He Said, She Said: The Impact of Gender and Marriage Perceptions on Self and Proxy Reporting of Labor

by Ervin Dervisevic and Markus Goldstein

He Said, She Said: The Impact of Gender and Marriage Perceptions on Self and Proxy Reporting of Labor

Ervin Dervisevic[1], Markus Goldstein[2]

[1 2] World Bank, 1818 H St NW, Washington, DC 20433, USA

Correspondence e-mail: edervisevic@worldbank.org

Abstract

Accurate estimation of agricultural employment and labor productivities are crucial in understanding opportunities for improvements in productivity and economic growth in developing countries. Available labor statistics mostly come from surveys in developing countries, which mostly rely on one person providing information about all household members. However, there are only a few empirical studies that provide a framework for understanding the potential advantages and disadvantages of using self and proxy reporters in developing countries. Using self and proxy reports from surveys in Ghana, we find significant differences in estimated labor productivities with very different implications in terms of policy-making. We find that differences in reporting are most likely due to gender and marriage satisfaction of both self and proxy reporters, so differences in reporting are due to bias and are not classical measurement errors.

Keywords

Gender Norms, Labor Productivity, Proxy Reporting, Self Reporting, Social Desirability Bias, Survey Methods

1 Introduction

Accurate information in surveys is vital in understanding labor market outcomes, and labor statistics in agriculture are particularly important given that three quarters of the poor in developing countries live in rural areas and most depend on agriculture for their livelihoods (Burch et al. 2008). Precise labor statistics help researchers and policy makers understand differences in labor productivity in agriculture and non-agricultural sectors, especially in terms of policies that may attempt to boost labor movements between different sectors of an economy.

When the standard surveys are conducted in developing countries, such as household budget surveys or income/consumption expenditure surveys, they mostly rely only on household heads to provide information about all household members. An alternative approach is taken within the Living Standards Measurement Study – Integrated Surveys on Agriculture (LSMS-ISA) initiative which attempts to conduct interviews with all household members above a certain age (Glewwe and Grosh 1999). The premise inherent in the LSMS-ISA approach to surveys is that self-reported information is more reliable than information provided by proxies. However, there is a gap in research that would support this logic, and there are only a few empirical studies that attempt to provide a framework for understanding the potential advantages and disadvantages of using these methods in developing countries.

Overall, most of the evidence on the differences, quality, and reliability of self/proxy reporting comes from the U.S. (Bardasi et al. 2010). There are several studies known to the authors which provide an analysis of differences in self and proxy reporting for adult labor outcomes in developing countries (Bardasi et al. 2010; Desiere and Costa 2019; Koolwal 2019; Kilic et al. 2020). However, neither of these studies examines the causes of differences in reporting and potential deliberate misreporting by respondents. Our analysis would fill in some of the gaps in development microeconomics by showing that gender and motivation matter in reporting using the available data from a survey conducted in Ghana in 1998.

Using the rich panel data from Ghana, we examine possible differences in marginal

labor productivity using self and proxy reports on labor hours. In an attempt to identify sources of differences in reports, we also examine the impact of gender and marriage perceptions on differences in self and proxy labor reporting. We expect male self respondents to underreport the labor they provided on their wives' plots. Such underreporting would be consistent with social desirability bias, which is a tendency for individuals to deny socially undesirable actions and exaggerate socially desirable ones (Chung and Monroe 2003). We expect men to be hesitant to admit helping their wives on their wives' plots due to prevailing gender norms in Ghana that are such that men are supposed to be providers for the family and not dependent on their wives (Quartey and Martin 2008). Also, in accordance with non-cooperative (or strategic) household models, we would expect a lower perceived quality of marriage to lead to significant differences in the reporting of both men and women, by both self and proxy reporters. A spouse with lower marriage satisfaction would have an incentive to engage in punitive misreporting which would manifest itself in an unwillingness to admit that help was provided to or by the spouse (Donni and Chiappori 2011).

To the best knowledge of the authors, there are no studies that examine the possibility of intentional misreporting of labor outcomes when conducting an analysis of self and proxy reports. Intentional misreporting is a particularly important issue, since it leads to biases which can invalidate studies using the data containing such biases. Our study attempts to demonstrate that data issues assumed to be classical measurement errors may be due to intentional misreporting by self and/or proxy reporters.

Section 2 of this paper provides the theoretical background and a short overview of empirical studies up to date that deal with self and proxy reporting. Section 3 provides the description of the data to be used and the empirical method. Section 4 will provide an overview of the results. Section 5 provides a discussion of our results and a conclusion.

2 Background

2.1 Measuring Labor Productivity

In the development context, structural change is necessary to shift labor away from less productive activities into more productive firms or activities. Such structural change would increase overall labor productivity which leads to higher rates of economic growth, especially if labor reallocation is between sectors with the largest differences in productivity. In poor countries, agriculture is typically the sector that employs the largest number of workers which means that the larger the productivity gap between agriculture and other sectors, the larger the opportunity to achieve productivity growth as labor shifts away from agriculture.

One of the major assumptions of development economics of the past few decades has been that developing countries are characterized by large productivity gaps between agriculture and non-agricultural sectors (Schaffner 2013). Dual economy models similar to Lewis (1954) see the economy composed of two sectors, the subsistence/agricultural sector and the modern/capitalist sector. The structural change that accompanies growth in such models is a shifting of labor out of the traditional sector into the modern sector, where that labor can be brought into productive employment (Kirkpatrick and Barrientos 2004). Therefore, higher labor productivity in non-agricultural sectors may imply that policies should be implemented to encourage agricultural workers to move to other sectors.

According to the data available, Sub-Saharan Africa has the highest share of the labor force employed in agriculture and the lowest labor productivity (WDI 2019). Within African countries, labor productivity in non-agricultural activities is six timer higher than in agriculture (Gollin et al. 2013). Other studies also show significant differences in labor productivity across sectors in Sub-Saharan Africa (McMillan and Harttgen 2014; Lele et al. 2013). However, the question remains if there is indeed higher productivity in non-agricultural sectors, and how high the difference may be. To answer these questions, there needs to be accurate assessment of labor productivities in all sectors.

A major source of information on labor outcomes and labor productivity in

Sub-Saharan Africa have been household surveys that ask only one respondent. If one respondent is not reliable in a systematic way and such household surveys mis-measure labor inputs as a key component of the productivity equation, estimates of labor productivity would be unreliable because the econometric models used in the subsequent analysis will contain biases.

2.2 Measurement Errors in Surveys

Measurement errors in surveys may be particularly problematic when errors are correlated in a causal way with explanatory variables in the analysis (Wooldridge 2009). Such biases may arise out of motivated misreporting which leads to respondents providing inaccurate information. In particular, when it comes to socially desirable (or undesirable) issues and social norms, there are sometimes reasons to suspect that respondents may lie. Theoretical considerations in sociology (Erving 1959) and psychology (Paulhus 1991) assert that in pursuing goals of social interaction, people will try to influence the opinion others have of them. Hence, people have a reason to lie, and they may do so in surveys by providing what they think are socially desirable answers (Krosnick and Presser 2010).

Social desirability may lead to patterns of underreporting (for socially undesirable behaviors and attitudes) or overreporting (for socially desirable attitudes and behaviors). The impact of social desirability on reporting may be particularly strong in issues related to gender, since we may expect that due to prevailing gender norms both self reporting women and proxy reporting men would be less likely to report perceived socially unacceptable behaviors of women.

In addition to social desirability, personal satisfaction or dissatisfaction with marriage or cohabitation conditions may provide a motivation for both men and women to provide inaccurate reports on matters related to their spouses. In accordance with non-cooperative strategic household models, we expect spouses to engage in retaliatory misreporting when they are not satisfied with their spouses and the state of their marriage (Donni and Chiappori 2011).Attitudes on marriage may arise from the behavior of partners or from power relations in marriage, and particularly dissatisfaction may provide

respondents with incentives to provide strategic information in surveys that they perceive to indirectly incur loss of utility in their partners.

2.3 Empirical Studies of Self vs. Proxy Reporting Differences

The question of the reliability of self and proxy reports is still an open one and still an unresolved issue. For the most part, research is limited to examining differences in self vs. proxy reporting, and what is being done is inadequate to conclude that self reporting is superior to proxy reporting (Moore 1988). This is possibly due to the fact that it would be difficult to compare self and proxy reports to the actual (or "true") values, and most often it is unfeasible to obtain objective measures of variables of interest.

In early examination of income, ownership reporting, and labor outcomes, results were often inconclusive on differences between self and proxy reporting (Hill 1987; O'Muircheartaigh 1991; Martin and Butcher 1982). Some studies found significant differences in reporting of job search (Boehm 1989), gross income (Dawe and Knight 1997), and employment and work hours (Lemaitre 1988). Press and Townsley (1998) and Achen and Stafford (2005) estimated that both men and women overreport time spent in housework in answers to direct questions and may do so for both self and proxy reporting. Conclusions in both studies are similar in that social desirability and social norms that require housework may be one of the causes of the overestimation of housework hours.

There are a few studies that deal with the issues of reliability and differences between self and proxy reporting in developing countries (Fisher et al. 2010; Bardasi et al. 2010; Dillon et al. 2012; Dammert and Galdo 2013; Desiere and Costa 2019; Koolwal 2019; Kilic et al. 2020). A major study which provides an analysis of differences in self and proxy reporting for adult labor outcomes in developing countries is Bardasi et al. (2010). In a survey experiment implemented in Tanzania, Bardasi et al. (2010) analyzed whether the characteristics of self and proxy respondents have an impact on labor statistics. They find that proxy respondents provide lower labor force participation and fewer working hours than self respondents. Several other studies find that there are significant differences in self and proxy reports of the employment rate (Desiere and Costa 2019; Kilic et al. 2020),

and of the desire to work and the hours worked (Koolwal 2019).

While differences in self vs. proxy reporting in surveys are examined in the existing literature, to the best knowledge of the authors, there are no studies that examine the possibility of intentional misreporting of labor outcomes when conducting an analysis of self and proxy reports.

It is our intention to provide an analysis that attempts to fill in some of these gaps in development microeconomics by using the rich panel data from surveys conducted in several waves in Ghana in 1998. In particular, we examine potential differences in estimated labor productivity that arise from reports on labor hours by self and proxy respondents. We examine the possibility that estimated labor productivities for men and women working on their spouses' plots is different when we use self and proxy reporting. Further, in order to understand where these differences come from, we look at the impact of gender and marriage perceptions on differences in self and proxy labor reporting by examining the impact of these variables separately on self and proxy reports and comparing those impacts.

