (0.122)	0.0726 (0.094)
Thursday	0.0589*** (0.014)	0.0557*** (0.015)	0.2020*** (0.071)	0.1868*** (0.036)	1.7449*** (0.090)	0.0176 (0.122)
Friday	-0.0320* (0.018)	-0.0362* (0.019)	0.1648* (0.096)	0.1442*** (0.045)	2.4324*** (0.061)	-0.0634 (0.117)
April	0.1092*** (0.021)	0.1003*** (0.021)	0.0937*** (0.022)	0.0667*** (0.024)	0.1050 (0.103)	1.0201*** (0.062)
May	0.2351*** (0.019)	0.2203*** (0.019)	0.1973*** (0.023)	0.1590*** (0.021)	0.1178 (0.087)	1.7792*** (0.055)
June	0.2361*** (0.020)	0.2212*** (0.019)	0.1854*** (0.026)	0.1455*** (0.020)	-0.0708 (0.086)	1.7506*** (0.062)
July	0.1620*** (0.027)	0.1483*** (0.027)	0.0569 (0.051)	-0.0040 (0.036)	-0.0953 (0.119)	2.7763*** (0.086)
constant		5.0354*** (0.028)	5.0524*** (0.031)	4.8329*** (0.024)	-0.6093*** (0.093)	-0.9359*** (0.078)
discret choice thresholds of the the dependent variables						
threshold 2					0.8249*** (0.055)	1.7533*** (0.024)
threshold 3					1.5780*** (0.066)	2.4907*** (0.031)
threshold 4					2.4827*** (0.071)	2.7712*** (0.033)
no. of parameters	11	13	13	49		
estimation	mean diff.	FGLS	2SLS	ML, simulated RE		

Significance: *: 10%, **: 5%, ***: 1%.

Table 1.2 reports estimates of planters productivity using the different estimators discussed above. All specifications condition on a set of dummy variables that control for the day of the week and the month of the year.

The first two columns present generalized least squares estimators of the productivity equation (1.1). Column (a) shows FE and column (b) RE. When *fatigue* increases there is a small but significant reduction in planters productivity. According to the FE estimates, each additional day of work is associated with a 1.25% reduction in productivity (significant at 1% level). The RE estimates show a 1.10% reduction in productivity (significant at 5%). Neither FE nor RE results show any linear relationship between rest and productivity. Their respective estimates are -0.0007 and -0.0017, none of them significant at 10% level. The bias in FE and RE estimates is misleading. The small correlation between productivity and *fatigue*, and the weak correlation between productivity and *rest* does not mean that their effects are not important. The true causal effect of these variables becomes evident once their endogeneity is taken into account.

The FE and RE estimates show that the days of the week and the months of the year are significantly correlated to productivity. From Monday to Tuesday planters' increase their productivity by 4%. Even though there is no significant change on Wednesday, on Thursday productivity increases by almost 6% with respect to the beginning of the week. On Friday, productivity draws back to -3.5% with respect to Monday. No clear pattern emerges from these parameters and their estimates change depending on the estimator used.

The parameters associated to the month of the year describe a hump-shaped curve and are all significant at 1% level. Productivity increases as the planting season goes on and slightly declines back towards the end. In April, productivity increases by 10% with respect to March. In May and June, planters reach their maximum of productivity: 22% on the FE specification and 23% on the RE. By the end of the season, in July, workers slightly reduce their productivity, but their outcome is still 15% or 16% larger than at the beginning of the planting season depending on the specification. These results could be associated to temperature variations over the summer, to an initial learning effect and latter exhaustion, or both.

Column (c) shows IV-RE estimates of model (1.3), which uses *holiday* and *contract* as instruments to identify *fatigue* and *rest*. These results are provided after verifying that RE estimators of *fatigue* and *rest* are consistent. A Hausman test shows that the equality between consistent IV-FE and efficient IV-RE cannot be rejected for the variables *fatigue* and *rest* ($p-$value $=0.16$).[1] This means that RE estimator is also consistent and that unobserved factors specific to individuals are orthogonal to our two variables of interest.

IV-RE results suggest that there is a considerable downward bias in the estimated effects of *fatigue* and *rest* when their endogeneity is ignored. The estimated impact of *fatigue* and

[1] Formal implementation of the Hausman test requires estimation of $\text{Var}(\hat{\theta}_{FE} - \hat{\theta}_{RE})$, where $\hat{\theta}_{FE}$ are IV-FE and $\hat{\theta}_{RE}$ IV-RE. We approximate this variance using bootstrap methods. We generated 400 bootstrap samples by drawing individuals with replacement from our original sample, estimate $\hat{\theta}_{RE}$, $\hat{\theta}_{RE}$, and calculate the difference for the variables *fatigue* and *rest*. Our estimate of the variance matrix is the sample variance of this difference.

rest on worker productivity using IV-RE is substantially higher than the estimated impact using FE or RE. Intuitively, this bias arises because observed productivity is higher that it would have been if workers were not allowed to take days off during the week. A worker who experiences a negative productivity shock in a given day (for example physical discomfort, pain, or a cold) tends to take a day off instead of working. This means that observed *rest* and *fatigue* are the result of a choice based on expected productivity. When this endogeneity is addressed, each additional day of work reduces planters productivity by 9.8% (at 5% level of significance). This means that tree planters who start working on Monday would have reduced their productivity by 39% by Friday. Moreover, an extra day of *rest* increases daily productivity by 4.2% (significant at 5% level). Even though IV addresses the endogeneity issue, it has some limitations. First, its two steps estimate does not take into account the discrete nature of the endogenous regressors, and second, the IV specification lacks of flexibility when capturing the relationship between worker productivity, *fatigue*, and *rest*. Simultaneous estimation of our three model equations overcomes these limitations.

The last three columns present the estimates of our system of equations (1.3) using ML and simulated RE. Column (d) shows productivity equation estimates, which is our main regression of interest. Despite the larger number of parameters in the model, estimates of this productivity equation are more precise with respect to the IV-RE. Column (e) corresponds to the equation of the endogenous variable *fatigue* and column (f) to the *rest* equation. Similarly to the IV estimation, the variables *holiday* and *contract* play the role of exclusion restrictions for identification.

Column (d) indicates that each additional day of work (*fatique*) reduces planters productivity by 9.13%, while an extra day of *rest* increases daily productivity by 5.78%. Both estimates are significant at 1% level.

ML estimates of the control variables are similar to the IV-RE. The days of the week describe a hump-shaped progression when compared with average productivity on Mondays. Productivity initially increases by 6.2% on Tuesday, then by 7.1% on Wednesday, 18.7% on Thursday, and slightly back to 14.4% on Friday. These estimates are quite large and all significant at 1%. Productivity also exhibits significant changes throughout the planting season. Workers initially improve their productivity by 6.7% in the first month of work and reach a pick in May, when their productivity becomes 16% larger. A slight decline starts in June, when the difference with respect to March is 14.5%, and by July, productivity goes back to a level similar to the observed at the beginning of the planting season.

Our two instruments *holiday* and *contract* exhibit strong correlations with the two endogenous variables. In the *fatigue* equation in column (e) the estimate of the instrument *holiday* is -0.24 and is significant at 1% level. As expected, worker *fatigue* is lower when a holiday took place the previous week. Inversely, the estimate of the instrument *contract* is 0.32, also significant at 1%. This shows that at the beginning of a contract planters tend to work for longer periods, probably to compensate for the additional *rest* induced by the relocation of the planting site. Workers may try to "recover" some of the forgone earnings corresponding to the days off in-between contracts by working more. As for the *rest* equation in column (f), the two instruments are positive and statistically significant at 1%. The parameter associated to *contract* is 0.77, corroborating that planters are forced to take a few days off before the beginning of a contract. The instrument *holiday* is 0.16, which means that statutory holidays increase workers *rest* as they extend the duration of the weekends.

The days of the week and the months of the year are also included as control variables in the last two regressions of the system. While planters significantly increase their *fatigue* throughout the workweek, the variable *rest*, measured by the number of days off during the last leave, remains statistically unchanged. Inversely, the months of the year do not affect the variable *fatigue* but increase workers' resting periods. This means that workers tend to take more days off as the planting season goes by, probably to cope with the effects of physical exhaustion.

Our simultaneous equations model takes into account the discrete nature of the measured *fatigue* and *rest* and therefore we can estimate the threshold parameters described in equation (1.2). They represent the points at which the latent counterparts change their observed value.

The variance of the unobserved components of our model (ε_{it} and η_i) are not of direct interest. Nonetheless, we can add flexibility to our system by allowing these components to correlate across equations. Table 1.3 reports the estimates of the nuisance parameters in our model. First, we discuss the variance of the error components ε_{it} and then the variance of the unobserved individual factors η_i.

Table 1.3 – Estimated nuisance parameters

	endogeneity ignored		endogeneity taken into account	
	FE	RE	IV with RE	syst. of eqns. with RE
	(a)	(b)	(c)	(d)
random error, covariance elements				
Var(ε^p)	0.2764*** (0.002)	0.2815*** (0.002)	0.2986*** (0.002)	0.2970*** (0.006)
Cov($\varepsilon^p, \varepsilon^f$)				0.3098*** (0.067)
Cov($\varepsilon^p, \varepsilon^r$)				-0.2254*** (0.040)
Cov($\varepsilon^f, \varepsilon^r$)				-0.0620** (0.028)
individual-specific effects, covariance elements				
Var(η^p)		0.0700*** (0.004)	0.0728*** (0.004)	0.1094*** (0.009)
Cov(η^p, η^f)				0.0512*** (0.009)
Cov(η^p, η^r)				0.0728*** (0.011)
Var(η^f)				0.0242*** (0.007)
Cov(η^f, η^r)				0.0301*** (0.009)
Var(η^r)				0.2046*** (0.028)

Standard errors computed using the Delta method.
Significance: *: 10%, **: 5%, ***: 1%.

The variance of the error term in the productivity equation $\text{Var}(\varepsilon^p)$ is similar across specifications and always significant at 1%. Its estimates range from 0.28 for the FE in column (a) and RE in column (b), to 0.30 for the IV-RE in column (c) and the system of equations in column (d). The latter estimator is the only one that allows us to estimate the full covariance structure between productivity, *fatigue* and *rest*. In general, error covariances are all statistically significant. Productivity shocks directly affect the number of days of consecutive work and inversely affect the number of days of rest. The estimated $\text{Cov}(\varepsilon^p, \varepsilon^f)$ is 0.31 and $\text{Cov}(\varepsilon^p, \varepsilon^r)$ is -0.22, both significant at 1%. This means that a worker who experiences a negative productivity shock in a given day tends to take more *rest* and cumulate less *fatigue*. We also find $\text{Corr}(\varepsilon^f, \varepsilon^r)$ to be -0.06, significant at 5%. This means that shocks that induce planters to work more and accumulate more *fatigue* (shocks in the *fatigue* equation) also tend to reduce workers' *rest*.

Unobserved individual characteristics η_i are clear determinants of workers' productivity. Regardless of the specification used the variance of these individual factors in the productivity equation $\text{Var}(\eta^p)$ is significant at 1%. For the RE and the IV-RE specifications in column (a) and (b) the estimate is around 0.07. When the model is simultaneously estimated in column (c) this variance is 0.11. Since we use difference of means to calculate FE, the individual effects cancel out and we do not have a direct estimate of this parameter. Unobserved individual characteristics also affect *fatigue* and *rest* equations. The variance in the *fatigue* equation $\text{Var}(\eta^f)$ is estimated to be 0.02, significant at 1% level. In the *rest* equation $\text{Var}(\eta^r)$ is 0.20, also significant at 1%. The strong significance of these two variances suggest that a simultaneous estimation of our system of equation is a more adequate approach to analyze workers' productivity. The two-stage approach (IV-RE) ignores these unobserved individual effects for workers *fatigue* and *rest*.

Unobserved individual factors are not only present in our model but also correlated across the equations. The estimates of $\text{Cov}(\eta^p, \eta^f)$ and $\text{Cov}(\eta^p, \eta^r)$ are respectively 0.05 and 0.07, both significant at 1%. Finally, $\text{Corr}(\eta^f, \eta^r)$ is 0.03, also significant at 1% level. This means that planters who accumulate more *fatigue* by working more days in a row also tend to take longer resting periods. In general, these correlations could be interpreted as evidence that the regular workweek and the standard weekend with two days of rest do not necessarily constitute an optimal schedule for the most productive workers. Individual preferences for work schedules seem to be heterogenous and alternative work schedules may be more appropriate to satisfy individual preferences and even increase worker productivity.

1.6 Model Predictions

We have provided estimates of the response of worker productivity to exogenous changes in *fatigue* and *rest*. Yet, it may be of interest to use these estimates to predict the potential gain in productivity for alternative work schedules.

A typical tree-planting contract lasts around fourteen days (two weeks). The regular schedule consists of two cycles of five days of consecutive work and two days of rest at the end of each cycle. However, there exist 1001 alternative schedules that can accommodate the same ten days of work and four days of rest in different ways. We use our model estimates to predict mean individual earnings for each one of the 1001 alternatives.[2] We find that 60% (596) of the alternative schedules lead to higher productivity (earnings) with respect to the regular workweek.

Table 1.4 – Predicted earnings in CAD under alternative work schedules

	regular workweek	alternative work schedules					
	(a)	(b)	(c)	(d)	(e)	(f)	(g)
Week 1							
Monday	213	213	213	213	213	213	213
Tuesday	207	207	207	207	207	207	207
Wesnesday	191		191		191	191	191
Thursday	195	242	195	242			
Friday	171	212	171	212	232		
Saturday				168	184	213	213
Sunday		201			168	195	195
Week 2							
Monday	213	184		201		178	
Tuesday	207			195	214		214
Wesnesday	191	216	257		197	216	197
Thursday	195	221	263	242		221	202
Friday	171		230	212	232		
Saturday		201	182	168	184	201	201
Sunday		184	166			184	184
total earnings	1954	2081	2075	2061	2021	2018	2016
mean earning differences in CAD with respect to the regular workweek		128	121	107	67	64	62
standard deviation		71.7	40.4	69.3	66.6	51.3	53.0
t-test statistic		56.2	94.9	48.8	32.0	39.5	37.2

Table 1.4 shows predicted earning of a two-week contract under some of the most compelling work schedules that consist of ten days of work and four days of rest. Column (a) shows average earnings under the regular workweek, this is, five days of consecutive work followed by two days of rest. Total earnings in this case are $1,954$ CAD. Columns (b) through (g) show the average individual earnings of the most productive alternative schedules. The table includes: average total earnings, difference in CAD with respect to the regular workweek, standard deviation (from individual variation), and a t-test statistic indicating whether a difference is statistically zero.

[2] $\widehat{\mathbf{E}}(earnings_{it}|fatigue_{it},\, rest_{it},\, \eta_i^p) = \displaystyle\int \exp\left(\hat{\alpha}\, fatigue_{it} + \hat{\beta}\, rest_{it} + \eta_i^p\, \frac{\hat{\sigma}_p^2}{2}\right) f(\eta_i^p)\, \mathrm{d}\eta_i^p$, where $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\sigma}^p$ are the ML estimates of our model.

According to the predictions in Column (b), taking one day off every two days of work increases earnings by 6.55% with respect to the regular workweek. Column (c) shows that gathering the 4 days of rest in the middle of the two weeks of work increases planters' earnings by 6.19%. Similarly, the following columns show alternative work schedules and their potential gain in earnings with respect to the regular workweek. All these differences are statistically significant at 1% significance level. In summary, our analysis predicts that merely changing the order of the workdays leads to significant improvements on individual earnings.

1.7 Conclusions

This paper measures how worker productivity is related to accumulated *fatigue* and *rest* using payroll records of a tree-planting firm. For the estimation, we propose an instrumental variable approach using national public holidays and the relocation of workers to planting sites as valid instruments. We find that *fatigue* and *rest* are important determinants of workers productivity and should be taken into account to correctly analyze real-world data. Each additional day of work reduces productivity by 9.8% or 9.1% depending on the specification, while an additional day of rest significantly increases productivity by 4.2% to 5.8%. Furthermore, we find that the tree-planting firm could increase productivity by up to 6.5% in a two week contract simply by rearranging workdays and the days of rest. These are inexpensive changes that do not increase the payroll cost and could be considered by firms that require workers to perform physically demanding task similar to tree-planting.

Our results can also be of interest for researchers conducting field experiments and labour economists interested in worker productivity. Issues concerning workers *fatigue* and *rest* into their analysis could harm the external validity of the results and create potential biases. As suggested by Levitt and List (2011), workers' fatigue and rest could interact with an experimental treatment and bias the estimates. As part of a future research, it could be interesting to measure the importance of the bias created by *fatigue* and *rest* in a controlled experiment. Furthermore, temperature related optimal working schedules might also be of interest.

1.8 Bibliography

Biddle, J. E. and Hamermesh, D. S. (1990). Sleep and the Allocation of Time. *Journal of Political Economy*, 98(5):pp. 922–943.

Greene, W. (1997). *Econometric Analysis*. Prentice-Hall international editions. Prentice Hall.

Hamermesh, D. S. (1990). Shirking or Productive Schmoozing: Wages and the Allocation of Time at Work. *Industrial and Labor Relations Review*, 43(3):pp. 121S–133S.

Levitt, S. D. and List, J. A. (2011). Was There Really a Hawthorne Effect at the Hawthorne Plant? An Analysis of the Original Illumination Experiments. *American Economic Journal: Applied Economics*, 3(1):224–38.

Rybczyński, W. (1991). *Waiting for the Weekend*. Viking Penguin.

Saez, M. M. (2011). Siesta: A theory of freelancing. ECON - Working Papers econwp055, Department of Economics - University of Zurich.

Train, K. (2003). *Discrete Choice Methods With Simulation*. Discrete Choice Methods with Simulation. Cambridge University Press.