Our hypothesis is that differences in labor hours reporting would lead to significant differences in implied labor productivity. The gender of self and proxy respondents would have an impact on how labor hours are reported, as we expect male self respondents to underreport the labor they provided on their wives' plots. We can foresee such an outcome because we expect men to be hesitant to admit helping their wives on their wives' plots, as gender norms in Ghana are such that men are supposed to be the providers for the family and not depend on their wives. Also, we would expect a lower perceived quality of marriage to lead to significant differences in the reporting of labor among both men and women, by both self and proxy reporters, because a person with lower marriage satisfaction would have an incentive not to admit that help was provided to or by the spouse. Dissatisfaction in marriage would provide incentives to men and women to engage in retaliatory misreporting by providing responses in surveys that withhold information which acknowledges the contributions of one's partner and reduces a partner's power in marriage.

3 Data and Methods

3.1 Ghanaian Household Data

In the late 1990's, Goldstein and Udry (1999) were involved in fieldwork collecting data from 435 farmers over a two year period in four village clusters in Southern Ghana. The survey was conducted in 15 rounds, each round 4 to 6 weeks apart. In each round, a core set of agricultural questionnaires were administered, complemented by a rotating set of modules.[1]

In all survey waves, men and women were interviewed separately to avoid the possibility of respondents' partners influencing responses. In this study, we will focus on rounds 10-15 administered from January to August 1998 and exclude the data obtained in previous rounds. In all rounds, the "Plot Activity Data" module was used and individuals were asked about work days and hours they did on the plots they manage and of the work of all others on these plots.

In all rounds the plot manager provided information on the work done by herself, her spouse, and anyone else. Also, the plot manager provided information on harvests and crops grown on each plot. However, only starting in round 10 was the "Work on Spouse's Plot Data" questionnaire used as part of the survey. As the name of the module implies, each participating individual was asked about the work he has provided on all the plots managed by his spouse. These two modules, administered jointly in rounds 10-15, enable us to compare self and proxy reports on work provided by spouses on the managers' plots.

In further text, we will use "worker report" as a label for self-reported work by managers' spouses and "worker characteristic" as a label for the personal specifics of a self reporter spouse. We will use "manager report" as a label for proxy report on a worker's work and "manager characteristic" as a label for the personal specifics of managers. An important limitation to our analysis is that interviews with managers and workers in each round were done at different times. This means comparing reports by rounds would be desirable but it is not feasible, and we need to collapse the data into

[1]Full information on the data and questionnaires, and the links to previous research that used this data are available at http://www.econ.yale.edu/ cru2/ghanadata.html.

reporting for total work hours and total production in all 6 rounds.

We will conduct our analysis in two parts. First, we will use harvest information and information on labor hours to compare marginal productivities of labor implied by workers' and managers' reporting. The dependent variable in this part of the analysis would be the reported value of crops harvested and those still in the field in the end of round 15, while the main explanatory variables of interest would be labor hours reported by managers and workers. In the second part of the analysis, we will examine what the factors contributing to differences in reporting would be. In this part of the analysis, reported labor hours would be the main dependent variables, while the main explanatory variables of interest would be the gender of the manager and reported perceived quality of marriage.

We assume that plot work done by the worker depends critically on the gender division of the manager-worker pair. We assume that in Ghanaian villages gender norms are a strong determinant of intra-household relations and, in accordance with the social desirability hypothesis outlined above, we hypothesize that gender division in the manager-worker pair would have a significant impact on reporting. We expect male managers to be less likely to report work done by their female workers and male workers to be less likely to report work done on female managers' plots. Such an occurrence would be due to the gender norms which are prevalent in rural Ghana, where men are supposed to be the primary providers for their families, while women would concentrate on domestic chores and small food production on their own plots (Quartey and Martin 2008).

Another variable of interest is a construct that we will label "perceived quality of marriage" (PQM). In round 10 of the survey, a "Marital Attitudes" questionnaire was administered which included the following three questions, where respondents were asked to provide answers on a five-point Likert scale:

- Overall, how do you get along with your husband/wife?
- On a scale one to five how much do you trust your husband/wife?

- On a scale one to five how does your husband/wife treat you?

To obtain a measure of internal consistency of answers and how correlated they are with each other, we used Cronbach's alpha (Cronbach 1951). The internal consistency for these questions is quite high for managers and workers, with Cronbach's alpha being above 0.9 for both. Since we are interested in the impact of individual perceptions of marriage on the reporting of labor outcomes, we use principle components analysis (PCA) to extract a single variable as a measure of perception from questions containing answers on various aspects of marriage.

Other variables to be included in the analysis are those that may have an impact value of the crop production or on the work performed by the worker. These include the ages of both the manager and the worker, their education expressed in years of education, number of children, and number of plots operated by each manager. Also, we include the managers' report of his own hours of labor on his plots, hours of labor by other family members (excluding the worker), and hours of labor by individuals with no relation to the managers (paid workers and/or friends). We had to exclude 24 polygamous marriages from our analysis, since it is not clear in the male managers' reports which of their wives did the work on the plots. Next, from the remaining observations we excluded all manager-worker pairs that have missing values for some core variable, such as the age of the manager or the worker.

In the first part of the analysis, we use only those manager-worker pairs where the manager reports positive values of harvests in rounds 10-15 and/or positive values for crops still in the fields at the end of round 15. For this part of the analysis, our sample is down to 256 manager-worker pairs. In the second part of the analysis, we exclude observations that have missing values of PQM, which leads to 216 manager-worker pairs.

3.2 Analysis of Labor Productivity

To understand how factors of production contribute to creation of output, we assume that output is a function of total number of labor hours L and physical capital K. Since we are dealing with different types of labor, we will use a production function with

labor decomposed into different types: workers' labor, managers' labor, family labor, and non-relative labor. Family labor includes all family members apart from manager and worker, while non-relative labor includes friends and paid labor.

In this study, we use a standard Cobb-Douglas functional form of production function with diminishing returns to labor and capital, and with several different types of labor:

$$P(L, K, A) = AK^{\alpha} L_m^{\beta_m} L_f^{\beta_f} L_n^{\beta_n} L_w^{\beta w} \tag{1}$$

where P is the monetary value of output for each manager, K is the value of capital inputs, A is a total factor productivity, while L_m, L_f, L_n, and L_w, are managers' labor hours, family labor hours, non-relative labor hours, and worker labor hours, respectively. Value of output P is the value of harvests reported by managers in rounds 10-15 and the value of crops in the fields at the end of round 15.

By taking the log of the function, we obtain the following functional form that can be used in a linear regression:

$$logP = logA + \alpha logK + \beta_m logL_m + \beta_f logL_f + \beta_n logL_n + \beta_w logL_w + \xi \tag{2}$$

where ξ is a normally distributed standard error. In such regression, coefficients on capital and labor would be marginal productivities of capital and labor, respectively. Replacing $logA$ with a set of personal and family variables X that could have an impact on production and productivity, we get for an individual i:

$$logP_i = \gamma + X_i\delta + \alpha logK_i + \beta_m logL_{mi} + \beta_f logL_{fi} + \beta_n logL_{ni} + \beta_w logL_{wi} + \xi_i \tag{3}$$

In this part of the analysis, the main explanatory variable of interest is log of workers' labor hours. We have workers' labor hours as reported by managers and workers, and we will conduct regressions using these different reports. The dependent variable and all explanatory variables apart from worker hours have the same values, so coefficients on workers' hours would be different for managers' and workers' reports only if those

reports on workers' labor hours are significantly different. In order to take into account the possibility that the difference in self and proxy reports may depend on the gender of the manager, all the regressions will be done using interactions of variables of interest and a "female manager" dummy that has a value of one when a manager is female.

3.3 Impact of Gender and Perceived Quality of Marriage

3.3.1 Decision to Work and Decision to Report

Preliminary analysis of the 216 manager-worker pairs to be used in this part of the analysis, shows that there are 33 cases where both manager and worker report zero hours of labor and 183 cases where at least one of them reports some positive number of hours. Hence, we can assume that 33 workers did no work on the managers' plots, and we label this as "zero-zero pairs". There are 118 cases where both the manager and his/her worker report some positive hours of work done by the worker, and we label this as "one-one pairs". There are also 23 cases where the manager reports positive hours of work, but the worker reports zero, and we label these cases as "one-zero pairs". There are 42 cases where the manager reports zero hours of work, while the worker reports some positive number, and we label these as "zero-one pairs".

The critical assumption on which this study rests is that work was actually done by workers whenever at least one person in the manager-worker pair reports positive hours. Hence, we assume that work was performed but not reported by either managers or workers in one-zero and zero-one pairs. Denoting the manager's report in household i as y_{im}, the worker's report as y_{iw}, and assuming that actual work has a real unobservable value y_i, we can express this assumption in the following way:

Assumption 1. *Whenever at least one person in the manager-worker pair reports positive hours of work, we assume that work was actually done, but may or may not have been reported by both the manager and the worker.*

$$y_{im} > 0 \vee y_{iw} > 0 \Longrightarrow y_i > 0; \qquad i = 1, ..., N, N = 216 \tag{4}$$

Under Assumption 1, households self-select into those in which workers did not ($y_i = 0$) and did perform work on managers' plots ($y_i > 0$). Among households in which we have $y_i > 0$, managers and workers self-select into not reporting work being done ($y_{im/w} = 0|y_i > 0$) and reporting some positive number of hours ($y_{im/w} > 0|y_i > 0$). In this sequential decision-making, the first decision is made by the manager-worker pair i who decides there is a need for worker's work, and we will refer to this decision as a decision to work. The second decision is made when the manager and worker in household i decide by themselves whether to report the work done by the latter, and we will refer to these as decisions to report.

Once managers and workers decide to report some positive number of hours worked by the worker, both reporters decide how many hours to report. We call the decision on how many hours to report "the amount decision". For the amount decision, we use standard ordinary least squares since the dependent variable is continuous. Therefore, with decision to work, decision to report, and the amount decision, we can use a following setup:

1. Decision to work:

$$Prob[y_i > 0] = \Phi(Z_i\delta) \tag{5a}$$

2. Decision to report positive work hours:

$$Prob[y_{im} > 0|y_i > 0|] = \Phi(Z_{im}\delta) \tag{6a}$$

$$Prob[y_{iw} > 0|y_i > 0|] = \Phi(Z_{iw}\delta) \tag{6b}$$

3. Number of work hours reported:

$$E[y_{im}|y_{im} > 0] = X_{im}\beta \tag{7a}$$

$$E[y_{iw}|y_{iw} > 0] = X_{iw}\beta \tag{7b}$$

Equations 5a, 6a, and 6b, are simple univariate probit models. The amount of hours equations 7a and 7b are ordinary least squares (OLS) regressions. The amount equations enable us to obtain expected values of $y_{im/w}$ conditional on $y_{im/w} > 0$. The analysis of unconditional expected value, that would take into account all parts of the model, is beyond the scope of this study. In this study, we are focused on the factors that lead to individuals making a decision whether to report labor being done and, when decision to report labor is made, differences in reporting between self (worker) and proxy (manager) respondents and how different personal and household characteristics impact their reports.

3.3.2 Comparison of Reports

We assume that there are essentially three underlying processes at work in our analysis, and there are observable and unobservable factors that influence each process. The first process is that there is some number of work hours invested by workers that we cannot observe, and we expect that several explanatory variables included in our analysis did have an impact on that work. The second process is that managers were or were not informed about the work that may have been performed by their workers and decide to report on that work in some way. Finally, workers are aware of the work they performed and decide how to report it. Some of the explanatory variables could have an impact on managers' reports, and the same or other variables may have had an impact on workers' reports.

In an attempt to distinguish between variables that have had an impact on work that was performed and those that have had an impact on reporting, we can compare $\hat{\beta}$ coefficients that appear on the same variables in two OLS regressions. This method of comparing coefficients has been extensively used in criminology (Brame et al. 1998). For example, Smith and Paternoster (1987) use a similar procedure to test whether the same causal factors are at work for male and female delinquency. In our study, comparison of coefficients across regressions is done in accordance with a procedure suggested by Clogg

et al. (1995) and Paternoster et al. (1998), and the z-score is computed as follows:

$$z = \frac{\hat{\beta}_m - \hat{\beta}_w}{\sqrt{\hat{SE}^2_{\beta_m} + \hat{SE}^2_{\beta_w}}} \tag{8}$$

For example, we can perform this test of equality of coefficients on the number of children in the household, so we test $H0 : \hat{\beta}_{m,children} = \hat{\beta}_{s,children}$. If those two coefficients turn out to be positive and significant in both regressions we can say that the number of children has a positive and significant impact on work performed by the worker. Also, if we cannot reject the null hypothesis of these coefficients being the same, it would imply that the number of children has no impact on how work was reported, because the equality of coefficients implies that the impact was only on the actual work or the respondent misreport in the same way. If, by contrast, we can reject the null hypothesis, we can say that the number of children has an impact on the reporting of workers' labor.

Since our goal is to be able to claim with some level of confidence that a particular variable is the same in both regressions, we will use p-values below 0.1 as indicative of statistically significant differences between $\hat{\beta}$'s.

4 Results

4.1 Differences in Labor Productivities

Summary statistics for the sample of 256 manager-worker pairs is given in Table 1. Total crop value is a sum of harvests in rounds 10-15 and current value of crops still in the fields at the end of round 15. We can see that there are 41.8% of female managers, and most of the work done in the fields is done by managers and non-relation (non-family) workers. In Table 2, we can see that there are 149 male managers and 107 female managers in our sample, and male managers report significantly higher values of crops produced, with higher values of labor and inputs invested in production of crops. In all categories of labor, male managers report more labor invested in their fields. Male managers also have more plots, are better educated, and are older.

Since we are looking at differences in labor outcomes implied by self and proxy reports, our main variable of interest is worker (aka spouse) hours, where we have self (worker) and proxy (manager) reports. All other labor variables, and individual and household characteristics, are primarily controls. We can see in Table 2 that proxy reporting male managers report worker hours that are lower than female self reporting workers, while female managers report higher workers' hours than their workers. For female managers, the difference in worker hours reports is statistically significant at $p < 0.05$ (not shown). But even with these differences in reporting between managers and workers, it seems that male managers use more workers' hours based on both their own and workers' reports.

The relevant part of production function estimations with log of crop value as a dependent variable is given in Table 3, while the full table is given in the Appendix. All

Table 1: Summary statistics for harvest data

	(1) Mean	(2) St.Dev.	(3) Min	(4) Max
Harvest and current value (log)	13.786	1.64	9.39	18.42
Female manager	0.418	0.49	0.00	1.00
Manager PQM	-0.010	1.00	-0.78	3.12
Worker PQM	-0.014	0.99	-0.89	3.03
Reported hours (logs):				
Manager hours	5.209	1.01	0.00	7.16
Family hours	2.266	2.15	0.00	7.47
No relation hours	3.035	2.44	0.00	8.29
Worker hours (manager report)	2.281	2.01	0.00	6.07
Worker hours (worker report)	2.239	2.18	0.00	7.58
Other variables:				
Inputs value (log)	8.492	4.73	0.00	16.36
Children	3.535	2.13	0.00	13.00
Manager plots	3.652	1.82	1.00	12.00
Manager educ.	7.033	4.77	0.00	16.00
Manager age	39.627	12.57	20.00	76.00
Worker educ.	6.848	4.73	0.00	16.00
Worker age	39.178	13.32	19.00	87.00
Village dummies:				
Village 1	0.285	0.45	0.00	1.00
Village 2	0.195	0.40	0.00	1.00
Village 3	0.281	0.45	0.00	1.00
Village 4	0.238	0.43	0.00	1.00
Observations		256		

Note: Summary statistics excludes missing values

regressions include a set of personal variables and village dummies. The basic setup is given in the first two columns. In order to check for the possibility that perceptions on marriage could have an impact on crop production, we include PQM variables in the fifth and sixth column. Missing values of PQM reduce the size of the sample to 198, but such a reduced sample does not significantly change our results with and without PQM in columns 3-6.

In Table 3, coefficients on interactions of female manager and all variables are given below the heading "Female manager × Variable". Sums of variables and interactions (and relevant p-values) are given under the heading "Female manager × Variable + Variable". Coefficients on variables are applicable to male managers, while coefficients

Table 2: Summary statistics for harvest data by gender of the manager

	(1)	(2)	(3)	(4)	(5)	(6)
	Male managers		Female managers		Male - Female	
	Mean	St.Dev.	Mean	St.Dev.	Diff.	T-stat
Harvest and current (log)	14.610	1.46	12.639	1.11	1.972***	(12.30)
Manager PQM	-0.372	0.71	0.495	1.13	-0.866***	(-6.88)
Worker PQM	0.226	1.03	-0.456	0.73	0.682***	(5.40)
Reported hours (log):						
Manager hours	5.462	0.92	4.857	1.03	0.605***	(4.85)
Family hours	2.460	2.08	1.995	2.22	0.466*	(1.70)
No relation hours	3.974	2.21	1.728	2.13	2.246***	(8.21)
Worker hours (m. report)	2.988	1.85	1.296	1.81	1.691***	(7.30)
Worker hours (w. report)	3.243	2.09	0.842	1.40	2.400***	(11.00)
Other variables:						
Inputs value (log)	10.604	3.42	5.549	4.75	5.055***	(9.40)
Children	3.430	2.09	3.682	2.17	-0.253	(-0.93)
Manager plots	4.443	1.54	2.551	1.59	1.892***	(9.50)
Manager educ.	8.178	4.58	5.439	4.60	2.739***	(4.71)
Manager age	41.191	12.27	37.449	12.72	3.743**	(2.36)
Worker educ.	5.564	4.52	8.636	4.43	-3.072***	(-5.42)
Worker age	35.721	11.60	43.991	14.11	-8.269***	(-4.97)
Village dummies:						
Village 1	0.275	0.45	0.299	0.46	-0.0239	(-0.41)
Village 2	0.201	0.40	0.187	0.39	0.0144	(0.29)
Village 3	0.282	0.45	0.280	0.45	0.00151	(0.03)
Village 4	0.242	0.43	0.234	0.43	0.00797	(0.15)
Observations	149		107		256	

Note: Summary statistics excludes missing values

* p<0.1, ** p<0.05, *** p<0.01

on sums are applicable to female managers. At the bottom of Table 3 we perform a t-test of coefficients on worker hours for manager and worker reports – the first and second t-tests provide tests of equality of worker hours coefficients for female workers and male workers, respectively.

In all specifications, coefficients on manager hours and no relation hours are positive and significant while family hours and worker hours are insignificant. Worker hours are negative but insignificant for both managers' proxy reporting and workers' self reporting. Hence, for male managers, marginal labor productivities are positive and significant for managers' hours and for hired and friends' labor. For female managers, marginal labor productivities are positive and significant for managers' hours, no relation hours, and for workers' hours using manager report. These results indicate that an increase in female workers' hours would not contribute to an increase in production for male managers, while an increase in male worker hours would contribute significantly to an increase in agricultural output for female managers. However, the latter result only applies to regressions with manager reports.

One of the most relevant results of our analysis is that workers' marginal labor productivities are different for male workers, as shown at the bottom of the table. In all three t-tests, null hypothesis is rejected at $p < 0.05$ for male workers. While it is possible that both measurement errors in reported values are correlated with real values, statistically significant differences in coefficients imply that female managers' or male workers' reported values (or both) are reported with a bias. Bias in reporting could arise from intentional misreporting by one or both reporters. If the bias is due to gender norms and/or marriage satisfaction, all coefficients could be biased and inconsistent in the analysis using managers' reports or the analysis using workers' reports (Wooldridge 2009).

Difference in estimated labor productivities is an important outcome with high policy relevance. With female managers' proxy reports, there is a positive and significant marginal productivity of male workers. Such positive marginal productivity of labor implies that workers' labor productivity gaps compared to other activities would be

smaller than if we relied on workers' self reporting. Moreover, with male workers' reports, there is no implication of output gain if using additional units of labor. Hence, policy implications of self and proxy reporting in our case would be quite different. In the former case additional units of male worker labor would lead to an increase in agricultural output, while in the latter case re-allocation of labor may be advisable since marginal productivity is zero.

A somewhat unexpected outcome of our analysis is that coefficients on male workers' hours are significantly higher for managers' reports than workers' reports. This result is surprising because worker hours reported by female managers are higher than hours reported by male workers, as shown in Table 2. Under the assumption of diminishing returns to scale, higher values of worker labor hours reported by female managers should lead to lower marginal productivity for managers' reports than workers' reports. It could be that the pattern of bias in worker hours reporting is such that it produces these unexpected results.