Appendix

1.A Estimation methods

We maximize the likelihood of the system of simultaneous equations described in (1.3) using simulated probabilities to approximate individual RE. Let $\mathbf{y}_{it} = [y^p_{it},\ y^f_{it},\ y^r_{it}]'$ be the vector of observed endogenous variables and $\mathbf{x}_{it} = [X_{it},\ Z_{it}]'$ the matrix of regressors and instruments from our model. For a given triplet of time invariant factors η_i, the individual contribution to the likelihood is

$$\Pr(\mathbf{y}_i |\, \mathbf{x}_i, \eta_i) = \prod_{t=1}^{T_i} \Pr(\mathbf{y}_{it} |\, \mathbf{x}_{it}, \eta_i).$$

The vector η_i is unobserved, but its distribution function is known. We can obtain the individual unconditional probability by integrating η_i over its three dimensions:

$$\Pr(\mathbf{y}_i |\, \mathbf{x}_i) = \int \prod_{t=1}^{T_i} \Pr(\mathbf{y}_{it} |\, \mathbf{x}_{it}, \eta_i).\phi(\eta_i)\, \mathrm{d}\eta_i, \tag{1.6}$$

where $\phi(\eta_i)$ is the multivariate normal distribution described in equation (1.4). The likelihood is a function of the observations $(\mathbf{y}_{it}, \mathbf{x}_{it})$ and a vector θ, containing the parameters of the model:

$$\mathrm{L}(\mathbf{y}_{it}, \mathbf{x}_{it}; \theta) = \prod_{i=1}^{N} \int \prod_{t=1}^{T_i} \Pr(\mathbf{y}_{it} |\, \mathbf{x}_{it}, \eta_i).\phi(\eta_i)\, \mathrm{d}\eta_i.$$

In logarithms,

$$\mathscr{L}(\mathbf{y}_{it}, \mathbf{x}_{it}; \theta) = \frac{1}{N} \sum_{i=1}^{N} \log \left(\int \prod_{t=1}^{T_i} \Pr(\mathbf{y}_{it} |\, \mathbf{x}_{it}, \eta_i) \phi(\eta_i)\, \mathrm{d}\eta_i \right).$$

For computational purposes, it is convenient to rewrite this three dimensional probability in terms of conditional probabilities

$$\Pr(\mathbf{y}_{it} |\, \mathbf{x}_{it}) = \Pr(y^f_{it}, y^r_{it} |\, y^p_{it};\ \mathbf{x}_{it}) \Pr(y^p_{it} | X_{it}),$$

where $\Pr(y^f_{it}, y^r_{it} |\, y^p_{it};\ \mathbf{x}_{it})$ is a bivariate normal density and $\Pr(y^p_{it} | X^p_{it})$ is a continuous univariate normal.

The maximization procedure consists in finding $\hat{\theta}$, a vector of estimated parameters that maximizes

$$\mathrm{L}(\mathbf{y}_{it}, \mathbf{x}_{it}; \theta) = \prod_{i=1}^{N} \int \prod_{t=1}^{T_i} \Pr(y^f_{it},\ y^r_{it} |\ y^p_{it};\ \mathbf{x}_{it},\ \eta_i) \cdot \Pr(y^p_{it} |\ X_{it},\ \eta^p_i) \phi(\eta_i)\, \mathrm{d}\eta_i.$$

The asymptotic covariance of the maximum likelihood estimates $\hat{\theta}$ is the negative inverse of the Hessian matrix (see Greene, 1997). We use the BHHH method to obtain its consistent estimate. This method uses first derivatives instead of analytic second derivatives, which are non linear. More precisely, the Hessian is approximated by the outer product of the score matrix

$$\widehat{\mathrm{Var}}(\hat{\theta}) = -\,\hat{\mathbf{E}}\left[\frac{\partial^2 \mathscr{L}(\hat{\theta})}{\partial\theta\,\partial\theta'}\right]^{-1} = -(S'S)^{-1}, \qquad\qquad S = \frac{\partial \mathscr{L}(\hat{\theta})}{\partial\theta}.$$

We approximate the integral in the objective function using simulation techniques (Train, 2003). This consists in drawing[3] C triplets $\eta_{i,c}$ from the multivariate normal distribution in equation (1.4) for each individual. Finally, we calculate the conditional probability of each $\eta_{i,c}$ and approximate the individual unconditional probability in (1.6) by

$$\check{\Pr}(\mathbf{y}_i|\,\mathbf{x}_i) = \frac{1}{C}\sum_{c=1}^{C}\prod_{t=1}^{T_i}\Pr(y^f_{it},\, y^r_i|\, y^p_{it};\, \mathbf{x}_{it},\, \eta_{i,c}).\,\Pr(y^p_{it}|\, X^p_{it},\, \eta^p_{i,c}). \tag{1.7}$$

[3] We use Halton draws.

Chapter 2

The Role of Productivity Shocks in the Effort Choice of Agents: Using experimental data to test for additivity in the production function

Charles Bellemare   María Adelaida Lopera   Bruce Shearer

2.1 Introduction

The *principal-agent* problem often refers to a situation in which a worker (the agent) chooses an effort level that optimally balances remuneration against an increasing cost of effort. In this context the employer (the principal) can be confronted with a problem of moral hazard because observed outcomes do not necessarily reflect effort on the part of the agent. This happens for example when performance depends on random productivity shocks unobserved by the firm. These shocks consist of unexpected unpredictable factors independent of the will of the agents that affect their outcome.

Productivity shocks are omnipresent: workers get sick, farmers experience pests, corporate managers suffer from unforeseeable economic fluctuations, etc. Surprisingly, little is known about the true impact of these shocks on the agent's choice of effort. Applied researchers have no guidance for modelling agent's production function. Most empirical studies assume a multiplicative relationship between productivity shocks and agents' effort (Shearer, 2004; Allen and Lueck, 1992; Dubois and Vukina, 2009; Bellemare and Shearer, 2013). This choice is presumably made because multiplicative production function makes the estimation of agency models a great deal simpler.

Productivity shocks are not atheoretic error terms that can be added to the productivity model *ex post* to suit the estimation requirements. Rather, their potential interaction with productivity incentives has behavioural implications. The form of the relationship between effort and productivity shocks determines the agent's optimal effort and ultimately, how agents may react to changes of their economic incentives. If the production function has a multiplicative structure, realizations of the productivity shock directly affect the agent's choice of effort. If instead the production function is additive, the agent's optimal effort will be exclusively determined by incentives and independent of shocks.

In this essay we propose a simple and innovative approach to select the most appropriate structure to model the relationship between agent's effort and productivity shocks in the production function. Generally speaking, specification errors may lead to an estimation bias. Failure to impose the correct assumptions about the production function may undermine the validity of the empirical analysis. Choosing a model that represents well the production function is a challenging task because neither effort nor productivity shocks are observed.

In our empirical analysis we use payroll records from a tree-planting company in British Columbia, Canada. We observe the payment received by each worker, as well as their daily productivity during the planting season of 2013. This data has two main advantages that facilitate our research. First, firm level analysis removes a lot of the potential unobserved heterogeneity among workers. Based on the evidence presented by Bellemare and Shearer (2010) about the tree-planting industry, we do not expect large variation on individual preferences between tree planters. This homogeneity comes from the fact that workers may self-select into this job according to characteristics such as risk preferences and physical effort cost. As opposed to Dubois and Vukina (2009), there is less need for us to model an environment with heterogeneous agents.

A second advantage of our case study is that agents observe productivity shocks before choosing effort. Considering *ex ante* productivity shocks implies that there is no uncertainty from the agents' perspective and we can abstract our analysis from their risk preferences. Tree planters observe most productivity determinants before choosing their effort. At the beginning of the day workers observe their individual health status, the weather, the steepness of of their own planting block, whether the soil is covered with underbrush, etc. Once workers know the planting conditions, we assume that they have observed the productivity shock. This might be particularly true for experienced tree planters who can very well anticipate their daily production after observing a set of productivity socks. The study of the productivity structure in contexts where the shock is unobserved is certainly an interesting track for future work. While introducing risk aversion in the utility function is possible in theory, empirical identification of the cost of effort from the risk aversion preferences may require additional assumptions.

Of particular concern when investigating the relationship between effort and productivity shocks is the endogeneity of effort incentives such as wages, bonuses, and piece rates. In our tree-planting firm piece rates and workers productivity hold a bidirectional relationship. The piece rate is chosen by the firm to incentivize planters to work hard, but at the same time it depends on external working conditions that determine planters productivity. Regular piece rates depend on factors unobserved to the econometrician that enter in the production function as random productivity shocks. The use of observational data to compare worker performance under "naturally changing" piece rates will fail to identify the effect of these variations on effort.

Our strategy to solve this endogeneity problem is to use a field experiment that permits the piece rate to vary exogenously, allowing direct measurement of the incentives' effect and eliminating the need for instruments. In our field experiment, 21 planters worked under three different piece rates.[1] In total, the experiment provided 270 observations on daily productivity and incentives.

Once we have properly identified the incentive effects with the experimental data, we can use a simple semi-parametric test to investigate the role of productivity shocks on planters'

[1] In total, 20 out of the 21 workers involved in the experiment were observed planting under all treatments.

effort choice. We describe how testing for additivity of the production function comes down to testing for scale effects in a quantile regression framework. Intuitively, if workers effort and the random productivity shocks hold a separable relationship, exogenous changes in effort induced by our experiment should not be able to predict output variance or any other distributional feature. In other words, the absence of scale effects is consistent with additivity in the production function, while the presence of scale effects points towards a more complex relationship between effort and the shocks, for instance a multiplicative function.

In our empirical analysis we reject the hypothesis that the production function of tree-planters is additive. The evidence suggests that planters' production can be better model by a multiplicative function. A direct consequence of our finding is that planters' optimal effort depends not only on economic incentives but also on working conditions as well as other productivity shocks. This result simplifies the estimation of agency models but at the same time leaves the door open for complex interactions between effort choice and potential uncertainty about the shocks. This question about the incidence of individuals risk preferences on incentive response is an interesting path for future research.

This paper is structured as follows. Section 2.2 provides institutional details of the tree-planting industry. It describes planters compensation scheme and the nature of the productivity shocks. Section 2.3 describes our field experiment on incentives change and its results. Section 2.4 presents the structural model used to analyze the data and describes how we test for separability of the production function. Section 2.5 summarizes our results and Section 2.6 concludes.

2.2 Tree-planting

The province of British Columbia is the largest timber producer in North America. Our data comes from a medium-sized tree-planting firm actively participating in this competitive market. For each contract, the firm divides the planting areas into *blocks* to separate different types of terrain. On each block, a price per tree planted is assigned depending on the soil conditions.

The *piece rate* paid to tree planters is endogenous because it depends on the block's characteristics and the expected number of trees that a regular worker can plant. For instance, since steep and rocky terrain slows workers and make planting more difficult, the firm sets a higher piece rate in these conditions to induce planters to put more effort into their jobs. The firm subdivides blocks into *plots* and allocates each planter for a day of work. Planters are naturally exposed to random productivity shocks within a given block. Even though all workers receive the same price per tree planted within a block there are random variations of planting conditions that are beyond the firms' control because it is not possible to completely know the undersoil conditions. A block may appear uniform on the surface, but some portions can have a rocky soil underneath which slows planting. As a result some planters end up working in more difficult conditions under the same piece rate. In this sense planters can be said to be exposed to random productivity shock.

Workers' earnings are jointly determined by a daily piece rate and their individual productivity measured by the number of trees planted. At the start of a normal field day, the manager assigns each planter to a single plot with no systematic matching of workers to planting conditions. Each worker receives a box full with seedlings and a shovel. A truck transports planters to their plots, where they spend the day. The task itself consists in digging a hole, pacing a seedling and covering its roots with soil. The manager evaluates the planted area afterwards to ensure quality. Poorly planted trees must be replanted at the planter's expense. However, incidents involving poor quality are rare.

2.3 Experimental Design and Data

We combine payroll data from our tree-planting firm with experimental outcomes from a field experiment, which exogenously incentivize planters to work hard by changing their piece rate. Our data set spans the period between April 28^{th} to May 22^{nd}, 2013. In our analysis we use an unbalanced panel of 270 observations from 21 tree planters in their natural work environment. We observe daily individual outcomes and the piece rate paid by the firm in each planting block.

The field experiment consists of two exogenous changes in planters' piece rate while holding all other conditions constant. These two changes incentivize workers' effort and consequently their productivity. For this experiment, a large planting block with homogenous soil conditions was fictitiously divided into three blocks with different piece rates, which correspond to different experimental treatments. The first block corresponds to a *baseline* treatment in which workers received their regular compensation of $0.14 CAD per tree planted. In the second block there was a *small increase* of 3¢ in the piece rate paid to planters ($0.17). Finally, in the third block there was a relatively *large increase* of 5¢ per three planted ($0.19). In order to avoid potential *Hawthorne effects*, the different piece rates were presented to planters in a context of normal daily operations, as if they were associated to different soil conditions. We have chosen to restrict our sample to the observations a few days before the treatments and exclude observations far away in time. Using a short counterfactual avoids strong seasonal weather variations that may affect workers productivity. We use only pre-treatment data to exclude potential biases created by persistent effects of the treatments.

Table 2.1 – Summary of the field experiment

	(a) all obs.	(b) baseline	(c) small increase	(d) large increase
piece rate in CAN dollars		0.14	0.17	0.19
no. of observations	270	149	61	60
no. of observed days	15	9	3	3
higher daily temperature in °C	18	21	14	14
lower daily temperature in °C	3	3	3	5
daily precipitations in mm.	1	0	2	1
productivity: number of trees planted				
average	1,528	1,424	1,502	1,813
standard deviation†	557	548	597	676
minimum	210	210	500	815
percentile 25^{th}	1,070	920	1,050	1,305
percentile 50^{th}	1,420	1,300	1,370	1,625
percentile 75^{th}	1,960	1,850	1,750	2,135
maximum	3,650	3,170	3,200	3,650

†Clustered at individual level.

Table 2.1 summarizes the observed data from our field experiment. It includes relevant information about the experiment itself, weather conditions,[2] and workers productivity measured by the number of trees planted. Column (a) shows statistics over all observations, while the other three columns summarize the information for each experimental treatment. Column (b) presents baseline data, when tree planters work are paid the regular piece rate of 0.14 CAD.

The first striking fact is the importance of productivity variation and its resulting income variation. The average number of trees planted in the baseline treatment is $1,424$, with a standard deviation of 548. In terms of income this represents average daily earnings of around 200 CAD with a standard deviation of 77 CAD.

The last two columns summarize the two experimental treatments, which incentivize workers' effort and their productivity by increasing their piece rate. Column (c) shows that when there is a small increase in the piece rate the average productivity increases by 78 trees (5.5%) with respect to the baseline treatment. This change is not statistically significant ($p-$value of a $t-test$ is 0.27), probably due to the large sample variance. As in all field experiments, the actual value of the incentives restricts the empirical analysis. The smaller the incentive, the larger the sample size required for measuring its effect with precision. Another possible explanation for this lack of statistical significance is that weather conditions offset the treatment effect. Rain may have a confounding effect during this first treatment because planters experienced 2mm of rain with respect to 0mm in the baseline. The treatment effect in column (d) is more salient. When there is a large increase in the piece rate the average productivity increases by 398 trees (27%). This change is statistically significant at 1% level. The effect of this last treatment should be sufficient to exogenously vary effort, enabling us to conduct our econometric analysis.

In addition to the effect of the treatment on the average productivity the increase in the piece rate could also affect other moments of the productivity distribution. In particular, we observe that the standard deviation increases from 548 trees in the baseline to 597 in the first treatment and 676 in the last treatment. This corresponds to a 9% and 23% increase respectively. A simple descriptive analysis of the data is not sufficient to assess whether these variations correspond to a treatment effect, or if they are the result of changes in other environmental factors such as weather conditions. A regression analysis would be more appropriate to explore the treatment effects on the productivity variance while controlling for climate factors.

The bottom of the Table 2.1 sketches the distribution of planters productivity conditional on the experimental treatments. Increasing the piece rate compresses the lower tail of the productivity distribution. The small increase in the piece rate shifts the minimum productivity upwards by 138% and the large increase by 305%. Similarly, the 25^{th} percentile moves upwards by 14% when there is a small piece rate increase, and there is a rise in productivity of 42% when there is a large increase in the piece rate. The treatment effect is not clear at the higher end of the distribution. Once again, it is necessary to control for other determinants such as weather conditions and unobserved individual effects to have a more accurate measure of the treatment impact on the conditional productivity distribution. This task requires more sophisticated tools such as conditional quantile analysis.

[2]Unfortunately, we do not have data on local weather conditions. We use data from the nearest weather station at Williams Lake (British Columbia, Canada) as a proxy for the true climate conditions.

2.4 Model

In this section we first develop a structural model of worker productivity under piece rate contract.[3] We describe the timing of our model and its empirical implications. In the second part of this section we discuss the structural form of the production function and why a test for homoskedasticity and a test for scale effects are suitable for choosing a good production model.

The timing of our model is as follows. For each plot of land to be planted:

1. The firm observes the distribution $f(\mu, \sigma)$ of a productivity shock S and selects a piece rate r.
2. Each worker observes a particular realization s of the random variable S, chooses the optimal effort level e that maximizes their utility and produces the equilibrium outcome y.
3. The firm pays ry to the worker.

The timing of our model makes explicit the endogeneity of the compensation system. In general, firms choose wages and piece rates based on the distribution of the productivity shock; therefore, variation observed in microeconomic data is endogenous. Higher piece rates are often correlated with tougher working conditions and depend on expected workers productivity. As explained by Paarsch and Shearer (1999), regression methods that directly use the observed covariation between workers' productivity and their payment will fail to provide a consistent estimate of the production function. Our identification strategy consists in using the three experimental treatments described in Section 2.3 to induce exogenous variation of planters' piece rate.