Tables 4-5 provide correlations for male and female managers of harvested and non-harvested crop values, and work hours. Correlations for worker hours with other variables among male managers are similar for managers' and workers' reports, with worker hours being positively correlated with values of harvests and values of crops at round 15. For female managers, correlations of worker hours with current crop values are insignificant but have the same signs. However, while not significant, worker hours correlations have different signs for correlations with harvest values. This difference in correlations may be indicative of a potential source of differences in reporting among female managers and male workers.

t could be that male workers are less likely to report their work on spouses' plots when work was done during a harvest season. In Ghana, an understanding of the farm roles of men and women is often "gendered", although these roles are not absolute. Land preparation and more physically demanding tasks are usually done by men, while planting is done by women and children with men assisting. Weeding and harvesting is primarily done by women (Raney et al. 2011). Also, there is most often division of duties such

Table 3: Regressions with log of harvest and log of reported work hours

	(1) Manager report	(2) Worker report	(3) Manager report	(4) Worker report	(5) Manager report	(6) Worker report
Female manager	0.2369	0.2997	-0.6387	-0.4750	-0.5408	-0.3355
	(0.804)	(0.811)	(1.056)	(1.054)	(1.094)	(1.088)
Manager hrs	0.3216***	0.3320***	0.2822***	0.3034***	0.3019***	0.3236***
	(0.089)	(0.089)	(0.101)	(0.099)	(0.103)	(0.101)
Family hrs	0.0123	0.0094	0.0083	0.0065	0.0147	0.0092
	(0.040)	(0.040)	(0.045)	(0.045)	(0.046)	(0.046)
No relation hrs	0.2398***	0.2417***	0.2182***	0.2140***	0.2113***	0.2040***
	(0.038)	(0.038)	(0.042)	(0.042)	(0.043)	(0.043)
Worker hrs	-0.0266	-0.0428	-0.0251	-0.0848*	-0.0387	-0.0978*
	(0.046)	(0.038)	(0.054)	(0.050)	(0.057)	(0.052)
Manager PQM					-0.1595	-0.1440
					(0.127)	(0.126)
Worker PQM					-0.0020	-0.0449
					(0.086)	(0.087)
Female manager × Variable:						
Manager hrs	0.0106	0.0329	0.1087	0.1269	0.0817	0.0960
	(0.134)	(0.135)	(0.175)	(0.175)	(0.179)	(0.178)
Family hrs	-0.0009	0.0034	-0.0117	0.0012	-0.0195	-0.0020
	(0.061)	(0.062)	(0.076)	(0.075)	(0.077)	(0.076)
No relation hrs	-0.0861	-0.0781	-0.0670	-0.0306	-0.0571	-0.0192
	(0.058)	(0.058)	(0.071)	(0.070)	(0.072)	(0.072)
Worker hrs	0.1397**	0.0314	0.1457*	0.0582	0.1604*	0.0760
	(0.067)	(0.073)	(0.082)	(0.087)	(0.085)	(0.089)
Manager PQM					0.1870	0.1436
					(0.172)	(0.170)
Worker PQM					0.0637	0.1305
					(0.182)	(0.182)
Observations	256	256	198	198	198	198
Female manager × Variable + Variable:						
Manager hrs	0.3322	0.3649	0.3909	0.4303	0.3836	0.4196
Manager hrs (p-val)	0.001	0.000	0.008	0.003	0.010	0.005
Family hrs	0.0114	0.0128	-0.0033	0.0077	-0.0048	0.0072
Family hrs (p-val)	0.809	0.791	0.957	0.901	0.940	0.909
No relation hrs	0.1537	0.1636	0.1512	0.1833	0.1542	0.1848
No relation hrs (p-val)	0.001	0.000	0.010	0.002	0.010	0.002
Worker hrs	0.1131	-0.0114	0.1206	-0.0266	0.1217	-0.0218
Worker hrs (p-val)	0.029	0.859	0.080	0.722	0.084	0.774
Manager PQM					0.0275	-0.0005
Manager PQM (p-val)	.	.	.	.	0.805	0.997
Worker PQM					0.0618	0.0856
Worker PQM (p-val)	.	.	.	.	0.693	0.586
H0: Manager report = Worker report (p-values):						
Female worker hrs	0.394		0.209		0.222	
Male worker hrs	0.019		0.023		0.032	

Household-clustered standard errors in parentheses.

* p<0.1 ** p<0.05 *** p<0.01

Table 4: Correlations for male managers with all values in logs

	Harvest and/or current values			**Worker work hrs**		**Other work hrs**		
	Harvest and current	Harvest	Current	Worker reports	Manager reports	Manager	Family	No relation
Harvest and current value	1.000							
Harvest value	0.478***	1.000						
Current value	0.729***	0.216***	1.000					
Worker hrs (worker reports)	-0.005	0.141*	0.108	1.000				
Worker hrs (manager reports)	0.045	0.265***	0.082	0.545***	1.000			
Manager hrs	0.426***	0.316***	0.453***	0.220***	0.279***	1.000		
Family hrs	0.184**	0.274***	0.142*	0.216***	0.287***	0.146*	1.000	
No relation hrs	0.585***	0.339***	0.353***	-0.047	-0.086	0.095	0.202**	1.000
Observations	149							

* p<0.1 ** p<0.05 *** p<0.01

Table 5: Correlations for female managers with all values in logs

	Harvest and/or current values			**Worker work hrs**		**Other work hrs**		
	Harvest and current	Harvest	Current	Worker reports	Manager reports	Manager	Family	No relation
Harvest and current value	1.000							
Harvest value	0.301***	1.000						
Current value	0.717***	-0.028	1.000					
Worker hrs (worker reports)	-0.022	-0.113	0.060	1.000				
Worker hrs (manager reports)	0.052	0.042	0.083	0.353***	1.000			
Manager hrs	0.501***	0.347***	0.355***	0.018	0.068	1.000		
Family hrs	0.132	0.334***	-0.007	-0.039	0.087	0.342***	1.000	
No relation hrs	0.377***	0.193**	0.305***	0.059	0.106	0.171*	0.225**	1.000
Observations	107							

* p<0.1 ** p<0.05 *** p<0.01

Table 6: Use of goods harvested

	Eaten	Sold	Other	Total
Yam	2	5	2	9
Other	1	8	1	10
Pepper	4	6	3	13
Plantain	15	8	2	25
Cocoyam	16	14	5	35
Firewood	17	6	17	40
Oil Palm	15	58	13	86
Cassava	34	51	131	216
Pineapple	1	135	149	285
Maize	282	122	24	428
Total	387	413	347	1147

Table 7: Harvests by male managers

	Eaten	Sold	Other	Total
Yam	1	4	2	7
Other	1	6	1	8
Pepper	2	6	3	11
Plantain	11	7	2	20
Cocoyam	11	9	5	25
Firewood	1	6	16	23
Oil Palm	13	55	13	81
Cassava	21	42	104	167
Pineapple	1	133	149	283
Maize	135	66	13	214
Total	197	334	308	839

Table 8: Harvests by female managers

	Eaten	Sold	Other	Total
Yam	1	1	0	2
Other	0	2	0	2
Pepper	2	0	0	2
Plantain	4	1	0	5
Cocoyam	5	5	0	10
Firewood	16	0	1	17
Oil Palm	2	3	0	5
Cassava	13	9	27	49
Pineapple	0	2	0	2
Maize	147	56	11	214
Total	190	79	39	308

Table 9: Harvests by gender of managers

	Female	Male	Total
Yam	2	7	9
Other	2	8	10
Pepper	2	11	13
Plantain	5	20	25
Cocoyam	10	25	35
Firewood	17	23	40
Oil Palm	5	81	86
Cassava	49	167	216
Pineapple	2	283	285
Maize	214	214	428
Total	308	839	1147

that men are in charge of cash crops such as pineapples and cocoa, while women work on smaller plots where they grow food crops such as maize and cassava (Drafor et al. 2005).

In tables 6-9, we can see the tabulation of the reported uses of agricultural produce from 1,147 plots. The "other" category is mostly crops not harvested and still in the fields, or used for seeds. We can see that pineapple is primarily a cash crop, and it is produced almost exclusively by men. Maize and cassava are both food and cash crops produced by both men and women. Female managers mostly use their produce for food, while men mostly sell their product. Therefore, the data on uses of produce in tables 6-9 provide an indication of gender roles according to which men primarily play a role of the providers in any household, while women focus on household chores and food crops. This is in accordance with prevalent gender norms in Ghana (Quartey and Martin 2008).

Given that men are supposed to be providers while women produce food, and gender roles are such that men do not provide support during harvests, it could be an explanation for a significant difference in reports. Men underreporting their work might lead to men reporting zero hours when they do some work to help their wives, or it could lead to men reporting positive but fewer work hours than the actual work they did.

In the next section, we conduct an analysis where we try to identify sources of differences in reporting. It is possible that these significant differences in marginal labor productivities are a consequence of intentional misreporting. As noted above, our main hypothesis in the next part of the analysis is that gender and marriage perceptions are major factors contributing to reporting differences.

4.2 Gender and Marriage Perceptions

In the previous section, analysis was done with logs of all major variables of interest. By taking the natural logs of all values, we did not need to correct for potential outliers. In this part of the analysis, we use nominal values of all reported labor hours, and preliminary analysis showed the presence of 8 extreme outliers in the reports of workers' labor, with 4 outliers in each manager and worker reports. The largest outlier was a worker report showing 1,962 hours of work, while the maximum manager report was 433 hours of work. Due to small sample size, these outliers had a major impact on the analysis, and all outliers were limited to an upper bound of the 98th percentile of the distribution.

Summary statistics for the sample of 216 manager-worker pairs is given in Table 10, and this sample is smaller than the one used previously because we use PQM in all specifications. Missing values for the PQM variable reduce the size of the sample. On average, workers report 18.84 total work hours more than managers' report for the workers. Table 11 shows the summary statistics of the manager-worker pairs by the gender of the manager. There are 129 manager-worker pairs with a male manager, and 87 pairs with a female manager. Male managers are older than female managers, better educated, manage more plots, report more own work hours on their plots, and report having more non-family labor working on their plots. An interesting fact is that females

Table 10: Summary statistics for all manager-worker pairs

	(1) Mean	(2) St.Dev.	(3) Min	(4) Max
Worker labor:				
Manager's report	46.977	67.23	0.00	271.00
Worker's report	65.818	98.04	0.00	419.00
Manager variables:				
Female manager	0.403	0.49	0.00	1.00
Manager age	39.414	12.79	20.00	76.00
Manager education	7.211	4.75	0.00	16.00
Manager's plots	3.625	1.83	1.00	12.00
Manager PQM	-0.000	1.00	-3.34	1.32
Worker variables:				
Worker age	38.463	12.88	19.00	78.00
Worker education	6.833	4.72	0.00	16.00
Worker PQM	-0.000	1.00	-3.16	0.85
Other variables:				
Children	3.472	2.13	0.00	13.00
Manager's work hrs	248.274	192.84	0.00	788.36
Family work hrs	54.091	104.98	0.00	504.50
Non-family work hrs	157.255	345.40	0.00	1757.50
Village 1	0.292	0.46	0.00	1.00
Village 2	0.181	0.39	0.00	1.00
Village 3	0.292	0.46	0.00	1.00
Village 4	0.236	0.43	0.00	1.00
Observations		216		

Note: Summary statistics excludes missing values

have lower estimates of PQM (perceived quality of marriage).

On average, male managers tend to report 38 less worker hours than their workers, while female managers tend to report 9 more work hours than their workers. However, summary statistics shown thus far include zero reports, and that may not show the full story. In Table A2 in the Appendix, t-tests of group means include only those observations in which positive worker labor hours were reported by the manager or the worker. When we exclude observations with zero reported work, there are 141 positive manager reports and 160 positive worker reports. Overall, managers tend to report fewer hours than their workers, and the difference is statistically significant at the p-value of 0.1. Female managers tend to report 30 more hours than their workers, and the difference is significant at the p-value of 0.05. Finally, male managers tend to report 36.03 fewer hours than their

Table 11: Summary statistics for all pairs with t-tests

	(1)	(2)	(3)	(4)	(5)	(6)
	Male managers		Female managers		Male - Female	
	Mean	St.Dev.	Mean	St.Dev.	Diff.	T-tests
Worker labor:						
Manager's report	63.15	73.90	22.99	46.86	40.16***	(4.89)
Worker's report	101.28	111.16	13.23	30.86	88.06***	(8.52)
Manager variables:						
Manager age	41.61	12.48	36.16	12.61	5.45***	(3.13)
Manager education	8.55	4.47	5.22	4.46	3.34***	(5.38)
Manager's plots	4.49	1.45	2.34	1.58	2.14***	(10.10)
Manager PQM	0.33	0.71	-0.48	1.16	0.81***	(5.79)
Worker variables:						
Worker age	35.58	11.79	42.74	13.30	-7.15***	(4.06)
Worker education	5.36	4.42	9.02	4.29	-3.67***	(6.08)
Worker PQM	-0.27	1.05	0.40	0.77	-0.66***	(5.36)
Other variables:						
Children	3.47	2.12	3.47	2.14	0.00	(0.01)
Manager's work hrs	312.14	183.33	153.57	166.59	158.58***	(6.59)
Family work hrs	61.99	111.75	42.39	93.46	19.60	(1.40)
Non-family work hrs	238.72	411.97	36.46	144.70	202.26***	(5.13)
Village 1	0.27	0.45	0.32	0.47	-0.05	(0.79)
Village 2	0.20	0.40	0.15	0.36	0.05	(1.00)
Village 3	0.29	0.45	0.30	0.46	-0.01	(0.19)
Village 4	0.24	0.43	0.23	0.42	0.01	(0.18)
Observations	129		87		216	

Note: summary statistics excludes missing values

* p<0.1 ** p<0.05 *** p<0.01

workers, and the difference is significant at the p-value of 0.01.

4.2.1 To Report or Not

As noted in Assumption 1, we assume that some work was performed by the worker when we have a positive report by either the manager or the worker. However, even though work was performed, some individuals have chosen not to report it.

The first-stage decision is the decision of the manager-worker pair as to whether the worker needs to do the work and whether there are means for her to do it. We have 33 cases where both manager and worker report no work and 183 cases where at least one of them reports positive hours of work. The second-stage decision is made by individuals in 183 manager-worker pairs in which at least one person reports positive work. Out of

these 183 pairs, we have 141 managers and 160 workers who provide positive reports. An overview of all reports is given in Table 12.

Based on the coding of reports in Table 12, decision to work and decision to report probit regressions are given in Table 13. As noted above, in sequential decision-making, the first decision in each household is a decision to work which is made by the manager-worker pair who decides whether there is a need for the worker's work. The decision to report is made when the managers and workers decide by themselves whether to report the work done by the latter. Both decisions are important parts of the decision-making process for self and proxy reports. In further analysis, both parts of sequential decision-making will be used to test the viability of Assumption 1, which states that work was done when at least one reporter states a positive number of worker's hours.

As shown in Table 13, the major factors that influence the decision to work are the number of plots, manager and worker age, and family work hours. Coefficients on manager's and worker's age indicate that older managers require more help and older workers are less likely to work, which is expected. Family work hours are negative and significant, possibly indicating that workers' help may not be needed when there are other household members that could and do work.

The decision to report regressions for managers and workers show the impact of different variables on the probability of reporting. We can see that PQM is positive and significant for managers and workers, showing that better perceptions of marriage

Table 12: Worker's Hours as Reported by Manager and Worker

		Manager reports positive		**Total**
		No	Yes	
Worker reports positive	No	33	23	56
	Yes	42	118	160
Total		75	141	216

increase the probability of reporting spousal labor. Also, we can see that both female managers and male workers are less likely to report labor provided by workers.

Our analysis relies on Assumption 1, which states that work was actually done by workers whenever at least one person in the manager-worker pair reports positive hours.

Table 13: Probit estimations of decision to work and decision to report positive

	(1) Decision to work	(2) Manager reports	(3) Manager reports	(4) Worker reports	(5) Worker reports
Female manager	-0.2359	-0.6652**	-0.8306***	-1.0826**	-0.7322*
	(0.489)	(0.278)	(0.261)	(0.483)	(0.443)
Manager age	0.0531***	-0.0127	-0.0113		
	(0.019)	(0.010)	(0.010)		
Manager education	-0.0260	0.0103	0.0107		
	(0.035)	(0.024)	(0.024)		
Manager PQM	0.0619	0.1866*			
	(0.134)	(0.113)			
Worker age	-0.0704***			-0.0064	-0.0043
	(0.019)			(0.013)	(0.013)
Worker education	-0.0231			0.0080	0.0098
	(0.034)			(0.039)	(0.036)
Worker PQM	0.2310			0.3721**	
	(0.161)			(0.167)	
Children	-0.0042	-0.0822	-0.0906*	0.0441	0.0298
	(0.088)	(0.053)	(0.054)	(0.059)	(0.057)
Manager's plots	0.4190***	-0.0243	-0.0362	0.1429	0.1259
	(0.162)	(0.082)	(0.082)	(0.118)	(0.115)
Manager's work hrs	0.0010	0.0017*	0.0016*	0.0007	0.0004
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Family work hrs	-0.0029**	-0.0002	-0.0003	0.0013	0.0017
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Non-family work hrs	0.0016	0.0000	0.0001	-0.0009**	-0.0007**
	(0.002)	(0.000)	(0.000)	(0.000)	(0.000)
Village 1	-0.1571	-0.4374*	-0.4647*	0.3321	0.2760
	(0.620)	(0.261)	(0.254)	(0.329)	(0.331)
Village 2	-2.6985***	-1.4780***	-1.4927***	0.8495	0.8790*
	(0.674)	(0.396)	(0.401)	(0.527)	(0.494)
Village 3	-2.6813***	0.0363	0.0259	-0.5167	-0.3837
	(0.643)	(0.324)	(0.326)	(0.414)	(0.386)
Constant	2.7265***	1.8578***	1.9735***	1.0746	0.9201
	(1.012)	(0.654)	(0.676)	(0.856)	(0.778)
Observations	216	183	183	183	183

Household-clustered standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

However, there is a possibility that there are cases in which work was done by the worker, but both a manager and a worker report zero work. Such cases of double false negatives could present an analytical problem – we should include these cases in the decision to report probits, but that is not possible since we cannot know which of the 33 cases are false negatives. Such false negatives may have a major impact on both decision to work and decision to report probit estimates, and we need to take into account this potential problem.

Unfortunately, a major model that can be used to control for misclassification in a binary dependent variable, developed by Hausman et al. (1998), cannot be used for small samples such as ours. Hence, in order to conduct a robustness check of our results by taking into account possible false negatives, we perform Monte Carlo (MC) simulations. In our simulations, we randomly pick a certain number M of observations and assume that work was actually done in those cases, but was not reported by either manager or worker. We perform 10,000 runs with such data and, in each of those runs, we conduct the estimation with the setup used in the Table 13.

We have 216 observations, with 33 "zero-zero" cases, 23 "one-zero" cases where positive work was reported by managers, 42 "zero-one" cases where positive work was reported by workers, and 118 "one-one" cases. We assume that there are M false negatives among 33 "zero-zero" cases, which means there are $N = 23 + 42 + 118 + M$ cases where work actually happened. When work is actually done, managers and workers report positive with probabilities p and q, respectively. Thus, we have: for "one-zero" cases, $p(1-q)N = 23$; for "zero-one" cases, $(1-p)qN = 42$; and for "one-one" cases, $pqN = 118$. Solving this system of three equations with three unknowns, we obtain the estimate of $M \approx 8$. Hence, in each simulation run we randomly pick 8 "zero-zero" observations, and assume that work was done in those cases but was not reported.

The result of MC regression simulations is given in Table 14. In the table, we report the average of regression coefficients obtained in simulations. In parentheses, we provide the percentages of regression simulations in which the null hypothesis for a given regression coefficient was rejected at the p-value of 0.1. For example, coefficients on manager's

plots in the first column were significant at 10% significance level in 84% of regressions. What is evident from this table is that coefficients of interest have the same signs as the coefficients in Table 13. Coefficients on the gender of the manager and on PQM are mostly significant in MC regressions for the manager and worker decision to report. Moreover, Figures A1-A5 in the Appendix provide distributions of regression coefficients of particular interest obtained in MC simulated regressions. In all figures, along with a mean value of regression coefficients obtained from MC simulations, "Reg. β" shows the value of the regression coefficient from Table 13. Hence, it seems that the results shown in Table 13 are quite robust to the possibility of a small number of false double negatives.

For the manager decision to report in columns 2 and 3 of the Table 13, significant factors in the decision to report are both PQM and the gender of the manager. This implies that the most important factors in the decision to report are gender and a personal view of the marriage; if the manager is not satisfied with the state of the marriage, he or she is likely not to admit that the worker helped. Also, the regression in column 4 shows that, if the worker is not satisfied with the state of the marriage, he or she is likely not to admit to helping the manager. Perhaps the more interesting result is the strong and negative coefficient on gender in both manager and worker decision regressions. It seems like female managers tend not to report their spouses' work, and male workers are less likely to admit doing the work and do not report helping on their wives' plots.

Due to a relatively small sample size, it may be problematic to make any conclusive arguments. However, our analysis thus far seems to imply that personal views of one's marriage have an impact on the reporting of labor for both managers and workers, meaning for both self and proxy reporters. Also, gender seems to matter in this case, and both males and female are less likely to report when husbands work for their wives.