In our model, firms know the distribution of the shock, whereas planters observe actual shock realizations. This assumption about the timing of the shock is crucial and it is mainly determined by the nature of the activity that we study. In the context of tree planting it seems reasonable to assume that workers observe a draw s from the distribution of S before selecting their effort level. First of all, working conditions can be treated as random draws because the assignment of workers to planting plots does not follow any systematic pattern. Second, s is observed because planters get to know the soil conditions at the very beginning of the day. Upon arrival to their planting plot, workers observe the steepness of the ground, whether the soil is covered with underbrush, etc. From the first plow into the ground planters can assess particular characteristics such as soil hardness and stiffness. There are situations in which the productivity shock is potentially unknown. For instance Dubois and Vukina (2009) use data from workers in a swine farm and argue that productivity shocks are unobserved in their context. In their case workers cannot foresee their own productivity because there is uncertainty regarding the feed conversion ratio that affect animals' gain weight. An advantage of our case study is that tree planters can choose their effort level after assessing the planting conditions, and thus we can abstract from uncertainty and workers preferences regarding these conditions. We use planting conditions as an illustrative example in our analysis, but any other random productivity shock (sickness, family problems, etc.) can be considered as long as they are observed *ex-ante*.

[3] Our analysis focuses on the choice of the agent and not the choice of the principal.

We assume that workers are rational and maximize a utility function defined over monetary incentives r and effort E

$$U(r, E) = rY - C(E). \tag{2.1}$$

The random variable Y is the technology of production and $C(E)$ is a convex cost function of effort. For exposition purposes we define this cost function as

$$C(E) = \frac{E^2}{2}. \tag{2.2}$$

Abstracting from uncertainty simplifies our analysis but does not impose additional constraints. Since workers can observe productivity shocks before they choose their effort level, risk preferences should not play any role in the process of choosing effort. If the value of the shock was unknown, the question of whether the utility function is risk neutral or not would become relevant. When there is uncertainty involved in the effort choice, workers maximize expected utility and the choice of effort could depend on their risk preferences. The advantage of focusing on risk neutral agents is that they don't need to be compensated for the loss of utility due to increased risk exposure. Risk neutral workers simply maximize their individual utility using the expected value of the shock instead of its actual realization. If the shocks were unobserved, risk neutrality would allow us to focus on the form of productivity while avoiding risk considerations. For other type of preferences towards risk, uncertainty may modify agents' cost of effort.

In our firm workers are paid in proportion to the number of trees planted per day (*i.e.* piece rates) rather than fixed wages. A discussion about the choice of the compensation system in the tree-planting industry has been well addressed by Shearer (2004). This issue is certainly interesting, but out of the scope of this essay.

2.4.1 Production Function

In general, planters productivity Y can be modelled as a function of effort and a productivity shock

$$Y = g(E, S). \tag{2.3}$$

We are not aware of any empirical evidence about the structural form of this function, even for particular industries. Most of the literature concerned with structural estimation of the principal-agent relationship assumes a multiplicative form, often without further explanation. An exception is Dubois and Vukina (2009) who make an effort to justify their choice of a multiplicative production function. They argue that the effort of agents working in a swine farm interacts with the distribution of the productivity shocks. In their example, an agent inspecting the swines too often (exerting more effort) may change the probability of infectious diseases and change the outcome distribution. The practical usefulness of this type of assumption in the context of mean regression relies on its convenience to perform non-linear transformations such as the log function. Logarithmic transformations make distributions appear more normal, achieve better model fit, and facilitates the estimation and the interpretation of coefficients as percentage changes.

A priori, the technology of production could also be separable with respect to the productivity shock. For instance, it could have an additive structure. These type of function is mainly used for theoretical analysis of the agency model (*e.g.* Prendergast, 2002). In a context of regression to the mean an additive structure implies that we can only estimate linear transformations of the production function.

The optimal effort derived from the utility function (2.1) depends on the piece rate r, the observed productivity shock s, and the structural form of the productivity function (2.3). For instance, if productivity had an additive structure of the form $Y = E + s$, the optimal level of effort[4] chosen by a worker would be the inverse of the marginal effort cost $e = m(r)$. Notice that this function does not depend on the productivity shock s. Replacing this optimal effort choice back into the production function we obtain a simple regression model of the production output

$$y = m(r) + s. \tag{2.4}$$

Thanks to the separability of the production function a change in the incentives r has a deterministic impact on the output and is independent of the productivity shocks.

If the production function had a non separable structure of the form $Y = Es$, the optimal level of effort would be a direct function of the productivity shock $e = m(rs)$. Notice that the shock interacts with the piece rate amplifying or decreasing its incentive effect on effort.

Making the right assumptions about the functional form of production is essential for a valid and coherent interpretation of our results. A priori, there is no reason to privilege one particular form over the other. Instead of choosing a multiplicative or an additive form *ad hoc*, we intend to use the empirical evidence to select the most appropriate structure.

2.4.2 Test

There are two potential mistakes that we can make when choosing between the multiplicative and the additive regression model. One is to use a multiplicative form when the true production function is additive. Unfortunately, the testable consequences of this error are rather difficult to derive. The second potential mistake is to assume an additive structure when the true production function is multiplicative. It is to test this second case that we aim.

Suppose for a moment that we have mistakenly chosen an additive structure $Y = E + S$. Under our particular parameterization of the cost function,[5] the maximization process of the true production function will lead to the optimal effort $e = rs$. Replacing this effort in the wrong production function we obtain the empirical model

$$y = rs + s. \tag{2.5}$$

The first term of this model is a stochastic component that leads to an heteroskedastic y. Changes in r affect the distributional shape of the dependent variable. In particular it affects its variance: $\mathrm{Var}(y|r) = \mathrm{Var}(s)(1 + r)^2$. In this equation we can clearly see that large values of the piece rate r amplify the impact of the productivity shocks s expanding the dispersion of the conditional productivity.

If the additive production function was correct, the regression model (2.4) would be homoskedastic because its conditional variance $\mathrm{Var}(y|r) = \mathrm{Var}(s)$ would not depend on r. In

[4]This result is easily derived by replacing the productivity function $Y = E + s$ into equation (2.1) and deriving the first order condition $r = C'(e)$, where C' is the marginal cost of effort and m its inverse.

[5]Our reasoning holds for any convex function of effort cost C. We use here the particular form in equation (2.2) to simplify the presentation.

principle, we can test the validity of the additive structure by using standard tests for homoskedasticity. However, the piece rate may affect in general the scale of the conditional distribution of productivity in a more general fashion, not only its variance. Quantile regression permits a broader investigation of scale effects.

In summary, additivity of the productivity shock is consistent with homoskedasticity in a linear regression model, while heteroskedasticiy and other distributional effects are consistent with a multiplicative structure.

The econometric literature has developed a variety of tests for heteroskedasticity that differ in flexibility and complexity. We begin our discussion by focusing on a linear scale model and the standard Wald test designed to test for homoskedasticity. Then, we introduce the quantile regression framework, which offers a more flexible environment to test for other distributional effects.

Mean Regression

In general, heteroskedasticity of a random variable s can be described as

$$\begin{aligned} s &= \sigma\varepsilon, \qquad \text{where} \\ \sigma^2 &= \mathrm{Var}(s|\mathbf{z}) = g(\mu + \delta\mathbf{z}) \end{aligned} \tag{2.6}$$

is the skedastic function, and ε is an *iid* random variable. The vector $\mathbf{z}$ contains potential sources of heteroskedaticity, and g is a a positive and monotonic function. Most formal tests for heteroskedasticity compare the null hypothesis $Ho : \delta = \mathbf{0}$ against the alternative hypothesis $H_a : \delta \neq \mathbf{0}$. A natural approach is to estimate model (2.6) and evaluate whether the estimates satisfy the null.

The Wald test for instance, rejects the null hypothesis if the estimates $\hat{\beta} = [\hat{\mu}, \hat{\delta}]'$ of the heteroskedasticity model (2.6) are statistically different form the estimates of a restricted regression $\hat{\beta}^R = [\hat{\mu}, \mathbf{0}]'$. The formal statistic of this test is

$$W = \left(\hat{\beta} - \hat{\beta}^R\right)' \hat{\Lambda}^{-1} \left(\hat{\beta} - \hat{\beta}^R\right), \tag{2.7}$$

where $\hat{\Lambda}$ is a weighting matrix, usually the difference between estimated variances of the unrestricted and restricted models. Under the null hypothesis this statistic distributes $\chi^2_{(q)}$, with q equal to the number of restrictions in $\hat{\beta}^R$.

Testing for homoskedasticity in a mean regression model tells us whether the variance of the regression errors is constant. However, covariates may influence the conditional distribution of the dependent variable in many other ways. Conditional quantile regression nests within the *iid* location shift model of classical linear regression and can provide a richer overview of the distributional effects. Quantile regression can capture the effect of covariates on the dispersion of the dependent variable (heteroskedasticity) as well as other scale effects.

Quantile Regression

Quantile regression is a semi-parametric technique first proposed by Koenker and Bassett (1978). It extends the notion of ordinary quantiles to a more general class in which the conditional quantiles have a linear form. Quantile regression provides a rich characterization of the conditional distribution of the endogenous variable, is robust to outliers, consistent, and more efficient under weaker assumptions about the shape of the error distribution.

Let $\theta \in (0, 1]$ represent a given quantile. The $\theta-$quantile of workers productivity is

$$\mathbf{Q}_\theta(y|r) = \beta_0(\theta) + \beta(\theta)r + \mathbf{Q}_\theta(s|r), \tag{2.8}$$

where $\mathbf{Q}_\theta(y|r)$ denotes the $\theta-$quantile of y conditional on an exogenous piece rate r. The distribution of the error term in this regression is left unspecified. The only requirement is that $\mathbf{Q}_\theta(s|r) = 0$ for at least one quantile θ.

In our application r is a vector of dummy variables for the two experimental treatments changing the piece rate. These variables are exogenous and uncorrelated with the productivity shock by construction, because in our field experiment the piece rate does not vary with the distribution of planting conditions and soil quality. Technically, this means that r is independent of $\mathbf{Q}_\theta(s|r)$.

The constant $\beta_0(\theta)$ denotes the $\theta-$quantile of productivity in the reference group. Our reference group is composed of observations under the baseline treatment. Naturally, the value of this constant becomes larger as the percentile θ increases because the 25^{th} percentile ($\theta = 0.25$) of productivity in the baseline group is different form the median ($\theta = 0.50$) and from the 75^{th} percentile ($\theta = 0.75$).

The vector of parameters $\beta(\theta)$ measures the effect of experimental (exogenous) variations of the piece rate r for a given quantile θ. These parameters capture two types of treatment effects: a *location effect*, which does not depend on the selected quantile; and a *scale effect*, which depends on the quantile.

An increase in the piece rate has a location effect when it shifts the conditional distribution of productivity. Location effects are uniform over the whole range of the distribution and can be captured by mean regression models as well as by quantile regression models.

Our interest in the quantile regression resides in the possibility of capturing scale effects. These are effect that regressors may exert on the dispersion of planters productivity and its conditional distribution in general. In the absence of scale effects, we should measure the same treatment effect regardless of the quantile we choose. For example, the difference $\mathbf{Q}_\theta(y|r = \text{T0}) - \mathbf{Q}_\theta(y|r = \text{T2})$ should be the same across θ. This restriction does not generally hold in presence of scale effects. Recall that the presence of heteroskedasticity and other distributional effects can inform us about the true form of the production function, which we have modelled as additive. If a multiplicative structure is more appropriate, we should detect scale effects. If the production function is additive, we should not observe scale effects.

Different estimates of β at distinct quantiles indicate the presence of scale effects. They reflect differences in the response of the dependent variable to changes in the regressors at various points in the conditional distribution of the dependent variable. The standard test for scale effects evaluates the differences between the slope coefficients of simultaneous quantile regressions. Systematic differences in the slope coefficient are consistent with scale effects and thus, with a multiplicative error structure in the production function. If instead the conditional distribution is independent of the regressors, all quantiles should have parameter vectors that differ only in their intercept. Similar to the approach used for the Wald test, we compare the parameters of an unrestricted model that allows scale effects and location effects, with a restricted model that only allows for location effects. The null hypothesis assumes that imposing scale effects to be zero is not a restrictive constraint.

The unrestricted vector $\hat{\beta}$ contains the estimates of the simultaneous quantile regression model (2.8), which allows for different slope coefficients in each quantile. The restricted parameters impose the slope coefficients to be equal over all quantiles in the same regression model. A first problem to implement this test is that the choice of these restricted parameters is not clear. Buchinsky (1998) suggests to use a minimum distance procedure to find a plausible value for $\hat{\beta}^R$. The vector of restricted coefficients is

$$\hat{\beta}^R = \arg\min_{\beta^R} \left(\hat{\beta} - \beta^R\right)' \hat{\Lambda}^{-1} \left(\hat{\beta} - \beta^R\right),$$

where $\widehat{\Lambda}$ is the estimated covariance matrix of the simultaneous quantile regression. Once we solve this optimization problem, we can simply compute the test statistic in (2.7) using $\hat{\beta}^R$ and $\widehat{\Lambda}$.

2.5 Estimation Results

In this section we estimate workers' productivity. We test the validity of an additive production function by computing parametric and semi-parametric tests for homoskedasticity. As discussed in Subsection 2.4.2, homoskedasticity is consistent with an additive production function, while heteroskedasticity and other distributional effects are consistent with a multiplicative form. The empirical evidence rejects the additive structure, which suggests an interaction between the observed productivity shocks and workers' choice of effort.

Table 2.2 shows the estimates of an additive production function using mean regression and quantile regression analysis. Column (a) shows mean productivity of tree planters conditional on the experimental treatments and weather conditions. These are the estimates of equation (2.4) using Correlated Random Effects (CRE), which control for unobserved individual characteristics in a way that approaches fixed effects models (FE). In addition to the standard linear regressors the CRE include individual time averages to capture worker-specific effects that may create serial correlation within indiviuals. These estimates result from the minimization of a simple squared error loss function and have at least two clear advantages. First, unlike FE, CRE allow to estimate time invariant factors. Second, they are less demanding than random effects models (RE), because they do not require unobserved individual factors to be independent of regressors. We also found that CRE estimates are similar to other point regression estimates. Appendix 2.6 compares CRE to other mean regression models such as OLS, FE, RE, and to the least absolute regression model.

The CRE estimates indicate that the mean productivity in the control group under average working conditions (average temperature, zero rainfall) is $1,164$ trees per day. Rising the piece rate from 14¢ to 17¢ ($T1$) is estimated to increase average productivity by 13% (150 trees per day). However large, this estimate is less than one third of the productivity standard deviation and it is not statistically significant at 10% level. The only way to reduce our uncertainty about the true effect of this treatment would be to reduce the confidence interval by increasing the sample size. Fortunately, the estimated effect of a large increase in the piece rate is more precise. Rising the piece rate from 14¢ to 19¢ ($T2$) increases productivity by 25% (291 trees) and it is significant at 5% level. Overall, we are confident that these two experimental variations of the piece rate effectively incentivize effort.

Moderate climate conditions improve workers productivity. Tree planting is a physically demanding job and extreme temperatures tend to reduce productivity. The parameter associated to maximum daily temperature is negative, suggesting that planters become less productive as maximum temperature rises. Nonetheless, maximum daily temperatures are mild; they rarely approach 20°C, which may explain why the estimate is small (-13 trees) and not statistically significant at 10%. Minimum temperatures and rainfall are more relevant for productivity. An increase of one degree Celsius of the minimum daily temperature increases average productivity by 14 trees (significant at 10%). Each additional millimetre of rainfall reduces planters productivity by 47 trees per day (significant at 5%).

Table 2.2 – Additive workers production function

	CRE†	quantile regression		
		$\theta = 0.25$	$\theta = 0.5$	$\theta = 0.75$
	(a)	(b)	(c)	(d)
constant	1163.6*	423.4**	1163.6***	2023.9***
	(610.6)	(201.6)	(257.8)	(489.7)
small PR increase $T1$	150.2	270.8***	-43.0	-231.2
	(129.5)	(80.5)	(100.2)	(157.6)
large PR increase $T2$	291.3**	351.3***	-20.6	55.5
	(114.7)	(94.2)	(148.8)	(234.0)
daily highest temp.	-12.7	-11.6	-31.0	-32.0
	(11.2)	(19.2)	(26.8)	(43.2)
daily lowest temp.	13.6*	13.2**	7.3	20.5**
	(7.8)	(5.7)	(6.2)	(8.8)
precipitations.	-45.6**	-31.1	-34.6	-6.2
	(21.8)	(30.2)	(44.0)	(75.4)
correlated effect $\overline{T1}_i$	-1587.4	66.4	-340.0	-6583.0*
	(4657.2)	(929.2)	(1237.5)	(3505.0)
correlated effect $\overline{T2}_i$	2520.8	1778.8***	2160.0***	6056.9**
	(2425.9)	(493.2)	(682.4)	(2913.9)
no. of parameters	8	24		
no. of observations	270			
no. of workers	21			

*** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1.
† standard errors clustered by worker.

Time invariant individual effects present in panel data may create time correlation and interact with the treatment effects. Mundlak (1978) proposes to model this relationship by adding individual time averages to the regression, which allows to consistently estimate all regression parameters. We include this correlated effects as individual treatment averages over time $\overline{T1}_i$ and $\overline{T2}_i$. Their estimates are $-1,587$ and $2,521$ respectively, and are not statistically different from zero. Their lack of significance suggest that unobserved individual characteristics are not major determinants of the mean productivity. Given this relative homogeneity of tree-planters, a richer specification of the individual effects as in Chamberlain (1980, 1982) is not required.

Separability of the production function with respect to the shock, and additivity in particular, are consistent with homoskedasticity in the mean regression model. A simple informal diagnostic procedure is to plot the fitted regression residuals against the variable assumed to be in the skedasticity function. We are interested in effort incentives and their interaction with the productivity shocks in the production function. Unlike weather conditions or other exogenous factors, piece rates and wages can be easily modified by a firm and constitute a powerful tool for inducing productivity in principal-agent situations.