The impact of marriage satisfaction may be understood as simply that humans are unlikely to admit positive contribution from and to others when those others are not as good to oneself as desired. Hence, marriage dissatisfaction in this case may reflect itself by reporters not attributing any positive contribution to their counterparts or not admitting that positive contribution was given. Impact of gender on worker (self) reporting may

be interpreted in a same way as above. It could be that we have social desirability bias in this, which leads to intentional misreporting. In agricultural communities in Ghana, gender norms may be such that wives perform domestic chores, while husbands are the providers for the family (Salm and Falola 2002). In such a context, it could be that

Table 14: Probit Estimation, Monte Carlo Simulation

	(1) Decision to work	(2) Manager reports	(3) Manager reports	(4) Worker reports	(5) Worker reports
Female manager	-0.2411	-0.7663	-0.9152	-1.1291	-0.7938
	(6.7%)	(99.8%)	(99.8%)	(99.3%)	(82.3%)
Manager age	0.0438	-0.0108	-0.0095		
	(78.7%)	(5.1%)	(1.4%)		
Manager education	-0.0255	0.0148	0.0147		
	(11.2%)	(0.0%)	(0.0%)		
Manager PQM	0.0391	0.1717			
	(5.2%)	(42.8%)			
Worker age	-0.0563			-0.0095	-0.0064
	(92.8%)			(0.9%)	(0.1%)
Worker education	-0.0211			0.0046	0.0075
	(9.8%)			(0.0%)	(0.0%)
Worker PQM	0.1747			0.3889	
	(25.9%)			(99.6%)	
Children	-0.005	-0.0788	-0.0854	0.0429	0.0288
	(2.4%)	(28.4%)	(40%)	(0.0%)	(0.0%)
Manager's plots	0.3429	-0.0049	-0.0171	0.1747	0.1595
	(84.3%)	(0.0%)	(0.0%)	(56.8%)	(42.3%)
Manager's work hrs	0.0009	0.0016	0.0015	0.0009	0.0005
	(11.7%)	(86.9%)	(76.6%)	(0.0%)	(0.0%)
Family work hrs	-0.0027	-0.0002	-0.0003	0.0008	0.0012
	(61.9%)	(0.0%)	(0.0%)	(0.8%)	(6.8%)
Non-family work hrs	0.0024	0.0001	0.0002	-0.0008	-0.0006
	(11.7%)	(0.0%)	(0.0%)	(99.8%)	(73.2%)
Village 1	-0.999	-0.4284	-0.4539	0.3218	0.2676
	(23.4%)	(48.2%)	(81.6%)	(0.0%)	(0.0%)
Village 2	-3.242	-1.5936	-1.6041	0.1696	0.1717
	(99.8%)	(99.8%)	(99.8%)	(0.9%)	(2.4%)
Village 3	-3.179	-0.2881	-0.2938	-0.7807	-0.6665
	(99.8%)	(0.3%)	(0.2%)	(97.3%)	(89.4%)
Constant	3.6422	1.7203	1.8256	1.1502	0.9461
	(94.2%)	(100.0%)	(100.0%)	(29.7%)	(19.2%)
Observations	216	191	191	191	191

Provided are average coefficients from 10,000 Monte Carlo simulations. In brackets percentage of cases in which null hypothesis was rejected for p¡0.1.

self-respecting men are not to help their wives in doing family-oriented work, so that both women and men do not want to report when that does happen. Thus, additional finding of this analysis is that both men and women may underreport work done by men on their wives' plots.

In Table 15 we test the assumption from the production function analysis that harvests by female managers may increase the likelihood of male workers underreporting their work. The basic setup in this case is the same as in the previous probit regressions, with an added dummy for reported harvests among female managers. Among 87 female managers which are tested in columns 2-6, there are 31 managers that report no harvest in rounds 10-15, and 56 managers that report at least one harvest. Female managers are more likely to report workers' labor when there is a harvest, while male workers seem to be less likely to report work when there is a harvest. This results further supports the conclusion reached in the previous section that gender roles have an impact on how harvest work is reported, since men are not supposed to help their spouses in harvest.

4.2.2 How Many Hours to Report

Once the decision to report work is made, we have 141 positive reports by managers and 160 positive reports by workers. There are 118 manager-worker pairs when both reported positive hours. In our analysis, we believe that errors are correlated for manager and worker reports, so we apply the seemingly unrelated ordinary least squares (OLS) regressions for positive reports with clustering at household level.

The results of OLS regressions for all positive reports are shown in the Table 16. We can see that coefficients on female managers in worker reports regressions are negative and significant, while positive and significant in at least one specification of manager reports in the third column. This implies that there is a significant difference in reporting based on managers' gender. Coefficients on manager plots and manager work hours imply that when there is a need for more work there is also more work being done or being reported, especially by managers.

As noted above, in an attempt to distinguish between variables that have had an

impact on work that was performed and those that have had an impact on reporting, we can compare $\hat{\beta}$ coefficients that appear on the same variables in OLS regressions. Tests of β coefficient equalities are given in Table 17. First, we can see that the difference

Table 15: Probit estimations with harvest dummy

	(1) Decision to work	(2) Manager reports	(3) Manager reports	(4) Worker reports	(5) Worker reports
Female manager	0.0402	-1.2615***	-1.4148***	-0.0226	0.3035
	(0.600)	(0.417)	(0.411)	(0.649)	(0.610)
Fem. manager × Harvest	-0.4569	0.8983**	0.8252**	-1.3441***	-1.3012***
	(0.434)	(0.390)	(0.370)	(0.501)	(0.488)
Manager age	0.0492***	-0.0128	-0.0114		
	(0.019)	(0.010)	(0.010)		
Manager education	-0.0239	0.0112	0.0106		
	(0.035)	(0.024)	(0.023)		
Manager PQM	0.0479	0.2212*			
	(0.136)	(0.119)			
Worker age	-0.0660***			-0.0069	-0.0044
	(0.019)			(0.013)	(0.012)
Worker education	-0.0206			0.0077	0.0094
	(0.033)			(0.040)	(0.037)
Worker PQM	0.2298			0.3864**	
	(0.158)			(0.174)	
Children	0.0116	-0.0856*	-0.0950*	0.0711	0.0562
	(0.087)	(0.051)	(0.052)	(0.056)	(0.056)
Manager's plots	0.4333***	-0.0594	-0.0692	0.1957	0.1797
	(0.165)	(0.080)	(0.080)	(0.133)	(0.132)
Manager's work hrs	0.0012	0.0013	0.0012	0.0015	0.0011
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Family work hrs	-0.0029**	0.0001	-0.0001	0.0013	0.0016
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
Non-family work hrs	0.0016	-0.0000	0.0001	-0.0010***	-0.0008**
	(0.002)	(0.000)	(0.000)	(0.000)	(0.000)
Village 1	-0.0491	-0.3938	-0.4185	0.3048	0.2355
	(0.591)	(0.264)	(0.256)	(0.322)	(0.327)
Village 2	-2.5922***	-1.4084***	-1.4198***	0.6326	0.6775
	(0.651)	(0.405)	(0.410)	(0.570)	(0.531)
Village 3	-2.5722***	0.0778	0.0674	-0.6436	-0.5028
	(0.613)	(0.328)	(0.324)	(0.426)	(0.396)
Constant	2.4582**	2.0634***	2.1871***	0.6700	0.5120
	(1.017)	(0.633)	(0.661)	(0.861)	(0.789)
Observations	216	183	183	183	183

Household-clustered standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

between β coefficients on gender is significant in both estimations, with the coefficient on "Female manager" being higher for managers in both estimations ("+/+" in second column of Table 17). Hence, it seems that male workers tend to underreport their own work, which is in accordance with the results obtained in the previous section. Manager's age is not highly significant in regressions and, at the significance level of 10%, the difference in reporting is insignificant. Manager's education and plots do not show significant differences in any specification, while worker age shows significant difference in both specifications.

Coefficients on PQM measures imply there is no significant impact of marriage perceptions, and differences are statistically insignificant for managers' and workers' PQM. Manager's work hours are significant and positive for managers, while being insignificant for workers. It could be that manager's own work makes him or her more knowledgeable about work done by the worker, and maybe even more appreciative of worker's effort. On the other hand, family work coefficients are negative for the manager, and insignificant for workers, which may be due to the manager mixing family work hours and worker work hours. Or managers may be providing more correct results, and the negative coefficient is due to a decreased need for workers' work.

Thus, based on the results shown in tables 16-17, it seems that the gender of the manager is a significant factor in reporting. On the other hand, PQM seems not to be a significant factor in this part of reporting, since all PQM coefficients are insignificant and differences in coefficients are also insignificant.

Looking at the extensive margin of reporting shown in probit estimations in tables 13-15, it seems that gender matter and male workers tend to report zero hours even when they have actually done the work to support their spouses. Also, in all probit regressions on decision to report, there is strong indication that higher marriage satisfaction of both workers and managers leads to higher probability of reporting of the work that was done.

Therefore, while gender matters at both the intensive and extensive margins of reporting, PQM seems to matter only for the latter. Our analysis thus implies there is a significant level of intentional misreporting, both among self (workers) and proxy

(managers) respondents, and misreporting may be particularly problematic when unequal gender norms and subsequent social desirability bias may be present.

Table 16: OLS regressions for positive reports

	(1) Manager reports	(2) Worker reports	(3) Manager reports	(4) Worker reports
Female manager	37.141	-95.556***	62.802***	-89.035***
	(23.609)	(30.367)	(23.717)	(33.511)
Manager age	0.8409	-0.7715	1.4901*	-0.1244
	(0.790)	(1.136)	(0.845)	(1.308)
Manager education	0.6826	1.5119	-0.9887	1.1013
	(1.220)	(2.007)	(1.273)	(2.086)
Manager PQM	9.7837	10.528	9.5291	5.8708
	(6.513)	(7.275)	(8.381)	(12.129)
Worker age	-1.0272	2.2956*	-1.7129**	2.7585*
	(0.843)	(1.296)	(0.755)	(1.477)
Worker education	-3.9465***	-2.4358	-4.7838***	-2.3526
	(1.309)	(2.144)	(1.263)	(2.268)
Worker PQM	-0.3045	7.6840	-0.2192	-1.0085
	(5.151)	(6.470)	(5.343)	(8.630)
Children	1.5442	1.4194	1.7922	7.8289*
	(2.844)	(3.527)	(3.038)	(4.228)
Manager's plots	6.4306	6.2660	13.017***	11.793*
	(4.561)	(5.797)	(4.684)	(7.165)
Manager's work hrs	0.1136***	-0.01709	0.1234***	-0.01239
	(0.041)	(0.046)	(0.042)	(0.052)
Family work hrs	-0.1483**	0.008366	-0.2242***	-0.01209
	(0.062)	(0.092)	(0.058)	(0.125)
Non-family work hrs	-0.03633***	-0.05949***	-0.03317**	-0.07380***
	(0.011)	(0.016)	(0.013)	(0.021)
Village 1	-47.528***	-77.595***	-49.036***	-89.610***
	(13.806)	(16.024)	(15.343)	(18.214)
Village 2	-88.032***	-100.80***	-101.03***	-122.55***
	(15.995)	(17.476)	(15.844)	(21.959)
Village 3	-25.512	-51.077**	-25.167	-60.790**
	(16.842)	(24.472)	(18.379)	(25.520)
Constant	70.254**	103.34**	62.945*	34.426
	(33.953)	(42.620)	(33.375)	(47.755)
Observations	183	183	118	118

Household-clustered standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 17: Comparison of coefficients in OLS regressions, all positive reports

H0: $\beta_{manager} - \beta_{worker} = 0$		P-value	P-value
Female manager	+/+	0.0003	0.0001
Manager age	+/+	0.1220	0.1499
Manager education	-/-	0.3620	0.1963
Manager PQM	-/+	0.4696	0.4020
Worker age	-/-	0.0158	0.0035
Worker education	-/-	0.2738	0.1745
Worker PQM	-/+	0.1670	0.4690
Children	+/-	0.4890	0.1231
Manager's plots	+/+	0.4911	0.4431
Manager's work hrs	+/+	0.0172	0.0206
Family work hrs	-/-	0.0789	0.0615
Non-family work hrs	+/+	0.1199	0.0482
Village 1	+/+	0.0776	0.0442
Village 2	+/+	0.2950	0.2134
Village 3	+/+	0.1947	0.1287
Constant	-/+	0.2719	0.3122
Observations		183	118

Note: Second column of the table provides the sign of the difference in coefficients, $\beta_{manager} - \beta_{worker}$.