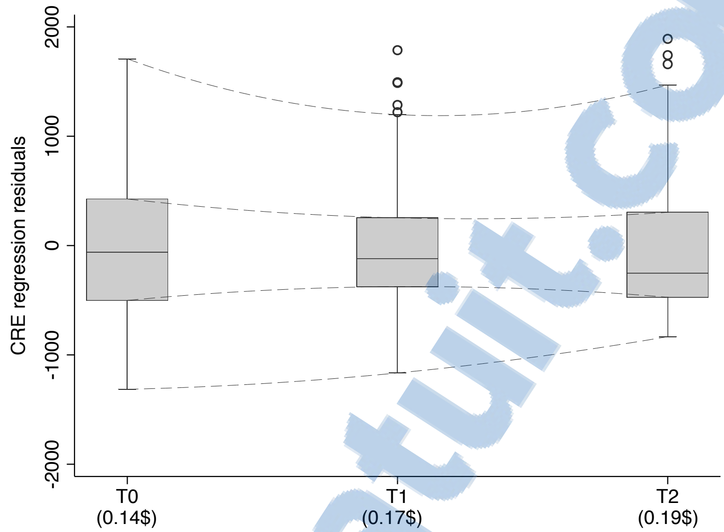

Figure 2.1 – Distribution of CRE regression residuals by treatment

Figure 2.1 plots the distribution of fitted residuals from the CRE model in Table 2.2 column (a), against the piece rate. The piece rate takes different values in each experimental treatment: 14¢ in the baseline treatment ($T0$), 17¢ in the first treatment ($T1$), and 19¢ in the second treatment ($T2$). In a homoskedastic context there should be no pattern to the mean regression residuals plotted against these or any other variables. It is difficult to determine from this graphic alone if the experimental changes in the piece rate are correlated with the error variance. We need to go beyond a visual representation to formally test for homoskedasticity. We defer the implementation of a formal test to the next Subsection.

The last tree columns of Table 2.2 present the quantile regression estimates of equation (2.8), controlling for weather conditions and unobserved individual effects. These estimates characterize the distribution of the conditional production function. Column (b) shows the 25$^{\text{th}}$ percentile, column (c) the median, and column (d) the 75$^{\text{th}}$ percentile. These quantile estimates result from minimizing an asymmetric absolute loss function. Since this type of objective function is not differentiable, standard gradient optimization methods are not applicable. Instead, we use a linear optimization algorithm proposed by Portnoy and Koenker (1997) to solve the minimization problem. Moreover, because the three quantile regressions are estimated using the same data with different weighting schemes, they ought to be correlated. We take this correlation into account by jointly estimating the asymptotic covariance matrix for the three quantiles using non-parametric techniques. See Appendix 2.6 for more details about the quantile regression estimation.

The constants in this quantile regression correspond to the productivity quartiles of the baseline treatment under average working conditions (in the reference group). Obviously, the constant over conditional quartiles indicates that the 25$^{\text{th}}$ percentile is 423 trees per day, the median is $1,164$, and the 75$^{\text{th}}$ percentile is $2,024$ trees.

The effect of the two experimental treatments is clear for the lower tail of the productivity distribution, but not for its higher end. In column (b), a small increase in the piece rate ($T1$) shifts upwards the 25$^{\text{th}}$ percentile by almost 65% (271 trees), while a larger increase in the piece rate ($T2$) increases productivity by 83% (351 trees). These scale effects compress the lower tail of the conditional productivity distribution, which will in turn affect the conditional mean ceteris paribus. The effects of the experimental treatments on the rest of the distribution are not statistically significant. In column (c) the treatment effects are -43 trees for $T1$ and -21 trees for $T2$. Both estimates are not statistically significant at 10% and represent only 3.7% and 1.8% of the median respectively. The treatment effects on the upper tails of the conditional productivity distribution are also not significant at 10% level. Changing the piece rate from 14¢ to 17¢ ($T1$) is estimated to shift downwards the 75$^{\text{th}}$ percentile by 11% (-231 trees), while changing the piece rate from 14¢ to 19¢ ($T2$) shifts it upwards by 2.7% (55 trees). A possible explanation is that at the upper end of the productivity distribution the cost of effort is very high. For these workers the income effect of the treatment dominates the effort incentive.

Overall, we observe scale effects that compress the conditional productivity distribution shifting the 25$^{\text{th}}$ percentile and the 75$^{\text{th}}$ percentile towards the median. These results suggest distributional effects of the experimental treatment and the rejection of a linear structure for the production function. However, there will always be numerical discrepancies on the treatment impact when measured at different points of a distribution. A formal test for heteroskedasticity is required to evaluate whether these observed differences in the slope coefficients across quantiles are statistically significant.

Similar to the mean regression, weather conditions have a moderate impact on the conditional productivity distribution. The coefficients associated to maximum daily temperatures and precipitations have the expected sign but are not statistically significant. If anything, high temperatures and rainfall tend to decrease productivity. The effect of minimum daily temperature affects the two tails of the productivity distribution, but not its median. An increase of one degree Celsius in the minimum temperature increases by 13 trees the value of the 25$^{\text{th}}$ percentile (significant at 5%), and by 20 trees the 75$^{\text{th}}$ percentile (significant at 5%).

The correlated random effects control for unobserved individual characteristics that may determine productivity and interact with the experimental treatments. These unobserved factors play a significant role in the case of the large piece rate increase. The CRE $\overline{T2}_i$ captures unobserved fixed effects of increasing the piece rate from 14¢ to 19¢. These effects are significant at 1% level for the 25$^{\text{th}}$ and 50$^{\text{th}}$ percentile of the conditional distribution, and significant at 5% for the 75$^{\text{th}}$ percentile. This suggests that the effect of the second experimental treatment depends on unobserved heterogenous characteristics of tree-planters. The CRE of the small piece rate increase $\overline{T1}_i$. These individual effects do not statistically affect the 25$^{\text{th}}$ and the 50$^{\text{th}}$ percentile, but reduce the value of the 75$^{\text{th}}$ percentile. This last effect is significant at 10% level.

In summary, the CRE model describes the conditional mean of an additive production function and their reaction to the two experimental increases of the piece rate $T1$ and $T2$. The quantile regression estimates complete the picture of the treatment effect on the conditional distribution. By simply looking at these regression results it is difficult to know if there is evidence of heteroskedasticity or other scale effects. We now turn to the formal test of these hypothesis.

2.5.1 Mean Regression Test

We are not directly interested in heteroskedasticity, but on the structural form of the production function. Separability of the productivity shock, and additivity in particular, are consistent with homoskedasticity with respect to the piece rate in the linear regression model. We concentrate our study on the potential heteroskedasticity associated to the experimental changes in the workers' incentives and how these changes interact with the productivity shocks in the production function. Unlike weather conditions or other exogenous factors, wages and piece rates are incentives that can be easily modified by a firm and constitute a powerful tool for inducing workers productivity in a principal-agent context.

Table 2.3 present the results of the homoskedasticity test (2.7) for the mean regression model in Table 2.2, column (a). The main conclusion is that the regression error can be assumed to be homoskedastic. The $p-$value is 0.59, which suggests that the calculated statistic ($\hat{W} = 1.06$) is likely to come from a distribution that respects the null hypothesis. This conclusion is robust to slight modifications of the test. We obtain similar results when implementing other version of the standard test for homoskedastcity such as the test proposed by Breusch and Pagan (1979) ($p-$value $= 0.574$), Wooldridge (2013) ($p-$value $= 0.591$), and Lu and White (2011) ($p-$value $= 0.440$).

Table 2.3 – Test results for the CRE regression

	$p-$value	
test statistic	$\mathrm{N} \to \infty$	$\mathrm{N} = 270$
1.06	0.588	0.369

One should be careful about overstating the results from a hypothesis test. There are at least two factors that could undermine the capacity of this test to capture heteroskedasticity. First, the limited sample size could reduce the *power* of the test. Second, heteroskedasticity may appear less evident when the treatment effects have a mild impact on the mean outcome.

The power of standard tests such as the test for homoskedasticity is ensured when the sample size tends to infinity ($\mathrm{N} \to \infty$). When the sample is finite there is no guarantee of a test capacity to reject a null hypothesis. We calculate the corrected size and the local power of the Wald tests for our specific sample size of 270 observations. Our results are based on the CRE estimates $\hat{\beta}$ from Table (2.2), column (a). We begin by calculating the vector of residuals $\hat{s} = y - \mathbf{x}\hat{\beta}$ and their standard deviation $\hat{\sigma} = \frac{\hat{s}'\hat{s}}{n-k}$, where $n = 270$ is the number of observations and k the number of regressors. Our data-generating process is $\breve{y} = \mathbf{x}\hat{\beta} + u$, with $u = \varepsilon \cdot \exp(\mathbf{r}\delta)$, and $\varepsilon \sim \mathbb{N}(0, \hat{\sigma})$.

To calculate the corrected test size we simulate $10,000$ samples under the null hypothesis $\delta = [0, 0]$. We find that $3,686$ of these samples generated a test statistic larger than our calculated value 1.06. We also find that the distribution of all test statistics under the null closely follows the theoretical asymptotic distribution $\chi^2_{(2)}$, which means that the test correction is not imperative in this case.

To study the power of the test we select two specific parameter values to form the vector $\delta = [\delta_{T1},\, \delta_{T2}]$ and build the alternative hypothesis

$$H_a:\ \mathrm{Var}(s|\mathbf{r}) = \exp(\mathbf{r}\delta) = \exp(T1\delta_{T1} + T2\delta_{T2})\,, \tag{2.9}$$

Where $T1$ and $T2$ are our two treatment variables. Then we use this skedastic function to simulate $1,000$ samples and perform the test for heteroskedasticity for each one of them. The test power for a particular alternative hypothesis is the proportion of times that we reject the null hypothesis at 5% confidence level. The difficulty with power analysis is that there are infinite alternative hypothesis against which we could challenge a test. Figure 2.2 shows the test power for different values of δ_{T1} and δ_{T2} in the unitary interval centred on zero.

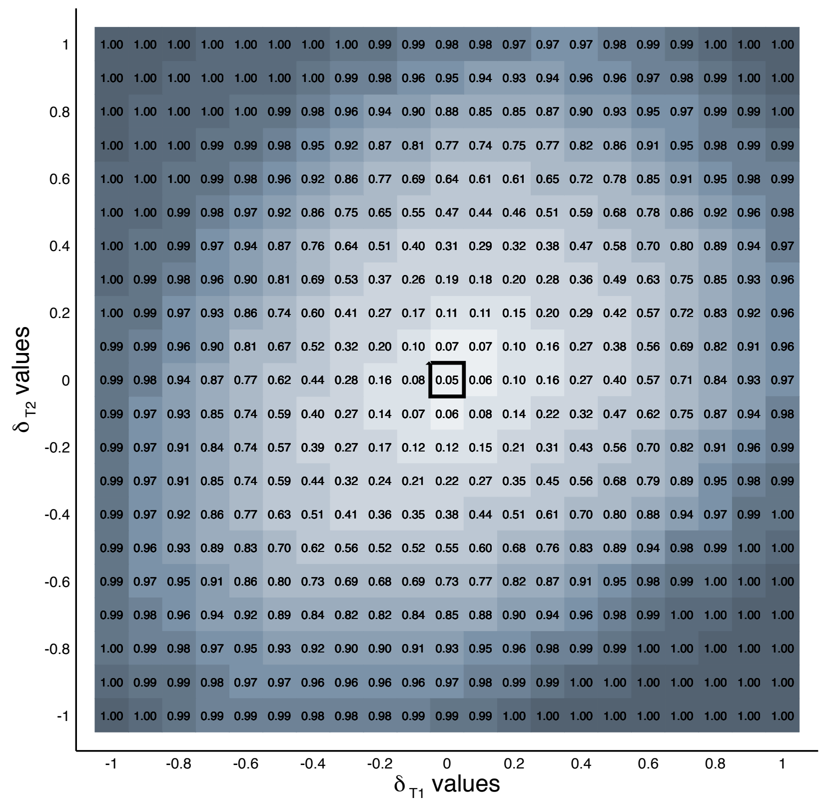

Figure 2.2 – Power of the Wald test across alternative hypothesis

As expected, the test rejects the null hypothesis 5% of the time when the null is true ($\delta = [0,0]$), and the test power increases as the alternative hypothesis departs from the null *i.e.* as δ_{T1} and δ_{T2} increase in absolute value. Naturally, power is lower in the vicinity of the null because the test is less successful in recognizing weak forms of heteroskedasticity.[6]

When testing for heteroskedasticity we evaluate whether the variance of the error term depends on the same variables as those in the regression model (2.4). In our example of tree

[6] In particular, the test power is 0.155 when the alternative hypothesis uses the values $\hat{\delta}_{T1} = -0.134$ and $\hat{\delta}_{T2} = 0.136$, which are the estimates of the augmented regression $(\hat{s}/\hat{\sigma})^2 = \alpha + T1\delta_{T1} + T2\delta_{T2}$ using our data.

planters, heteroskedasticity could be muted by the moderate effect of treatment variables on the mean regression model. It would be worth testing for distributional effects on alternative points other than the mean, where the treatment effects are more salient. Moreover, changes in the piece rate may not only affect the dispersion of the conditional productivity but the entire conditional distribution, stretching one tail of the distribution, compressing the other tail, or even inducing multimodality. We explore broader forms of heteroskedasticity and scale effects using a quantile regression framework.

2.5.2 Quantile Regression Test

A quantile regression test for scale effects evaluates the differences between the slope coefficients of a simultaneous quantile regression. We test the equality of the coefficients associated to $T1$ and $T2$ in Table 2.2. The test is between the three quartiles 25, 50 and 75, in columns (b), (c) and (d) respectively. Systematic differences in the slope coefficient are consistent with scale effects and thus, with a multiplicative error structure in the production function. If instead there are no scale effects, all quantiles should have parameter vectors that differ only in their intercept. In our application to tree planters, the null hypothesis of the quantile regression test can be written as

$$\mathrm{H}_0: \begin{bmatrix} \beta_{T1}(25) \\ \beta_{T2}(25) \end{bmatrix} \begin{matrix} = \\ = \end{matrix} \begin{bmatrix} \beta_{T1}(50) \\ \beta_{T2}(50) \end{bmatrix} \begin{matrix} = \\ = \end{matrix} \begin{bmatrix} \beta_{T1}(75) \\ \beta_{T2}(75) \end{bmatrix}. \tag{2.10}$$

The alternative hypothesis is that at least one of the equalities does not hold. Table 2.4 summarizes the results of the quantile regression test for scale effects on the coefficients of $T1$ and $T2$. The asymptotic theory predicts that there is almost zero percent probability of obtaining a test statistics of 25 when the null hypothesis is true. This rejection of the additive structure of the production function is based on the assumption that the quantile regression test statistic follows a $\chi^2_{(4)}$. This is exact when the number of observations tends to infinite ($\mathrm{N} \to \infty$), but may not be a good approximation when the sample is finite.

Table 2.4 – Test for scale effects
in the simultaneous quantile regression

	p−value	
test statistic	$\mathrm{N} \to \infty$	$\mathrm{N} = 270$
24.98	0.000	0.061

Figure 2.3 compares the theoretical distribution of this test statistic under the null hypothesis with the distribution of the same test statistic corrected for our specific sample of 270 observations. In order to approximate the corrected distribution of the test statistic we simulate $10,000$ samples under the null (with linear and homoskedastic quantiles). The data-generating process is $\breve{y} = \mathbf{x}\hat{\beta} + s; \quad s \sim \mathbb{N}(0, \hat{\sigma})$, where $\hat{\beta}$ is the vector of estimates of the median linear regression model in Table 2.2 column (c), and $\hat{\sigma}$ the estimated standard deviation of the residuals. For each simulated sample we calculate a test statistic and draw their distribution.

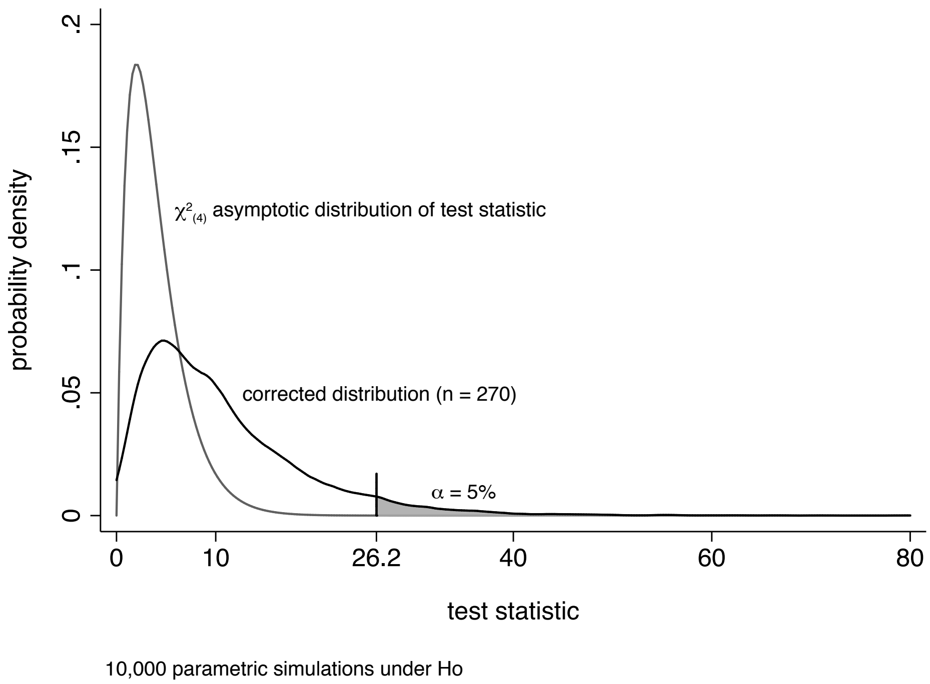

Figure 2.3 – Theoretical and corrected test distribution under the null hypothesis

Unlike the standard test for heteroskedasticity, the asymptotic distribution and the corrected distribution of this test are substantially different. While the critical value at 5% significance is 26.2 in the corrected distribution, the same critical value for the asymptotic distribution is 9.49. The corrected $p-$value is 0.061 in comparison of the theoretical 0.00. Despite the differences between the finite sample and the asymptotic analysis, the results still suggest that the treatment effects (the slope estimates) are statistically different across quantiles, which implies that the piece rate has a scale effect on the conditional distribution of productivity, not only a location effect.

Similar to our approach used to challenge the mean regression test we calculate local power of the quantile regression test with respect to various alternative hypothesis that imply scale effects. We select two parameter values to construct a specific alternative hypothesis $\delta = [\delta_{T1},\, \delta_{T2}]$ and simulate 1,000 data sets using the data-generating process $\breve{y} = \mathbf{x}\hat{\beta} + u$, where $u = \varepsilon\ [\exp(\mathbf{r}\delta)]^{1/2}$; $\varepsilon \sim \mathbb{N}(0, \hat{\sigma})$. For each simulated sample we perform a quantile regression test for heteroskedasticity. The local test power of a particular alternative hypothesis is the proportion of times that we reject the null hypothesis at 5% confidence level (using the corrected test distribution).

Figure 2.4 shows the test power for values of δ_{T1} and δ_{T2} within a two units interval centred around zero. These results suggests that the quantile regression test requires the alternative hypothesis to be relatively distant from the null to attain reasonable power. However, low power is not a problem for our particular application because the test does reject the null hypothesis of no scale effects.

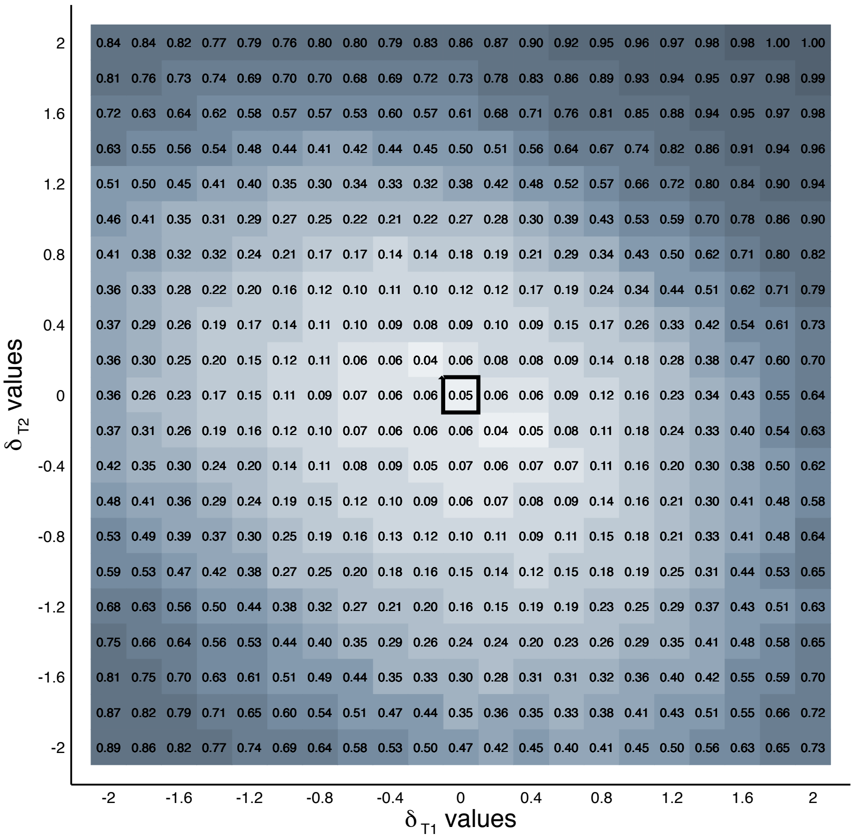

Figure 2.4 – Power of the quantile test for scale effects across alternative hypothesis

Our empirical analysis of the quantile regression test is specific to our application of tree planters. These results can be considered preliminary because they cannot be extrapolated to other contexts. Understanding general test features such as size or power require a more judicious investigation. Future analysis should explore the test results under different sample sizes and alternative types of scale effects.

A common exercise in empirical studies consists in exploring how the main results change when analyzing particular subsamples of the data. Identifying the core attributes of the data that drive the main results improves our understanding of the research results and their interpretation. In our application the main treatment effects as well as the scale effects are located at the bottom of the productivity distribution. These observations rise questions about the characteristics of the corresponding workers and whether they meet the minimum productivity standards required by the firm.

In 2013 all adult workers in British Columbia were entitled to be paid at least a minimum wage of $82 per day.[7] This means that if a worker does not reach an average productivity of 585 trees planted per day the firm will have to top-up their payment to the minimum wage required by law. Our data contains 31 observations that do not meet the minimum daily productivity requirements. In normal conditions these productivity levels generate extra cost

[7]Source: http://labour.gc.ca

and the firm will certainly dismiss the worker in the long run. However, these observations of low productivity may corresponds to workers who are not planting full time (foremen and supervisors), or planters who suffer temporary productivity shocks (for example a physical injury or transportation problems in a given day).

Table 2.5 shows the results of the quantile regression test excluding observations that correspond to earnings lower than the minimum wage. This small change in the empirical sample overturns the previous result. The evidence of scale effects disappear when we exclude low productivity observations. This means that the distributional effects of the treatments are exclusively driven by the observations on the lower tail of the productivity distribution. Another possible interpretation is that the interaction between incentives and productivity shocks depends on the value of the shock but also on its direction. Negative productivity shocks may lead to a reoptimisation of the agent's effort choice considering effort incentives, while positive shocks may not. The possibility that observations at the lower end of the productivity distribution in our data correspond to outliers caused by measurement errors instead of true negative productivity shocks remains present.

Table 2.5 – Test for scale effects
excluding low productivity observations

	$p-$value	
test statistic	$N \to \infty$	$N = 239$
13.57	0.0087	0.269

If we assume that our data reflects true productivity outcomes, we can conclude that a production function with a multiplicative error term is the more adequate model to describe workers productivity. The observed values of the productivity shock do play a role on workers choice of effort and interact with productivity incentives such as the piece rate. This means that workers may choose their effort not only based on the piece rate offered by the firm, but also on the value of their own shock. A potential form of interaction would be for example that planters tend to work harder when they found out that the soil in a given planting plot is hard (negative productivity shock). More technically, the rejection of scale effects in the additive production means that a separable production function cannot model the true relationship between effort and the shocks. A multiplicative structure seems to be more adequate.

2.6 Conclusions

Structural assumptions about the production function have important consequences on the analysis of agents' optimal behaviour and ultimately on the choice of adequate incentives to enhance their productivity. Using statistical tools to test between models with additive and multiplicative shocks is a relevant question for agency models. On the one hand, a separable production function rules out uncertainty considerations. In this case the optimal effort choice is independent of productivity shocks, which are the only source of randomness. For example under an additive structure, both the principal and the agent would gain from contracts that induce effort by only rewarding performance. On the other hand, a multiplicative production function would lead to an effort choice that is sensitive to productivity shocks. This form may imply potential gains in designing state dependent incentives that change with the realization of the productivity shock, offering workers a wider range of piece rate depending on the working conditions, insuring their risk, sorting them across working environment, or even designing contracts according to their specific risk preferences. Moreover, a multiplicative structure would allow for convenient transformations of the production function such as the logarithmic transformation, which facilitate the empirical estimation of workers productivity.

The current empirical literature offers little guidance to applied researchers as to which production structure may be more adequate. There are two major difficulties in choosing between an additive and a multiplicative structure. First, the two factors of interest, effort and productivity shocks are unobserved. Second, and a more serious drawback for empirical applications, effort determinants are endogenous. In general, piece rates, wages, and other effort incentives result from choices made by the firms. These choices that determine effort are based on factors unobserved to the econometrician that also affect workers' productivity.The originality of our approach consists in using data from a field experiment to overcome this endogeneity problem.

We use two experimental treatments to induce exogenous variation on workers incentives to effort. The experiment ensures that there are no endogenous changes on the piece rate or any other productivity incentives chosen by the firm. Our approach is an example of how field experiments can be used as tools to answer questions that can hardly be addressed with observational data. Not only they may serve to evaluate specific treatment effects, but as intermediate tools to respond to a broader range of research questions. Once the endogeneity problem is solved and workers' incentives are correctly identified, the complex choice between an additive or a multiplicative production function can be settled by a a standard test for scale effects. We showed how testing for the additivity between effort and the productivity shocks in the production function boils down to testing for scale effects in a quantile regression. The central idea is that a multiplicative structure leads to scale effects when ignored, while an additive production function is consistent with the absence of scale effects.

Our data comes from a tree-planting firm, located in British Columbia, Canada. We observe the piece rate that workers receive as well as their daily productivity measured by the number of trees planted. We find that a multiplicative production function is more suitable to model planters productivity. This means that the effort exerted by workers is directly related to the productivity shocks they experience. As in any applied research, the external validity or our conclusions is limited by the particularities of our case study. Replication studies in other industries would allow for further comparison. There is no indication that agents in other industries integrate productivity shocks in their effort choice in the same way as do tree planters.

Relaxing some of the assumptions that underlie our analysis would be an interesting direction for future research. In particular, it would be important to test our assumption about the timing of the shock. In this essay we considered optimal effort as independent of risk preferences. This is due to the fact that tree planters observe working conditions before selecting their effort level. There is no risk once the productivity shock is observed. Now that we have evidence that optimal effort is a function of the shock, it seems natural to further explore the question of its timing.

2.7 Bibliography

Allen, D. and Lueck, D. (1992). Contract Choice in Modern Agriculture: Cash Rent versus Cropshare. *Journal of Law and Economics*, 35(2):397–426.

Bellemare, C. and Shearer, B. (2010). Sorting, incentives and risk preferences: Evidence from a field experiment. *Economics Letters*, 108(3):345–348.

Bellemare, C. and Shearer, B. (2013). Multidimensional heterogeneity and the economic importance of risk and matching: evidence from contractual data and field experiments. *RAND Journal of Economics*, 44(2):361–389.

Breusch, T. S. and Pagan, A. R. (1979). A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica*, 47(5):1287–94.

Buchinsky, M. (1998). Recent Advances in Quantile Regression Models: A Practical Guideline for Empirical Research. *Journal of Human Resources*, 33(1):88–126.

Chamberlain, G. (1980). Analysis of Covariance with Qualitative Data. *Review of Economic Studies*, 47(1):225–38.

Chamberlain, G. (1982). Multivariate regression models for panel data. *Journal of Econometrics*, 18(1):5–46.

Dubois, P. and Vukina, T. (2009). Optimal incentives under moral hazard and heterogeneous agents: Evidence from production contracts data. *International Journal of Industrial Organization*, 27(4):489 – 500.

Koenker, R. W. and Bassett, Gilbert, J. (1978). Regression Quantiles. *Econometrica*, 46(1):33–50.

Lu, X. and White, H. (2011). Testing for Separability in Structural Equations. Working Papers.

Mundlak, Y. (1978). On the Pooling of Time Series and Cross Section Data. *Econometrica*, 46(1):69–85.

Paarsch, H. J. and Shearer, B. S. (1999). The Response of Worker Effort to Piece Rates: Evidence from the British Columbia Tree-Planting Industry. *Journal of Human Resources*, 34(4):643–667.

Portnoy, S. and Koenker, R. (1997). The Gaussian Hare and the Laplacian Tortoise: Computability of Squared- Error versus Absolute-Error Estimators. *Statistical Science*, 12(4):pp. 279–296.

Powell, J. L. (1984). Least absolute deviations estimation for the censored regression model. *Journal of Econometrics*, 25(3):303–325.

Powell, J. L. (1986). Censored regression quantiles. *Journal of Econometrics*, 32(1):143–155.

Prendergast, C. (2002). The Tenuous Trade-off between Risk and Incentives. *Journal of Political Economy*, 110(5):1071–1102.

Shearer, B. (2004). Piece Rates, Fixed Wages and Incentives: Evidence from a Field Experiment. *Review of Economic Studies*, 71(2):513–534.

Appendix

2.A Linear Regression

We have judiciously selected the econometric specification that best fits our data. Table 2.A.1 presents different estimates of the additive production function in equation (2.4).

Table 2.A.1 – Linear regression model

	OLS† (a)	FE (b)	RE (c)	CRE† (d)	LAD (e)
constant	1334.0***		1383.4***	1163.6*	1559.1***
	(255.9)		(193.0)	(610.6)	(341.7)
small PR increase ($T1$)	180.7	130.1	132.5	150.2	31.4
	(148.0)	(84.1)	(83.6)	(129.5)	(162.6)
large PR increase ($T2$)	359.8**	290.5***	293.7***	291.3**	165.1
	(164.7)	(93.8)	(93.2)	(114.7)	(249.7)
daily highest temp.	-10.4	-11.9	-11.8	-12.7	-31.6
	(31.5)	(17.8)	(17.7)	(11.2)	(44.6)
daily lowest temp.	12.8*	12.6***	12.6***	13.6*	4.0
	(7.0)	(4.0)	(4.0)	(7.8)	(10.7)
daily precipitations	-41.2	-37.4	-37.5	-45.6**	-34.3
	(57.1)	(32.2)	(32.0)	(21.8)	(91.5)
correlated effect $\overline{T1}_i$				-1587.4	
				(4657.2)	
correlated effect $\overline{T2}_i$				2520.8	
				(2425.9)	
no. of parameters	6	5	6	8	6
number of observations	270				
number of individuals	21				
Hausman FE vs. RE	1.687				
Hausman $p-$value	0.891				

*** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1.
† Standard errors clustered by worker.

Column (a) shows the results of a basic ordinary least square (OLS) regression with panel-corrected standard errors. The main disadvantage of this approach is that it ignores the almost certain serial correlation of individuals' productivity over time. A fixed effect model (FE) estimated by mean differences in column (b) is better suited to model the special characteristics our panel sample. FE are less restrictive and allow each planter to have a different intercept, their drawback is that these constant "absorbs" all individual time-invariant factors and thus, the model constant can no longer be estimated. The random effects model (RE) in column (d) provides a potential solution. It accounts for the panel nature of the data and estimates the effect of time invariant characteristics. The RE approach produces efficient estimates under the so called *strict exogeneity* hypothesis. The validity of this assumption is endorsed by a Hausman test, which suggest the independence between the unobserved individual characteristics and the regression error (p-value$=$ 0.992). The correlated random effects model (CRE) (Mundlak, 1978) in column (c) is a middle ground between FE and RE. It controls for the planter-specific factors by including individual averages and imposes less restrictive assumptions than RE. In addition, the CRE can be used for mean regression analysis as well as in a quantile regression framework. Finally, column (d) corresponds to the least absolute deviations estimator (LAD). This is also a central tendency model with a single intercept, but it refers to the conditional median instead of the conditional mean. With respect to the mean regression analysis, LAD estimates have the advantages of being insensitive to outliers.

2.B Quantile Estimates

Let X be a matrix containing the regressors in our quantile regression model and $\gamma(\theta) = [\beta_0(\theta), \beta'(\theta)]'$ the vector of associated parameters, including the constant, the experimental indicators and the control variables for weather and correlated random effects. The main idea of the quantile regression (Koenker and Bassett, 1978) is to find a parameter vector $\widehat{\gamma}(\theta)$ that minimizes the continuous and piecewise linear loss function

$$\rho_\theta(y - X'\gamma(\theta)).$$

Unfortunately, this optimization problem has no elegant closed-form solution. We use the algorithm proposed by Portnoy and Koenker (1997) to find our quantile estimate $\widehat{\gamma}(\theta)$.

2.B.1 Asymptotic Distribution

Let $\gamma_\theta = [\gamma'(\theta_1), \gamma'(\theta_2), \ldots, \gamma'(\theta_P)]'$ denote the vector of parameters form a sequence of P quantile regressions $\{\theta_1, \ldots, \theta_P\}$. Powell (1984) shows that, under some regularity conditions

$$\sqrt{N}(\widehat{\gamma}_\theta - \gamma_\theta) \xrightarrow{L} \mathbf{N}(0, \mathbf{\Lambda}), \qquad \text{with } \mathbf{\Lambda} = \{\mathbf{\Lambda}_{p,\,l}\}_{p,\,l\,=\,\theta_1,\,\ldots,\,\theta_P}\,.$$

To ensure the validity of our inference we use a general form of the covariance matrix that is valid under any dependence structure (homoskedasticity or heteroskedasticity)

$$\mathbf{\Lambda}_{p,\,l} = (\min\{\theta_p, \theta_l\} - \theta_p\theta_l)\left[\mathbf{E}\left(f_{s_{\theta_p}}(0|X)X'X\right)\right]^{-1}\mathbf{E}(X'X)\left[\mathbf{E}\left(f_{s_{\theta_l}}(0|X)X'X\right)\right]^{-1}_{.}$$

Problems in estimating the covariance matrix $\mathbf{\Lambda}$ arise mainly with regard to $f_{s_\theta}(0|X)$, the distribution of the regression error evaluated at zero. We have chosen to estimate a covariance matrix $\hat{\mathbf{\Lambda}}$ using the so called kernel estimator approach (Powell, 1986), which consists in approximating the asymptotic covariance matrix with non parametric estimates of its components

$$\widehat{\mathbf{E}}(X'X)^{-1} = \left(\frac{1}{N}X'X\right)^{-1}, \quad \text{and}$$

$$\widehat{\mathbf{E}}\left(\hat{f}_{s_{\theta_p}}(0|X)X'X\right) = \frac{1}{Nh}\sum_{i=1}^{N} K\left(\frac{y_i - X_i\widehat{\gamma}(\theta_p)}{h}\right)X_iX_{i.}'$$

We use the Gaussian kernel density K and the Silverman's rule-of-thumb to determine its optimal bandwidth h. Our choice is mainly driven by the relatively low computational power required by this non parametric approach. We are confident in choosing this method over the alternative bootstrap approach since both are valid and equivalent under any dependence structure. It is worth noticing that kernel estimators provide smaller standard errors, at least in our application.