5 Conclusion

In this study, we have made a concentrated attempt to identify factors that may lead to systematic differences in self and proxy reporting on labor outcomes. In an effort to do that, we have used an unusual setup of surveys that were done in Ghana in the second half of the 1990's. Using the data from Ghana, which contains reports on worker labor provided by both plot managers and workers, we were able to compare differences in reports provided by both.

An important finding of this study is that there are significant differences in estimated marginal labor productivities using manager and worker reports. When using self reports by male workers, it seems that we would be underestimating the hours they spend providing labor on their spouses' plots. We could also be underestimating male workers' labor hours when using female managers' reports, albeit to a smaller extent. In both cases, measurement error could be significant enough to invalidate the estimates of labor productivity. Estimates of labor productivity may be especially problematic since we

have indications of intentional misreporting by self or proxy reporters, or even by both.

Our results have particular policy significance in the context of Sub-Saharan Africa, where work on household farms remains a dominant economic activity in rural areas. Much of the workforce supplied in industry and services sectors includes processing and trade in agricultural commodities for consumers whose income is dominated by the agricultural sector activities. Moreover, non-agricultural activities involving rural households are closely linked to agriculture. In order to create and implement policies that would use these links and promote growth-enhancing structural changes, we need reliable information on current agricultural activities. As McCullough (2015) points out, if we are underestimating labor supply, the implication is that current estimates of the productivity of labor may overestimate true productivity of agricultural labor relative to other sectors. Results we obtain for male workers productivity also imply that it is possible to underestimate labor productivity when analysis uses self reporting by men. Analysis based on such findings would underestimate potential for improvements in women's opportunities and farm-based businesses.

Our results indicate that both gender and marriage satisfaction have an impact on reporting at the extensive margin, when managers and workers decide whether to report work being done. On the other hand, only gender seems to be a significant factor at the intensive margin when managers and workers make a decision on how many working hours to report. It could be that both social desirability bias and strategic reporting have an impact at the extensive margin when respondents make a decision whether to report any work being done by them or by their spouses, while social desirability plays a role when deciding how many hours of work to report.

However, due to the relatively small sample size, it is very hard to make any strong arguments about factors that may or may not influence reporting on labor. It is also questionable if our analysis is applicable outside of Ghana, in the context of other Sub-Saharan countries, or even other West African countries. With those caveats in mind, we can say that our data and this analysis do indicate that there are differences in reporting, and gender and marriage perception have an impact on reporting. Moreover,

if our results are seen as indicative of anything, our analysis implies there is a significant level of intentional misreporting of labor outcomes, both among self (spouses) and proxy (managers) respondents, and that misreporting may be particularly problematic when unequal gender norms and subsequent social desirability bias may be present.

Therefore, our results indicate the possibility that surveys being done in Ghana (and elsewhere) these days could be providing data that underestimate labor hours in agriculture. In our modest attempt, this study has demonstrated methods that can be used to analyze similar data and may provide future researchers with a framework on which to build their analyses.

References

Achen, A. C. and Stafford, F. P. (2005). Data quality of housework hours in the Panel Study of Income Dynamics: Who really does the dishes. Technical report, University of Michigan, Ann Arbor, MI.

Bardasi, E., Beegle, K., Dillon, A., and Serneels, P. M. (2010). Do labor statistics depend on how and to whom the questions are asked? Results from a survey experiment in Tanzania. Technical report, The World Bank, Washington, DC.

Boehm, L. M. (1989). Reliability of proxy response in the current population survey. In *Proceedings of the Survey Research Methods Section, American Statistical Association*.

Brame, R., Paternoster, R., Mazerolle, P., and Piquero, A. (1998). Testing for the equality of maximum-likelihood regression coefficients between two independent equations. *Journal of Quantitative Criminology*, 14(3):245–261.

Burch, D., Lawrence, G., Green, G. P., Ichijo, K., Nonaka, I., Pimentel, M., Bower, J., Gilbert, C., Couto Filho, V., Flavio, L., and others (2008). World Development Report 2008: Agriculture for Development. Technical report, The World Bank, Washington, DC.

Chung, J. and Monroe, G. S. (2003). Exploring social desirability bias. *Journal of Business Ethics*, 44(4):291–302.

Clogg, C. C., Petkova, E., and Haritou, A. (1995). Statistical methods for comparing regression coefficients between models. *American Journal of Sociology*, pages 1261–1293.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3):297–334.

Dammert, A. C. and Galdo, J. (2013). Child labor variation by type of respondent: evidence from a large-scale study. *World Development*, 51:207–220.

Dawe, F. and Knight, I. (1997). A study of proxy response on the Labour Force Survey. *Survey Methodology Bulletin*, pages 30–36.

Desiere, S. and Costa, V. (2019). *Employment Data in Household Surveys: Taking Stock, Looking Ahead*. The World Bank.

Dillon, A., Bardasi, E., Beegle, K., and Serneels, P. (2012). Explaining variation in child labor statistics. *Journal of Development Economics*, 98(1):136–147.

Donni, O. and Chiappori, P.-A. (2011). Nonunitary models of household behavior: a survey of the literature. In *Household economic behaviors*, pages 1–40. Springer.

Drafor, I., Kunze, D., and Al-Hassan, R. (2005). Gender roles in farming systems: An overview using cases from ghana.

Erving, G. (1959). *The presentation of self in everyday life*. Anchor Books, Garden City, NY.

Fisher, M., Reimer, J. J., and Carr, E. R. (2010). Who should be interviewed in surveys of household income? *World Development*, 38(7):966–973.

Glewwe, P. and Grosh, M. (1999). Designing household survey questionnaires for developing countries: Lesson from 10 years of LSMS experience. Technical report, The World Bank, Washington, DC.

Goldstein, M. and Udry, C. (1999). Agricultural Innovation and Resource Management in Ghana. Technical Report MP17, IFPRI, Washington, DC.

Gollin, D., Lagakos, D., and Waugh, M. E. (2013). The agricultural productivity gap. *The Quarterly Journal of Economics*, 129(2):939–993.

Hausman, J. A., Abrevaya, J., and Scott-Morton, F. M. (1998). Misclassification of the dependent variable in a discrete-response setting. *Journal of econometrics*, 87(2):239–269.

Hill, D. (1987). Response errors in labor surveys: Comparisons of self and proxy. In *Proceedings of the Bureau of Census, Third Annual Research Conference*. US Department of Commerce, Bureau of the Census.

Kilic, T., Van den Broeck, G., Koolwal, G., and Moylan, H. (2020). Are you being asked? impacts of respondent selection on measuring employment.

Kirkpatrick, C. and Barrientos, A. (2004). The lewis model after 50 years. *The Manchester School*, 72(6):679–690.

Koolwal, G. B. (2019). *Improving the Measurement of Rural Women's Employment: Global Momentum and Survey Research Priorities*. The World Bank.

Krosnick, J. A. and Presser, S. (2010). Question and questionnaire design. In Marsden, P. V. and Wright, J. D., editors, *Handbook of survey research*, pages 263–314. Emerald Group Publishing Ltd, Bingley, UK, 2 edition.

Lele, U., Agarwal, M., and Goswami, S. (2013). Lessons of the global structural transformation experience for the east african community. In *East African Community International Symposium and Exhibition on Agriculture, Kampala, Uganda*.

Lemaitre, G. (1988). A look at response errors in the Labour Force Survey. *Canadian Journal of Statistics*, 16(S1):127–141.

Lewis, W. A. (1954). Economic development with unlimited supplies of labour. *The Manchester School of Social Studies*, 22(2):139–191.

Martin, J. and Butcher, B. (1982). The Quality of Proxy Information-Some Results from a Large-scale Study. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 31(4):293–319.

McCullough, E. B. (2015). Labor productivity and employment gaps in Sub-Saharan Africa. World Bank Policy Research Working Paper, The World Bank, Washington, DC.

McMillan, M. S. and Harttgen, K. (2014). What is driving the'african growth miracle'? Technical report, National Bureau of Economic Research.

Moore, J. C. (1988). Self/proxy response status and survey response quality. *Journal of Official Statistics*, 4(2):155–172.

O'Muircheartaigh, C. (1991). Simple response variance: Estimation and determinants. In Biemer, P. P., Groves, R. M., Lyberg, L. E., Mathiowetz, N. A., and Sudman, S., editors, *Measurement errors in surveys*, pages 551–574. Jon Wiley and Sons.

Paternoster, R., Brame, R., Mazerolle, P., and Piquero, A. (1998). Using the correct statistical test for the equality of regression coefficients. *Criminology*, 36(4):859–866.

Paulhus, D. L. (1991). Measurement and control of response bias. In Robinson, J. P., Shaver, P. R., and Wrightsman, L. S., editors, *Measures of personality and social psychological attitudes*, volume 1, pages 17–59. Academic Press, Cambridge, MA.

Press, J. E. and Townsley, E. (1998). Wives' and husbands housework reporting: gender, class and social desirability. *Gender & society*, 12:188–218.

Quartey, Q. Q. and Martin, A. (2008). Gender and diversity report: Ghana. Technical report.

Raney, T., Anríquez, G., Croppenstedt, A., Gerosa, S., Lowder, S. K., Matuschke, I., and Skoet, J. (2011). The role of women in agriculture.

Salm, S. J. and Falola, T. (2002). *Culture and customs of Ghana*. Greenwood Publishing Group, Westport, CT.

Schaffner, J. (2013). *Development economics: Theory, empirical research, and policy analysis*. Wiley.

Smith, D. A. and Paternoster, R. (1987). The gender gap in theories of deviance: Issues and evidence. *Journal of Research in Crime and Delinquency*, 24(2):140–172.

WDI (2019). *World Development Indicators 2019*. The World Bank.

Wooldridge, J. (2009). *Introductory econometrics: A modern approach*. South-Western, Nashville, TN, 4th ed. edition.