Chapter 3

Conditional and Unconditional Cooperation in a Public Goods Game: Experimental evidence from Mali

3.1 Introduction

When government institutions cannot guarantee the provision of public goods, social welfare relies on community cooperation. Behavioural economics can contribute to the search for incentives that influence individual actions in a desired direction. In particular, it can help us find better incentives for contributing to public goods without having to impose regulations that are costly to enforce and may create conflict. Experimental economists have identified various motives that affect voluntary contributions to public goods aside from direct pecuniary benefits. The choice of contributing in a public goods game depends on intrinsic preferences as well as on expectations about the actions of other individuals. Understanding the influence of expectations on the contribution choice can be useful for public policy purposes.

This paper disentangles between two rationales underlying the decision to contribute or not to a public good. First, *unconditional cooperation* results from intrinsic individual preferences that are independent of others' behaviour, such as altruism or egoism. The choice of contributing may also come from *conditional cooperation.* This principle is inherently related to the behaviour of others or to expectations about their actions when behaviour is not directly observed. Conditional cooperation arises for example when individuals try to match others' actions. Behind conditional and unconditional cooperation may lie other type of preferences, for instance a concern or indifference for equity, or the desire to conform to social norms. Most preferences can be classified as conditional or unconditional cooperation depending on their correlation with other people's actions. For instance equity concerns motivate unconditional cooperation when they are independent of the actions of others', but may very well be at the heart of conditional cooperation when the choice of other individuals is taken into account, which is often case in practice. Taken together, conditional and unconditional cooperation determine the decision of contributing to public goods.

I estimate a decision-making model that uses conditional and unconditional cooperation to describe individual choices in a public goods game. I use random coefficients to allow heterogeneous preferences to depend on individual characteristics and unobserved factors such as social norms and cultural practices. This model also allows for correlations between preferences. For example, participants who weight more the actions of others in their own decision to cooperate may have a systematic tendency to be less altruistic, leading to a correlation between conditional and unconditional cooperation.

A distinctive feature of the behavioural analysis in this paper is that instead of relying on strong assumptions such as *rational expectations* I use elicited beliefs to measure the relative importance of conditional cooperation. In other words, I measure the extent to which the expected actions of other individuals affect the decision to cooperate. I estimate the model using data from $2,697$ individuals who participated in a contextual field experiment conducted in 121 rural communities in Mali in 2011.

The experiment is a repeated binary linear public goods game in which a group of participants simultaneously and individually decide whether or not to invest in a public good. In addition to the game, participants reveal their expectations about total public goods provision. As most public goods games, this experiment poses a social dilemma: on the one hand, participants maximize total social benefit by contributing to the public good; on the other hand, they maximize their own private benefit by not contributing.

The experiment consists of three choice periods that correspond to three different treatments. The first period is a *baseline* treatment in which participants remain physically distant from each other. The second and the third period randomly alternate between a *discussion* treatment and a *leader* treatment. In the discussion treatment participants are invited to hold an open conversation among them. After this talk, everyone makes a private cooperation choice and state their subjective beliefs. In the leader treatment, one of the participants is randomly chosen from the group and is given the mandate to convince everyone else to cooperate in order to maximize social welfare.

I find that unconditional cooperation can be partially explained by common preferences shared by all participants, but also depends on observed individual characteristics. Younger and wealthier participants are less inclined towards unconditional cooperation. Conditional cooperation is much more heterogeneous and depends on individual factors, mainly unobserved. The discussion treatment increased cooperation by 7.6%. The estimation of the structural model indicates that this effect is primarily driven by conditional cooperation. When using my structural model to predict individual behaviour, I find that conditional cooperation is responsible for almost 24% of the observed public goods provision. Moreover, I find that the leader treatment increases total public goods provision by 14%. According to the model, this improvement is mainly due to unconditional cooperation. Even in the most pessimistic scenario, in which all participants expect zero public goods provision, the structural model predicts that 60% of the group will stills choose to cooperate if a local leader is present to motivate them.

These findings may be useful in other regions of sub-Saharan Africa, in contexts similar to rural Mali. They could be helpful to improve the probability of success of projects that require community cooperation and for which supervision is difficult or too costly. The involvement of local leaders appear to be the most effective tool to incentivize cooperation and could be used in health campaigns to promote choices such as using condoms, sleeping under bed nets, chlorinating water, or hand washing. However simple, these actions can make a real difference. In a single year, 1.8 million people die from AIDS, $655,000$ die from malaria, and 1.5 million children die from diarrhea (WHO, 2012; UNICEF and WHO, 2012; UNICEF et al., 2011).

The rest of the paper is organized as follows. Section 3.2 discusses conditional cooperation, unconditional cooperation and expectations in the context of public goods. Section 3.3 describes the experimental design and Section 3.4 presents the structural choice model. Section 3.5 presents the data and the estimation results. Finally, Section 3.8 concludes.

3.2 Background

Unconditional cooperation is based on intrinsic preferences regardless of the actions of other individuals. In the context of public goods games, unconditional cooperation may include motivations such as egoism or altruism. Egoism motivates participants to *free ride* - to benefit from the cooperation of other participants without contributing to the public good. Numerous public goods experiments have shown that egoism alone cannot explain observed contributions to public goods. Unconditional cooperation is consistent with this evidence because it includes preferences like altruism, which does not depend on others' actions. Individuals may have an intrinsic taste for giving (Becker, 1974) or they may get a *warm glow* from giving (Andreoni, 1989, 1990). The utility of altruism comes from the action of contributing itself and it always motivates individuals to cooperate. Goeree et al. (1999) describe an alternative type of altruistic preferences that depend on the utility of others, but not on their actions. Another motivation consistent with unconditional cooperation is efficiency. In a public goods game, efficiency considerations incentivize participants to choose the action that maximizes the total net benefits and not their own private benefits.

Conditional cooperation requires individual choices to depend on the actions of others (Gächter, 2007). The experimental evidence suggests that in public goods games individuals are often willing to contribute more the more the others contribute (*e.g.* Fischbacher and Gächter, 2010). Conditional cooperation can be motivated for instance by inequality concerns (Fehr and Schmidt, 1999), when individuals dislike a particular distribution of payoffs; or by fairness concerns (Rabin, 1993), when individuals seek to reward contributors and to punish non contributors.

This paper approaches the cooperation choice as a complex decision in which a variety of preferences intervene in potentially opposite directions. I measure the relative importance of conditional and unconditional cooperation when a group of individuals decide whether or not to contribute in a public goods game. Since decisions are taken simultaneously, participants make their choices without knowing the actions of the rest of the group. Due to this uncertainty, conditional cooperation requires participants to form expectations, or beliefs about the others' behaviour.

There is an identification problem in estimating conditional cooperation. When only final contribution decisions are observed, different combinations of preferences and expectations can lead to identical choices (Manski, 2002). One possible solution consists in assuming rational expectations; in other words, assuming that participants can predict the actions of the rest of the group on average. Avoiding such assumptions on participants' behaviour is the main justification for using experimental data on beliefs instead of the regular observational data on individual choices. Bellemare et al. (2008) show that a structural model of decision making (for a ultimatum game) generates much better predictions when estimated with elicited beliefs instead of assuming rational expectations. Following their result, I prefer to use data on expectations to estimate my contribution-choice model.

Concerns have been raised in the experimental literature that eliciting beliefs may lead to more strategic thinking and therefore affect behaviour. Rutström and Wilcox (2009) find that asking subjects their beliefs during a repeated game changes the way those subjects play only when using a *scoring rule* to incentivize accuracy. Not rewarding accuracy improves the likeliness that beliefs affect choices exclusively through the expected action taken by other participants. This avoids potential biases in participant choices, minimizes hedging opportunities, and improves cognitive simplicity of the instructions.

3.3 Experimental Design

According to the taxonomy of Harrison and List (2004), the experiment studied in this paper can be classified as a *contextual field experiment* - a controlled laboratory design adapted to the Malian cultural context.[1] The experiment involves three treatments or periods. Each treatment includes a *public goods game* and a *beliefs elicitation* question in which participants privately report their expectations about the unknown public goods provision. Total outcomes are revealed at the end of the three periods, but elicited expectations are never made public.

The public goods game is a simplified linear game of binary choices that closely follows the design of Cárdenas et al. (2009). All participants receive an endowment of one token and take the simultaneous and anonymous decision of cooperating or not. Cooperating means investing the entire endowment into a common account that is a public good. The total amount of this account is multiplied by the number of participants and the returns are shared equally among all group members, cooperators and non-cooperators. This means that the marginal per capita return from the public good is constant. Choosing not to cooperate means investing the entire endowment in a private account that has a fixed private return of nine tokens. In total, cooperators receive the amount of the public goods provision, while non-cooperators receive the amount of the public goods provision plus ten additional tokens from their private account.

Monetary payoffs depend on individuals' actions as well as on the actions of the rest of the group. If nobody contributes to the public good, all participants receive ten tokens from their private account. Inversely, if everyone contributes each participant receives as many tokens as the group size. Since there are always more than ten participants in this experiment, social returns of the public good are always greater than the total returns of the private account.

As in most public goods games, participants' face a social dilemma. On the one hand, the behaviour that maximizes individual payoffs is to free ride, to invest in the private account to receive private returns and also receive returns from the public good. On the other hand, contributing to the public good gives greater social returns, and thus is the optimal strategy to maximize social welfare.

The beliefs question in this experiment seeks to elicit participants' subjective expectations about the proportion of group members that contributed to the public good. In order to simplify communication (Manski and Molinari, 2010), participants reveal their beliefs by choosing one out of five alternatives as depicted in Figure 3.1. The meaning of each alternative from left to right are worded as follows: *none* of the participants contributed, a *few* contributed, around *half* of the participants contributed, *many* contributed, *all* participants contributed.

[1] A detailed experimental protocol is available under request (in French).

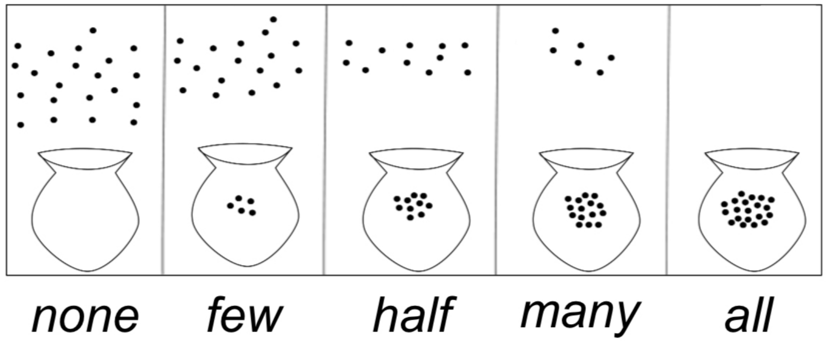

Figure 3.1 – Graphic question for beliefs elicitation

The experiment did not reward the accuracy of elicited expectations to prevent stated beliefs from becoming part of the game strategy and potentially affect the contribution choice. Armantier and Treich (2013) show theoretical and empirical evidence that paying individuals for their predictions can lead to a significant bias. They find that incentivizing beliefs through a scoring rule when individuals have a financial stake in the predicted event, as they do in public goods games, produces systematic differences between subjective and reported beliefs. Palfrey and Wang (2009) also find empirical evidence that scoring rules can create significant complex distortions in the observed outcomes when there are prominent hedging opportunities. A potential solution to make the experimental design "hedging proof" is to randomly reward either the accuracy of elicited beliefs or the game outcomes (Blanco et al., 2010). This solution however, carries the price of adding cognitive complexity to the experiment instructions, which can be a major issue when individuals have low literacy levels as they do in rural communities in Mali.

The experiment includes three treatments or periods that correspond to three different versions of the public goods game. The first period is always a *baseline* treatment in which participants remain physically distant from each other the entire time. In this treatment, participants play a standard public goods game and state their beliefs about total public good provision. The second and the third period randomly alternate between a *discussion* treatment and a *leader* treatment.

In the discussion treatment, participants are allowed to have an open discussion among them; they communicate freely and potentially make non-binding and non-verifiable agreements. After 5 to 10 minutes, participants are asked to make their own contribution decisions in private and state their beliefs. In the leader treatment, one of the participants is randomly chosen from the group to lead the discussion. This person is brought apart and told that the group's optimal solution to the game is to cooperate. The leader's explicit mandate is to convince everyone else to contribute and maximize social welfare, this person has a few minutes to convince the group before all participants take their own private contribution decisions and state their beliefs.

In this experiment the leader's decision is simultaneous and identical to the rest of the group. The selected leader has no power to punish or reward or even verify the actions of the group members. This setting departs from other experimental treatments where the leader's decision differs from the decisions of the rest of participants. While some experiments reveal the leader's contribution ex-ante, others simply give the leader special capacities such as communicating (Koukoumelis et al., 2012), monitoring the decisions of others, or rewarding and punishing the rest of the group (Van der Heijden et al., 2009; Rivas and Sutter, 2011).

Finally, the total public good provision of each game is revealed at the very end of the experimental session after the three periods. According to Costa-Gomes and Weizsäcker (2008), postponing the feedback about outcomes until the end of the experiment reduces the dynamics between outcomes and decisions or expectations. In my analysis, I assume that cooperation decisions are based solely on preferences and beliefs, and not on the actual outcome of previous periods.

At the end of the experimental session, participants use their tokens to "buy" prizes from a temporary shop managed by experimenters. The articles available are gender free and consist of pens, lighters, matches, notebooks, razors, batteries, and lamps.[2].

3.4 Model

Based on the premise that conditional and unconditional cooperation motivates individual choices, I propose a model that describes the cooperation decision in the experiment described above. The interest of using an economic model is to recognize behavioural patterns and stylized facts about individual preferences that go beyond simple correlations. In particular, the model allows me to identify the channels through which the *discussion* treatment and the *leader* treatment affect individual choices.

In each experimental session $k = 1, 2, \dots, 121$, a group of N_k participants interact together over the three treatments or periods $t = 1, 2, 3$. In a given period, each individual $i = 1, 2, \dots, N_k$ receives a unitary endowment and makes a private binary choice $c_{it} \in \{1, 0\}$. Participants who choose to contribute to the public good ($c_{it} = 1$) invest their endowment in a common account that returns one unit to each one of the N_k group members. Participants who choose not to contribute to the public good ($c_{it} = 0$) invest their endowment in a private account that gives an individual private return of ten, and zero returns to the rest of the group. The individual payoff of this game can be written as a linear function of the choice variable

$$m(c_{it}) = 10\,(1 - c_{it}) + c_{it} + (N_k - 1)\,c'_{it,} \tag{3.1}$$

where $c'_{it} \in [0, 1]$ denotes the average contribution of the rest of the group. Since participants interact anonymously, the rest of the group can be modeled as a unique player with a continuous contribution within the unit interval.

This model defines the utility of contributing to the public goods as a broad function that includes not only individual monetary payoffs, but also the notions of unconditional and conditional cooperation

$$u_{it} = m(c_{it}) + [\alpha + \gamma(N_k - 10)]\,c_{it} - \theta\,|c_{it} - c'_{it}|. \tag{3.2}$$

The component $m(c_{it})$, which coefficient is normalized to one, is simply an expression of standard preferences for individual monetary payoffs. This first component of the utility function follows the theory applied to the early studies of voluntary contributions, which assumes that participants are selfish payoff maximizers. In the laboratory, there is always

[2]Item prices in tokes: pen=5; matches=10; notebook=20; razor=30; batteries=50; lamp=80.

a fraction of subjects whose behaviour is consistent with this notion. Andreoni and Miller (2002) found that a quarter of subjects participating to a dictator game were not willing to share their payoff with another participant. Nevertheless, the assumption of completely selfish players typically fails in public goods games, and thus it is necessary to adjust the utility function accordingly.

The parameter α can be interpreted as a preference capturing the Andreoni (1989) *warm glow giving*: the individual satisfaction from the act of contributing per se, regardless of the actions of the others. Existing experimental evidence clearly shows that subjects in the laboratory have an interest in behaving unselfishly. Multiple studies of the dictator game provide evidence of altruistic preferences (*e.g.* Robert et al., 1994; Elizabeth et al., 1994; Bolton et al., 1998; Andreoni and Miller, 2002).

In this experiment, total returns from the public good increase with group size because each participant receives the total amount of the common account. The parameter γ can be interpreted as representing preferences for efficiency, or any other utility related to the group size. If the number of participants was ten ($N_k = 10$), there would be no gain in efficiency for contributing to the pubic good, because investing into the private account would generate the same net returns as investing into the common account. In this experiment, the social returns are always larger than the private returns ($N_k > 10$). Consequently, the parameter γ measures the benefit of the additional total returns from contributing to the public goods ($N_k - 10$).

Lastly, the parameter θ preceded by a minus sign represents the cost of deviating from the average contribution of the rest of the group. The term $|c_{it} - c'_{it}|$ is a linear and symmetric function that relates individual choices to other participants' choices. If $\theta > 0$, the conditional cooperation parameter is a penalty to deviations from the actions of the majority. This structure conveys the idea that the more a group contributes to the public goods, the more each participant is willing to contribute himself. Motives such as fairness, inequality concerns and reciprocity are often evoked as explanations for this conditional cooperation behaviour (Keser and van Winden, 2000; Offerman et al., 2001; Fischbacher et al., 2001).

Replacing the monetary payoff (3.1) in the utility function (3.2), the utility function can be written as

$$u_{it} = [\alpha - 9 + \gamma(N_k - 10)]\, c_{it} - \theta\, |c_{it} - c'_{it}|.$$

The term $\alpha - 9 + \gamma(N_k - 10)$ captures unconditional cooperation preferences, which do not depend on the actions of the group. The nine units of utility subtracted represent the opportunity cost of contributing to the public good, which is the forgone return of the private account. If the altruism parameter was the only preference in play, participants would contribute when $\alpha > 9$. More realistically, α is expected to be positive if the act of contributing is gratifying.

This analysis assumes participants' rationality throughout. Individuals choose to cooperate if their expected net benefit from doing so is at least as great as the expected net benefit from not cooperating. Consequently, their choice is based on the utility differential between the two actions. To make explicit that the cooperation choice depends on the actions of the group, I write the utility differential as a function of the average group contribution

$$\Delta u(c'_{it}) = u\,(c'_{it}|c_{it}=1) - u\,(c'_{it}|c_{it}=0) = \alpha - 9 - \theta + \gamma(N_k - 10) + 2\theta\, c'_{it.}$$

Since all participants take their decisions simultaneously, individuals have to rely on expectations or beliefs about the actions of rest of the group. This uncertainty can be modeled as a censored distribution function over the interval $[0, 1]$, with expected value $\mathbf{E}\, c'_{it}$. The expected utility differential is

$$\Delta u^e_{it} = \alpha - 9 - \theta + \gamma(N_k - 10) + 2\theta\ \mathbf{E}\, c'_{it.}$$

The difference in expected utility Δu^e_{it} is not directly observed. I observe participants' binary contribution decisions and assume that

$$c_{it} = \begin{cases} 1 & \text{if } \Delta u^e_{it} > 0, \\ 0 & \text{otherwise.} \end{cases}$$

3.4.1 Econometric Model

In principle the random utility approach permits to estimate the choice model (3.3) by simply pluging reported beliefs into the expected utility differential, adding an error term that follows a specific distribution, and running a logit or a probit regression. This is an interesting approach when expectations ($\mathbf{E}\, c'_{it}$) are continuous. In this particular experiment however, given the discrete nature of the elicited beliefs, the econometric model needs to be adapted. I assume that each participant has a subjective probability distribution of the public good provision and reports the alternative belief that is closest to their mean. Although the beliefs question provides little guidance on how to associate each alternative to a numerical scale, it seems plausible to assume that participants interpret the category *none* as 0% cooperation and the category *all* as 100% cooperation. From this perspective, I set the first alternative to zero and the last one to one. This restriction allows me to identify conditional cooperation. For the remaining alternatives I use dummy variables D_{few}, D_{half}, D_{many}. The estimated model is

$$\begin{aligned} \Delta u^e_{it} &= \alpha - 9 - \theta + \gamma(N_k - 10) \\ &\quad + 2\theta \left(\theta^{few} D_{few} + \theta^{half} D_{half} + \theta^{many} D_{many} + D_{all}\right) + \varepsilon^u_{it.} \end{aligned} \tag{3.3}$$

The parameter θ captures participants' conditional cooperation. It measures the "desirability" of contributing when the expected contributions change; in other words, the relative utility of contributing when a participant goes from thinking that 0% of the group will contribute to thinking that 100% of the group will contribute. The parameters θ^{few}, θ^{half}, and θ^{many} determine changes in the utility of cooperating for each alternative in the beliefs question. Naturally, they have to be interpreted with respect to the first omitted category (0% expected cooperation).

In random utility models, the contribution choice is not purely deterministic. The expected utility differential is influenced by a variety of factors modeled here as random errors iid

$$\varepsilon_{it} \sim N(0, 1).$$

The error variance is normalized to one because, as in all discrete choice models, the coefficients are identified up to a scale factor.

3.4.2 Heterogeneous Preferences

Conditional and unconditional cooperation are likely to be determined by individual factors as well as social norms unobserved by the researcher. I use a random coefficients model that provides an explicit characterization of the heterogeneity that exists among participants and across the communities. I model the two main parameters of interest $\beta_i \in \{\alpha_i, \theta_i\}$ as combinations of deterministic components and random components

$$\beta_i = \beta_0 + X_i'\beta + \eta_i^\beta + v_k^\beta. \tag{3.4}$$

The deterministic components include a constant β_0, which represents preferences common to all participants, and a vector β, which represents heterogeneous preferences associated to observed individual characteristics X_i. Moreover, *individual* factors η_i^β and *cultural* factors v_k^β specific to each community capture other elements unobserved to the researcher which may also determine preferences.

Unobserved individual factors account for the fact that two participants with identical observed characteristics can still have different preferences. I model these factors as a vector of random variables specific to each participant i and potentially correlated across preferences

$$\eta_i \begin{pmatrix} \eta_i^\alpha \\ \eta_i^\theta \end{pmatrix} \sim \mathbf{M}\mathbb{N}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \mathrm{Var}(\eta_i^\alpha) & \\ \mathrm{Cov}(\eta_i^\alpha, \eta_i^\theta) & \mathrm{Var}(\eta_i^\theta) \end{pmatrix}\right]. \tag{3.5}$$

Cultural diversity and geographic isolation of the rural communities in Mali shape the protocol of social interactions. These ethnographic factors may determine the choice of contributing to a public good. Even though no empirical model can hope to capture all these features, the random coefficients approach allows preferences to vary across communities, in an attempt to capture some of this cultural heterogeneity. I model these and other unobserved factors specific to each village as a random vector also correlated across preferences

$$v_k = \begin{pmatrix} v_k^\alpha \\ v_k^\theta \end{pmatrix} \sim \mathbf{M}\mathbb{N}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \mathrm{Var}(v_k^\alpha) & \\ \mathrm{Cov}(v_k^\alpha, v_k^\theta) & \mathrm{Var}(v_k^\theta) \end{pmatrix}\right]. \tag{3.6}$$

In summary, the model described in this section is a probit model of the expected utility differential Δu_{it}^e. The discrete nature of the elicited expectations requires normalizing some of the parameters in order to identify conditional and unconditional cooperation. These assumptions are thought to be less restrictive than relying on strong hypothesis such as rational expectations.

3.5 Data and Descriptive Analysis

The data used in this paper was collected in 2011. It consists of experimental observations as well as information on a household survey conducted in 121 rural villages in central Mali. The survey reveals little presence of government institutions and the necessity of cooperative actions to ensure the provision of essential public goods such as health and sanitation. Over all, 43% of these villages have a primary school and 5% have a health center. There is evidence of total coliform bacteria in 70% of the water sources, yet only 45% of households report treating their drinking water. Malaria and "birth complications" are the two leading causes of death. Nonetheless, only 9.7% of the survey respondents report using birth control methods. In this context, understanding individual choices in a public goods game may reveal valuable information on the best way to enhance individual cooperative practices related to health and sanitation. Hand-washing, using latrines, chlorinating water, sleeping under bed nets or using condoms, represent simple but essential contributions that can help improve community welfare but require minimum cooperation rates to lead to significant changes.

Table 3.1 summarizes some demographic characteristics of the 2,697 individuals who participated in the experiment. Their average *age* is 34 years, which corresponds to late adulthood in rural sub-Saharan regions. 43% are males, half of them work in agriculture, and 22% declare to be able to read or to have attended school. In an effort to measure relative wealth I construct the variable *assets*, a continuous positive index based on ownership of agricultural and non-agricultural assets and land.

Table 3.1 – Summary statistics

variables	mean	std. dev.	missing	description
age	34.4	11.85	0	age in years
male	0.43	0.49	0	1 man; 0 woman
agriculture	0.50	0.50	151	1 agricultural occupation; 0 otherwise (statistic only for men)
education	0.22	0.41	196	1 can read or attended school; 0 otherwise
assets	2.69	0.70	154	index measuring capital and land possession
contributions	0.71	0.34	0	total average cooperation rates
baseline	0.67	0.47	0	average cooperation in the baseline treatment
discussion	0.72	0.45	0	average cooperaiton in the discussion treatment
leader	0.76	0.43	0	average cooperation in the leader treatment
group size	22.5	4.01		number of participants in the experiment
communities	121			number of experimental sessions (villages)
participants	2,697			total number of participants

This experiment was replicated in 121 villages, the average number of participants was 22. Over all, individuals contributed to the public good 71% of the time. In the baseline treatment 67% of participants contributed. This result clearly differ from full cooperation or zero cooperation, suggesting that participants are confronted with a true social dilemma in the experiment. Even though experimental evidence is not directly comparable, typical designs of public goods games in the laboratory lead to cooperation proportions that range between 40% and 60% in the first period (Davis and Holt, 1993).

In public good games it is well known that communication enhances cooperation in the laboratory (*e.g.* Isaac and Walker, 1988a) and the field (*e.g.* Cardenas et al., 2000). The experimental results show 72% cooperation in the discussion treatment, which represents an increase of 7.5% with respect to the baseline game (p-value = 0.004). The presence of a leader is also known to increase cooperation(Guth et al., 2007; Moxnes and Heijden, 2000; Koukoumelis et al., 2012; Van der Heijden et al., 2009). In this experiment the leader treatment resulted in 76% contribution to the public good, a relative increase of 13% with respect to the baseline treatment (p-value = 0.000). Contributions in the leader game were 4.1 percentage points larger than in the discussion game. This difference is also statistically significant (p-value = 0.027).[3]

[3]Standard errors used in the $t-$tests were clustered by village.

In general, these results are consistent with the existing evidence in the policy evaluation literature and speak in favour of development initiatives that enhance community involvement, like the *community led projects* currently implemented by UNICEF (Pickering et al., 2016). The observed treatment show that public debates and integration of local leaders are relatively inexpensive tools that can be used to increase cooperation and improve the probability of success of community projects. The remaining question is why are these tools successful. The argument of this paper is that expectations play a major role in the transmission mechanism and their role can be studied and measured through a structural model.

3.5.1 Time Effects

Repeatedly playing a game could potentially give rise to undesired experimental effects such as *learning effects* or *intertemporal strategies.* While the experimental design intends to avoid intertemporal choices by creating a *restart effect*, I find no evidence of learning over time or any other type of correlation across periods.

One of the stylized facts of repeated public goods games is that contributions decrease over time. According to the empirical evidence, repetition "drags down" contributions over periods. This feature has been largely documented in the laboratory (Andreoni, 1988, 1995; Croson, 2007; Davis and Holt, 1993; Fischbacher et al., 2001; Fischbacher and Gächter, 2010), and more recently in field experiments (Walker, 2011).

Table 3.2 shows average cooperation rates and the standard deviations in each period for this experiment. There is no evidence of any decreasing time effect or learning effect on cooperation between the second and the third period.[4] One could imagine a similar results over the three periods. Unfortunately, it is not possible to separately test for time effects between the first and the second period because the baseline treatment always takes place first. If the reader considers that time effects might decrease cooperation between the first and the second period, the estimates of the leader and the communication treatment with respect to the baseline should be interpreted as lower bounds for their true values.

Table 3.2 – Average contributions by period

	first $t = 1$ (a)	second $t = 2$ (b)	third $t = 3$ (c)
Contribution	0.67	0.73	0.75
Standard deviation	0.47	0.45	0.43

Being subsequently exposed to a game could also give rise to intertemporal strategies. For example, participants may consider a set of treatments over many periods as a single choice. These type of time effects are unlikely in this experiment for two reasons. First, because participants did non know in advance how many rounds they will play. They were invited to attend a social activity which included "games" prizes, and a "celebration" of the end of the household survey conducted by enumerators the previous week. This means that it would have been difficult to predict the number of rounds played. Moreover, the existing literature suggest that individuals exposed to different treatments may experience a so called restart effect. Andreoni (1988), found that the restart effects tend to reset contributions towards

[4] A F test of equality of coefficients in the regression of observed contributions on periods (clustered by experimental session) gives a $p-$value of 0.201 .

initial levels (not higher). This experiment was designed to enhance these type of effects by presenting each game as a separate one: the common pot and the expectations question were of different color, and the individual endowments (a set of token coins) were new each time.

The descriptive analysis above shows that the discussion and the leader treatment successfully enhance contributions to the public good, the leader treatment being somewhat more effective. There is little evidence, if any, of time effects affecting the observed choices.

3.6 Estimation Results

The behavioral model makes more explicit and flexible the role of expectations, it allows individuals to emphasize on conditional or unconditional cooperation depending on the experimental treatment. For example, a local leader might have the power to stimulate unconditional cooperation by convincing participants to cooperate regardless or what others do. Instead, discussing contribution choices with neighbors and friends might result in a collective agreement, mainly supported by conditional cooperation.

Table 3.3 contains simulated maximum likelihood estimates[5] of the random utility model in Section 3.4. Column (a) reports parameter values under the hypothesis of homogenous preferences. This specification assumes that conditional and unconditional cooperation are identical for all participants. An alternative interpretation is that these are the expected preferences of a participant randomly drawn from the sample.

Parameters are separated into groups. The first group corresponds to preferences for unconditional cooperation, the second group is associated to conditional cooperation, and the third group contains the covariance elements or nuisance parameters. Column (b) exploits the panel aspect of the data using the random coefficients approach presented in Subsection 3.4.2. In this specification, conditional and unconditional cooperation consists of a constant that is common to all participants and various heterogeneous components specific to individuals and local communities.

Unconditional Cooperation

In the model, α_i and γ are the two parameters associated with preferences for unconditional cooperation. α_i could be interpreted as an altruism parameter because it captures the utility of cooperating itself. Its common component shared by all participants (α_o) is 9.39 according to the homogenous preferences model in column (a) and 9.07 according to the more flexible model with heterogeneous preferences in column (b), both estimates are significant at 1%. These two estimates are greater than the pecuniary opportunity cost of not contributing, that is to say, the nine forgone tokens from investing in the private account. Moreover, the fraction of this preference that is associated to material possessions (α_{assets}) is negative and significant at 10%, and the fraction associated to age (α_{age}) is positive and significant at 1% level. This means that wealthier and younger participants have more egoistic preferences and thus, they are less willing to cooperate. This result corresponds well to the cultural patterns of the rural communities in Mali, where the elders are references of desired social behaviour. Not surprisingly, they often play the role of community counsellors or village chiefs. This result

[5]Standard errors in parenthesis are calculated using the BHHH method, which approximates the covariance matrix with the outer product of the gradient. Results were generated using Ox version 7.00 ©.

is also inline with other experimental studies using a representative sample of the population in Netherlands. Bellemare and Kröger (2007); Bellemare et al. (2008) and Bellemare et al. (2011) find that young and highly educated individuals have weaker social preferences.

There is no statistical difference between males and females with respect to the unconditional cooperation preferences, a result consistent with the preponderant evidence on gender effects in public goods experiments (Ledyard, 1994). With respect to the unobserved factors, both, individual unobserved characteristics and factors specific to each village influence α_i. Their estimated variations $\text{Var}(\eta_i^{\alpha}) = 0.73$ and $\text{Var}(v_k^{\alpha}) = 0.40$ are significant at 1%.

The estimates of γ are small but robust across specifications. This preference parameter is constant in the model because there is no variation in the group size within individuals. In column (a), γ is 0.014 and significant at 1% level. In the more flexible specification in column (b) its value is 0.013, but its significance is reduced to 15%. While the model with homogenous preferences contains a total of six parameters, the model with heterogeneous preferences requires estimating twenty-two parameters from the same variation. This loss in the degrees of freedom may harm the precision of the estimates. The estimated preferences associated with group size are in line with earlier findings (Isaac and Walker, 1988b), which suggest that the cooperation is weakly motivated by the number of participants. Even though in this experiment total social returns increase with group size, participants do not tend to contribute much more to the public good as the returns increase. This could be interpreted as an absence of efficiency concerns.

Conditional Cooperation

The second group of estimates is associated with conditional cooperation. In column (a) conditional cooperation θ_o is estimated to 0.42 and is significant at 1%. For an average participant, the utility of cooperating increases when expectations go from the lowest level (*none* of the group members will contribute) to the highest level (*all* group members will contribute). The estimates of the model with heterogeneous preferences in column (b) suggest that the simplifications imposed by the previous model can be misleading. First, common preferences for conditional cooperation are less important than suggested. The common component θ_o is estimated to 0.18 and it is not statistically significant at 10% level. Second, preferences for conditional cooperation are highly heterogeneous and their variation is mainly associated to unobserved factors specific to each participant. This idea is supported by the large variance of the time invariant component $\text{Var}(\eta_i^{\theta}) = 4.49$, which is significant at 1%. The role of unobserved cultural factors is less relevant in the case of unconditional cooperation, their variance $\text{Var}(v_k^{\theta}) = 0.37$ is significant only at 10% level.

Preferences associated to alternative beliefs are similar across specifications. The parameter θ^{few} is negative and weakly significant. When few group members are expected to contribute to the public good, participants behave as if they expected zero cooperation. This result follows the principle that individuals try to match the actions of the majority and is consistent with conditional cooperation. Incentives for cooperation start rising when participants expect half of the group to contribute. The value of θ^{half} is 0.22 or 0.24 depending on the specification, and both estimates are significant at 1%. The utility of cooperating increases even more when the majority of the group is expected to contribute to the public good. The estimates of θ^{many} are 0.35 and 0.43, significant at 1%. Unlike unconditional cooperation, conditional cooperation does not depend on observed individual characteristics.

Table 3.3 – Probit estimates

	homogenous preferences (a)	heterogenous preferences (b)
unconditional cooperation:		
common preferences α_o	9.3872*** (0.025)	9.3074*** (0.263)
assets: α_{asset}		-0.0865* (0.044)
age: α_{age}		0.0089*** (0.003)
male: α_{male}		0.0435 (0.054)
discussion: $\alpha_{\text{disc.}}$		0.1702*** (0.038)
leader: $\alpha_{\text{lead.}}$		0.3679*** (0.031)
group size: γ	0.0140*** (0.002)	0.0132 (0.015)
conditional cooperation:		
common preferences θ_o	0.4245*** (0.015)	0.1792 (0.192)
assets: θ_{assets}		0.0957* (0.051)
age: θ_{age}		-0.0037 (0.003)
male: θ_{male}		0.0371 (0.066)
discussion: $\theta_{\text{disc.}}$		0.1502*** (0.058)
leader: $\theta_{\text{lead.}}$		0.0901 (0.061)
few: θ^{few}	-0.1005* (0.059)	-0.0595 (0.100)
half: θ^{half}	0.2237*** (0.048)	0.2398*** (0.080)
many: θ^{many}	0.3527*** (0.042)	0.4267*** (0.062)
nuisance parameters:		
$\text{Var}(\eta_i^\alpha)$		0.7343*** (0.071)
$\text{Cov}(\eta_i^\alpha, \eta_i^\theta)$		-1.3769*** (0.263)
$\text{Var}(\eta_i^\theta)$		4.4895*** (1.623)
$\text{Var}(v_k^\alpha)$		0.3969*** (0.062)
$\text{Cov}(v_k^\alpha, v_k^\theta)$		-0.2537*** (0.087)
$\text{Var}(v_k^\theta)$		0.3753* (0.203)
log-likelihood	-33.484	-29.888
participants	2499	2499
observations	7074	7074
parameters	6	22

*** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1.

In summary, while unconditional cooperation is easily captured by common preferences shared by all participants, conditional cooperation is much more heterogeneous and sensitive to unobserved individual factors.

Treatment Effects

One of the advantages of estimating an structural model is that it shows additional information on how the discussion treatment and the leader treatment affect the cooperation choice. The estimates show that while the leader treatment enhances unconditional cooperation, the discussion treatment fosters a significant increase in conditional cooperation, making expectations more relevant.

The leader treatment increases public good provision by incentivizing unconditional cooperation. The parameter $\alpha_{\text{leader}} = 0.37$ and is significant at 1%. The presence of a local leader may be an effective way to increase awareness of optimal group behavior. Furthermore, the leader effect on unconditional cooperation is statistically zero ($\theta_{\text{leader}} = 0.09$), which means that the impact on participants' beliefs about other decision makers' remains unchanged. It is natural to think that a leader with particular characteristics or strengths may influence preferences for cooperation in a different way (Gächter et al., 2012; Guth et al., 2007; Bruttel and Fischbacher, 2013); however, in this experiment controlling for the characteristics of the leader does not alter the main results, nor it provides additional information on participants' choices.[6]

A second result from the model is that the discussion treatment promotes conditional cooperation. When participants are allowed to communicate, expectations about the actions of the group become a relevant factor in the cooperation choice. I find $\theta_{\text{discussion}} = 0.15$ and significant at 1%. This communication effect adds-up to the shared preferences for conditional cooperation θ_o, which result in common preference for conditional cooperation equal to 0.33. The discussion treatment has a moderate effect on unconditional cooperation, $\theta_{\text{discussion}} = 0.17$ and is significant at 1%.

Policy-wise, these results suggest that involving local leaders and promoting community discussion are both effective tools to incentivize cooperation. Nonetheless, community discussions may not be an adequate tool in a context of low expectations.

Nuisance Parameters

Even though covariance elements of the unobserved factors η_i and v_k are not of direct interest, they contain relevant information about conditional and unconditional cooperation. Obviously, column (a) is empty because the model with homogenous preferences does not account for these variations. In column (b), covariances of individual factors and village specific factors are negative and significant at 1%. This suggests that more altruistic individuals care less about the actions of the others ($\text{Cov}(\eta_i^{\alpha}, \eta_i^{\theta}) = -1.38$). The same relationship holds for the unobserved cultural factors specific to the villages but to a smaller extent ($\text{Cov}(v_i^{\alpha}, v_k^{\theta}) = -0.25$). It is important to notice that ignoring these covariances between conditional and unconditional cooperation may result in misestimation of the causal relation between expectations and observed choices.

[6] Regression results are available upon request.

3.7 Model Predictions

Predicting contribution probabilities has a dual purpose. First, estimated choice probabilities ensure that the model provides a good fit for the data. Second, estimates can be used to predict individual choices in hypothetical situations that are unlikely to be observed. The first section of Table 3.4 reports the proportion of cooperators observed across treatments and compares the experimental results with the average probabilities of cooperating predicted by the model. The similarities between observed and predicted outcomes suggest a good model fit. In particular, the predicted probabilities capture the increase in cooperation caused by the two experimental treatments.

Table 3.4 – Observed and predicted average contributions†

contributions	baseline treatment (a)	discussion treatment (b)	leader treatment (c)	over all sample (d)
observed	0.669	0.720	0.765	0.719
	(0.019)	(0.020)	(0.019)	(0.016)
predicted	0.668	0.723	0.766	0.720
	(0.005)	(0.006)	(0.006)	(0.005)
predicted average contributions under alternative beliefs				
none	0.560	0.550	0.606	0.572
	(0.003)	(0.003)	(0.003)	(0.003)
few	0.550	0.535	0.592	0.559
	(0.003)	(0.003)	(0.003)	(0.003)
half	0.610	0.624	0.675	0.637
	(0.003)	(0.003)	(0.003)	(0.003)
many	0.651	0.685	0.729	0.689
	(0.003)	(0.003)	(0.002)	(0.003)
all	0.766	0.838	0.863	0.823
	(0.002)	(0.002)	(0.002)	(0.002)

†Standard deviations in parenthesis clustered by village.

Structural parameters estimates can be used to predict participants' choices under hypothetical beliefs and obtain estimates of unobserved counterfactuals. In a pessimistic scenario in which participants expect *none* or *few* of the group members to contribute to the public good, the model predicts 55% to 56% cooperation. This proportion is very similar if not lower in the discussion treatment. A first clear message is that communication does not necessarily ameliorate social outcome and may even worsen it when expectations are weak. In fact, when only a *few* participants are expected to cooperate, the discussion treatment is predicted to drags down cooperation to 53%. Another possible interpretation is that in the discussion treatment expectations are more relevant, they account for 23.6% of the observed public good provision (compared to the observed 72% cooperation). In a context of low expectations, the presence of a local leader seems to be a more appropriate tool to promote cooperation. The leader is predicted to increase cooperation from 55% in the baseline to 60.6%. In general, model predictions corroborates the earlier finding that conditional cooperation plays a major role when communication is allowed.

3.8 Conclusions

This paper estimates a structural microeconometric model that separately identifies conditional and unconditional cooperation in a public goods game. The regression includes not only observed choices, but also information on participants' expectations about total public good provision. The model integrates a random coefficient approach to account for the potential heterogeneity in participants' preferences. While unconditional cooperation is easily captured by common preferences shared by all subjects, conditional cooperation is much more heterogeneous and depends on individual factors unobserved to the researcher.

I find that unconditional cooperation is sensitive to the presence of local leaders and to community discussions, and that the former is a more robust tool to enhance public goods provision. I also find that the efficiency of communication in promoting cooperation largely depends on expectations. This result may be of interest for policy purposes, because in particular social environments communication may not enhance public good provision and can even worsen the social outcome when expectations are negative. Nonetheless, community involvement can be an effective tool for inducing cooperative behavior in presence of positive expectations.

Finally, the results obtained in this research and the additional data available from the experiment open new questions that are left for future work. For instance, it would be interesting to investigate the role of social connections in the cooperation choice. The structural model used here assumes that each individual sees the rest of participants as a unique homogenous group. However, participants' perceptions of the rest of the group may depend on who is participating: close friends, extended family members, or detractors. There exists data on social networks within village households and this information could be used to measure the importance of peer effects in the cooperation decision. Furthermore, expectations are assumed to be exogenous through out the analysis. It would be interesting to instrument beliefs in order to test their potential endogeneity.

3.9 Bibliography

Andreoni, J. (1988). Why free ride? : Strategies and learning in public goods experiments. *Journal of Public Economics*, 37(3):291–304.

Andreoni, J. (1989). Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence. *Journal of Political Economy*, 97(6):1447–58.

Andreoni, J. (1990). Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving. *The Economic Journal*, 100(401):pp. 464–477.

Andreoni, J. (1995). Cooperation in Public-Goods Experiments: Kindness or Confusion? *American Economic Review*, 85(4):891–904.

Andreoni, J. and Miller, J. (2002). Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism. *Econometrica*, 70(2):737–753.

Armantier, O. and Treich, N. (2013). Eliciting beliefs: Proper scoring rules, incentives, stakes and hedging. *European Economic Review*, 62(C):17–40.

Becker, G. S. (1974). A Theory of Social Interactions. *Journal of Political Economy*, 82(6):1063–93.

Bellemare, C. and Kröger, S. (2007). On representative social capital. *European Economic Review*, 51(1):183–202.

Bellemare, C., Kröger, S., and van Soest, A. (2008). Measuring Inequity Aversion in a Heterogeneous Population Using Experimental Decisions and Subjective Probabilities. *Econometrica*, 76(4):815–839.

Bellemare, C., Kröger, S., and van Soest, A. (2011). Preferences, intentions, and expectation violations: A large-scale experiment with a representative subject pool. *Journal of Economic Behavior & Organization*, 78(3):349–365.

Blanco, M., Engelmann, D., Koch, A., and Normann, H.-T. (2010). Belief elicitation in experiments: is there a hedging problem? *Experimental Economics*, 13(4):412–438.

Bolton, G. E., Zwick, R., and Katok, E. (1998). Dictator game giving: Rules of fairness versus acts of kindness. *International Journal of Game Theory*, 27(2):269–299.

Bruttel, L. and Fischbacher, U. (2013). Taking the initiative. What characterizes leaders? *European Economic Review*, 64(C):147–168.

Cárdenas, J. C., Chong, A., and Ñopo, H. (2009). To What Extent Do Latin Americans Trust, Reciprocate, and Cooperate?: Evidence from Experiments in Six Latin American Countries. *Journal of LACEA Economia*.

Cardenas, J. C., Stranlund, J., and Willis, C. (2000). Local Environmental Control and Institutional Crowding-Out. *World Development*, 28(10):1719–1733.

Costa-Gomes, M. A. and Weizsäcker, G. (2008). Stated Beliefs and Play in Normal-Form Games. *Review of Economic Studies*, 75(3):729–762.

Croson, R. T. (2007). Theories Of Commitment, Altruism And Reciprocity: Evidence From Linear Public Goods Games. *Economic Inquiry*, 45(2):199–216.

Davis, D. and Holt, C. (1993). *Experimental Economics*. Princeton University Press.

Elizabeth, H., Kevin, M., Keith, S., and Vernon, S. (1994). Preferences, Property Rights, and Anonymity in Bargaining Games. *Games and Economic Behavior*, 7(3):346–380.

Fehr, E. and Schmidt, K. M. (1999). A Theory Of Fairness, Competition, And Cooperation. *The Quarterly Journal of Economics*, 114(3):817–868.

Fischbacher, U. and Gächter, S. (2010). Social Preferences, Beliefs, and the Dynamics of Free Riding in Public Goods Experiments. *American Economic Review*, 100(1):541–56.

Fischbacher, U., Gächterr, S., and Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters*, 71(3):397–404.

Gächter, S. (2007). Conditional cooperation: Behavioral regularities from the lab and the field and their policy implications. In Frey, B. and Stutzer, A., editors, *Economics and Psychology. A Promising New Cross-Disciplinary Field*, CESifo seminar series, pages 19–50. MIT Press.

Gächter, S., Nosenzo, D., Renner, E., and Sefton, M. (2012). Who Makes A Good Leader? Cooperativeness, Optimism, And Leading-By-Example. *Economic Inquiry*, 50(4):953–967.

Goeree, J. K., Holt, C. A., and Laury, S. K. (1999). Altruism and Noisy Behavior in One-Shot Public Goods Experiments. Virginia Economics Online Papers 331, University of Virginia, Department of Economics.

Guth, W., Levati, M. V., Sutter, M., and van der Heijden, E. (2007). Leading by example with and without exclusion power in voluntary contribution experiments. *Journal of Public Economics*, 91(5-6):1023–1042.

Harrison, G. W. and List, J. A. (2004). Field Experiments. *Journal of Economic Literature*, 42(4):1009–1055.

Isaac, R. M. and Walker, J. M. (1988a). Communication and Free-Riding Behavior: The Voluntary Contribution Mechanism. *Economic Inquiry*, 26(4):585–608.

Isaac, R. M. and Walker, J. M. (1988b). Group Size Effects in Public Goods Provision: The Voluntary Contributions Mechanism. *The Quarterly Journal of Economics*, 103(1):179–99.

Keser, C. and van Winden, F. (2000). Conditional Cooperation and Voluntary Contributions to Public Goods. *Scandinavian Journal of Economics*, 102(1):23–39.

Koukoumelis, A., Levati, M. V., and Weisser, J. (2012). Leading by words: A voluntary contribution experiment with one-way communication. *Journal of Economic Behavior & Organization*, 81(2):379–390.

Ledyard, J. O. (1994). Public Goods: A Survey of Experimental Research. Public Economics 9405003, EconWPA.

Manski, C. F. (2002). Identification of decision rules in experiments on simple games of proposal and response. *European Economic Review*, 46(4-5):880–891.

Manski, C. F. and Molinari, F. (2010). Rounding Probabilistic Expectations in Surveys. *Journal of Business & Economic Statistics*, 28(2):219–231.

Moxnes, E. and Heijden, E. v. d. (2000). The Effect of Leadership in a Public Bad Experiment. Discussion Paper 2000-102, Tilburg University, Center for Economic Research.

Offerman, T., Sonnemans, J., and Schram, A. (2001). Expectation Formation in Step-Level Public Good Games. *Economic Inquiry*, 39(2):250–69.

Palfrey, T. R. and Wang, S. W. (2009). On eliciting beliefs in strategic games. *Journal of Economic Behavior & Organization*, 71(2):98–109.

Pickering, A. J., Djebbari, H., Lopez, C., Coulibaly, M., and Alzua, M. L. (2016). Effect of a community-led sanitation intervention on child diarrhoea and child growth in rural Mali: a cluster-randomised controlled trial . *The Lancet Global Health*, 3(11):pp. 701–711.

Rabin, M. (1993). Incorporating Fairness into Game Theory and Economics. *The American Economic Review*, 83(5):pp. 1281–1302.

Rivas, M. F. and Sutter, M. (2011). The benefits of voluntary leadership in experimental public goods games. *Economics Letters*, 112(2):176–178.

Robert, F., L., H. J., E., S. N., and Martin, S. (1994). Fairness in Simple Bargaining Experiments. *Games and Economic Behavior*, 6(3):347–369.

Rutström, E. E. and Wilcox, N. T. (2009). Stated beliefs versus inferred beliefs: A methodological inquiry and experimental test. *Games and Economic Behavior*, 67(2):616–632.

Train, K. (2003). *Discrete Choice Methods With Simulation*. Discrete Choice Methods with Simulation. Cambridge University Press.

UNICEF, UNAIDS, and WHO (2011). Global HIV/AIDS response. Epidemic update and health sector progress towards universal access: progress report 2011. Report, UNICEF, UNAIDS and WHO, Geneva.

UNICEF and WHO (2012). Progress on Drinking Water and Sanitation: 2012 Update. Report, UNICEF and WHO, Geneva.

Van der Heijden, E., Potters, J., and Sefton, M. (2009). Hierarchy and opportunism in teams. *Journal of Economic Behavior & Organization*, 69(1):39–50.

Walker, T. F. (2011). Community involvement in public goods provision: evidence from a field experiment in Ghana. Working Paper.

WHO (2012). World Malaria Report 2012. Report, World Health Organization, Geneva.

Appendix

3.A Model

This section details the probabilities and the likelihood function of the model introduced in Section 3.4. The probability of observing a decisions c_{it} given some characteristics specific to the individualX_i, some subjective beliefs b_{it}, and some unobserved factors η_i and v_k

$$\Pr\left(c_{it}|X_i,\, b_{it},\, v_k,\, \eta_i\right) = \Pr\left(\Delta u^e_{it}\middle|X_{it},\, v_k,\, \eta_i\right)$$

$$= \quad \Pr\left(\varepsilon^u_{it} < [\alpha_i - 9 - \theta_i + \gamma(N_k - 10)]c_{it} + 2\,\theta_i\, b_{it}\middle|X_i,\, b_{it},\, v_k,\, \eta_i\right)$$

Since I assume normality of the error term, the probability is a univariate standard normal.

3.A.1 Likelihood

Suppose for a moment that the vector of individual characteristics $\eta_i = (\eta_i^\alpha,\, \eta_i^\theta)'$ and the vector of villages specific characteristics $v_k = (v_k^\alpha,\, v_k^\theta)'$ are observed. The likelihood of the choices of an individual is a function of the observed variables X_i, b_{it}, η_i, v_k and β, a vector containing all the parameters of the model:

$$\Pr\left(c_i|X_i,\, b_i,\, v_k,\, \eta_i;\, \beta\right) = \prod_t \Pr\left(c_{it}|X_i,\, b_{it},\, v_k,\, \eta_i\right). \tag{3.7}$$

For a given village k with characteristics v_k, we can calculate the probability of the observed choices $\mathbf{c}_k = (\mathbf{c}_1, \ldots, \mathbf{c}_{N_k})$ given a set of beliefs $\mathbf{b}_k = (\mathbf{b}_1, \ldots, \mathbf{b}_{N_k})$ by integrating out individual probabilities (3.7) over their bivariate distribution function f:

$$\Pr\left(\mathbf{c}_k|X_i,\, \mathbf{b}_k,\, v_k; \beta\right) = \prod_i^{N_k} \int \Pr\left(c_i|X_i,\, b_i,\, v_k,\, \eta_i;\, \beta\right) f(\eta_i)\, \mathrm{d}\eta_{i.} \tag{3.8}$$

To obtain the unconditional likelihood of all observations $\mathbf{c} = (\mathbf{c}_1, \ldots, \mathbf{c}_{121})$ across villages, we integrate again (3.8) over the two dimensional distribution of the village characteristics g:

$$\mathbf{L}(\mathbf{c}\,|X_i,\, b_{it}; \beta) = \prod_k^{121} \int \Pr\left(\mathbf{c}_k|X_i,\, \mathbf{b}_k,\, v_k; \beta\right) \quad g(v_k)\, \mathrm{d}v_{k.} \tag{3.9}$$

Moreover, f and g are assumed to be multivariate normal functions, which facilitates the approximation of the integrals by simulation methods (Train, 2003, Chap. 9).