Appendix: All Tables and Figures

Table A1: Regressions with log of harvest and log of reported work hours

	(1) Manager report	(2) Worker report	(3) Manager report	(4) Worker report	(5) Manager report	(6) Worker report
Female manager	0.2369	0.2997	-0.6387	-0.4750	-0.5408	-0.3355
	(0.804)	(0.811)	(1.056)	(1.054)	(1.094)	(1.088)
Manager hrs	0.3216***	0.3320***	0.2822***	0.3034***	0.3019***	0.3236***
	(0.089)	(0.089)	(0.101)	(0.099)	(0.103)	(0.101)
Family hrs	0.0123	0.0094	0.0083	0.0065	0.0147	0.0092
	(0.040)	(0.040)	(0.045)	(0.045)	(0.046)	(0.046)
No relation hrs	0.2398***	0.2417***	0.2182***	0.2140***	0.2113***	0.2040***
	(0.038)	(0.038)	(0.042)	(0.042)	(0.043)	(0.043)
Worker hrs	-0.0266	-0.0428	-0.0251	-0.0848*	-0.0387	-0.0978*
	(0.046)	(0.038)	(0.054)	(0.050)	(0.057)	(0.052)
Inputs value	0.0735***	0.0723***	0.0653**	0.0667**	0.0689**	0.0700**
	(0.025)	(0.025)	(0.027)	(0.027)	(0.028)	(0.027)
Children	-0.0081	-0.0047	0.0207	0.0221	0.0209	0.0246
	(0.035)	(0.036)	(0.041)	(0.040)	(0.041)	(0.041)
Manager plots	0.2681***	0.2753***	0.2433***	0.2581***	0.2313***	0.2502***
	(0.056)	(0.057)	(0.066)	(0.067)	(0.068)	(0.068)
Manager educ.	0.0298*	0.0319*	0.0348*	0.0332*	0.0356*	0.0332
	(0.017)	(0.017)	(0.020)	(0.020)	(0.020)	(0.020)
Worker educ.	-0.0015	-0.0054	0.0048	0.0050	0.0029	0.0030
	(0.019)	(0.019)	(0.023)	(0.023)	(0.023)	(0.023)
Manager age	0.0209**	0.0213**	0.0092	0.0096	0.0113	0.0113
	(0.010)	(0.010)	(0.013)	(0.013)	(0.013)	(0.013)
Worker age	-0.0129	-0.0134	-0.0037	-0.0029	-0.0069	-0.0059
	(0.011)	(0.011)	(0.014)	(0.014)	(0.015)	(0.015)
Manager PQM					-0.1595	-0.1440
					(0.127)	(0.126)
Worker PQM					-0.0020	-0.0449
					(0.086)	(0.087)
Female manager × Variable:						
Manager hrs	0.0106	0.0329	0.1087	0.1269	0.0817	0.0960
	(0.134)	(0.135)	(0.175)	(0.175)	(0.179)	(0.178)
Family hrs	-0.0009	0.0034	-0.0117	0.0012	-0.0195	-0.0020
	(0.061)	(0.062)	(0.076)	(0.075)	(0.077)	(0.076)
No relation hrs	-0.0861	-0.0781	-0.0670	-0.0306	-0.0571	-0.0192
	(0.058)	(0.058)	(0.071)	(0.070)	(0.072)	(0.072)
Worker hrs	0.1397**	0.0314	0.1457*	0.0582	0.1604*	0.0760
	(0.067)	(0.073)	(0.082)	(0.087)	(0.085)	(0.089)
Inputs value	-0.0855***	-0.0805**	-0.0773**	-0.0683*	-0.0788**	-0.0684*
	(0.032)	(0.032)	(0.038)	(0.037)	(0.039)	(0.038)
Children	0.0147	0.0089	-0.0382	-0.0416	-0.0418	-0.0489
	(0.056)	(0.056)	(0.068)	(0.068)	(0.069)	(0.069)
Manager plots	-0.0858	-0.0999	-0.0793	-0.1218	-0.0750	-0.1151

	(0.090)	(0.091)	(0.113)	(0.112)	(0.117)	(0.116)
Manager educ.	-0.0160	-0.0145	-0.0236	-0.0145	-0.0248	-0.0152
	(0.028)	(0.029)	(0.037)	(0.037)	(0.038)	(0.037)
Worker educ.	0.0084	0.0162	0.0092	0.0029	0.0113	0.0028
	(0.029)	(0.029)	(0.037)	(0.038)	(0.038)	(0.039)
Manager age	-0.0365*	-0.0418**	-0.0373	-0.0425	-0.0386	-0.0432
	(0.020)	(0.020)	(0.027)	(0.027)	(0.027)	(0.027)
Worker age	0.0360**	0.0382**	0.0408	0.0402	0.0442*	0.0426*
	(0.018)	(0.019)	(0.025)	(0.025)	(0.025)	(0.026)
Manager PQM					0.1870	0.1436
					(0.172)	(0.170)
Worker PQM					0.0637	0.1305
					(0.182)	(0.182)
Observations	256	256	198	198	198	198
Female manager × Variable + Variable:						
Manager hrs	0.3322	0.3649	0.3909	0.4303	0.3836	0.4196
Manager hrs (p-val)	0.001	0.000	0.008	0.003	0.010	0.005
Family hrs	0.0114	0.0128	-0.0033	0.0077	-0.0048	0.0072
Family hrs (p-val)	0.809	0.791	0.957	0.901	0.940	0.909
No relation hrs	0.1537	0.1636	0.1512	0.1833	0.1542	0.1848
No relation hrs (p-val)	0.001	0.000	0.010	0.002	0.010	0.002
Worker hrs	0.1131	-0.0114	0.1206	-0.0266	0.1217	-0.0218
Worker hrs (p-val)	0.029	0.859	0.080	0.722	0.084	0.774
Inputs value	-0.0121	-0.0082	-0.0121	-0.0016	-0.0100	0.0015
Inputs value (p-val)	0.533	0.673	0.647	0.949	0.715	0.954
Children	0.0066	0.0042	-0.0176	-0.0195	-0.0210	-0.0243
Children (p-val)	0.878	0.922	0.751	0.725	0.710	0.667
Manager plots	0.1823	0.1754	0.1640	0.1363	0.1563	0.1351
Manager plots (p-val)	0.009	0.012	0.065	0.119	0.094	0.144
Manager educ.	0.0137	0.0174	0.0112	0.0187	0.0108	0.0180
Manager educ. (p-val)	0.543	0.443	0.718	0.540	0.736	0.566
Worker educ.	0.0069	0.0108	0.0141	0.0079	0.0142	0.0058
Worker educ. (p-val)	0.752	0.624	0.636	0.788	0.651	0.852
Manager age	-0.0156	-0.0205	-0.0281	-0.0329	-0.0273	-0.0318
Manager age (p-val)	0.364	0.243	0.220	0.160	0.238	0.178
Worker age	0.0231	0.0248	0.0371	0.0373	0.0373	0.0366
Worker age (p-val)	0.110	0.096	0.065	0.069	0.068	0.079
Manager PQM					0.0275	-0.0005
Manager PQM (p-val)	.	.	.	.	0.805	0.997
Worker PQM					0.0618	0.0856
Worker PQM (p-val)	.	.	.	.	0.693	0.586
H0: Manager report = Worker report (p-values):						
Female worker hrs	0.394		0.209		0.222	
Male worker hrs	0.019		0.023		0.032	

Household-clustered standard errors in parentheses.

* p<0.1 ** p<0.05 *** p<0.01

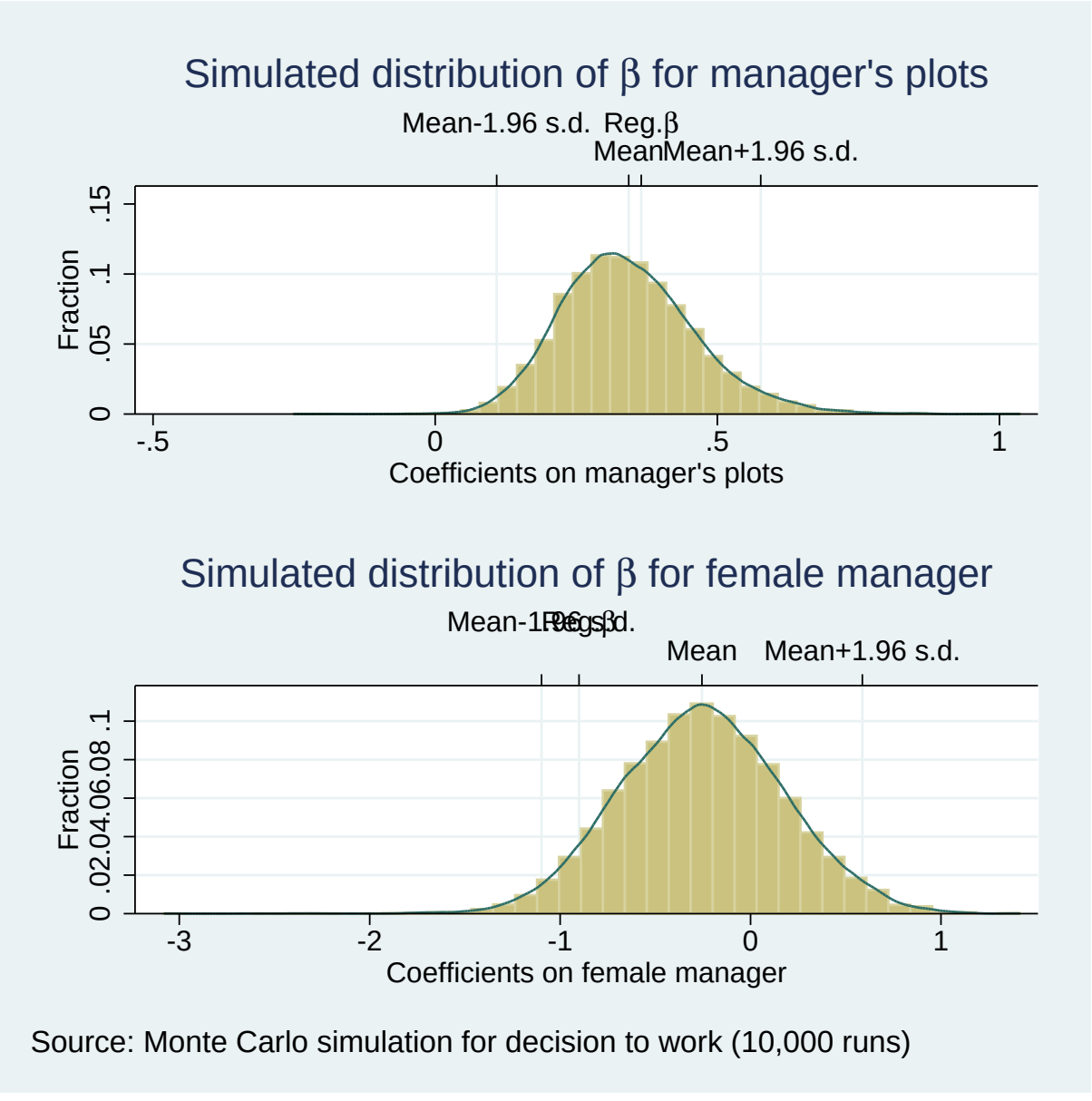
Table A2: T-test of differences in reporting of labor hours

T-test of differences for all managers					
Variable	Obs	Mean	Std. Err.	[95% Conf.	Interval]
Manager's report	141	71.965	6.032	60.04	83.89
Worker's report	160	88.854	8.268	72.525	105.183
Difference		-16.889	10.234	-37.034	3.256
H0 : Difference = 0 t=-1.650					
Ha : Difference < 0		Ha : Difference! = 0		Ha : Difference > 0	
Pr(T < t) = 0.050		Pr(|T|>|t|) = 0.100		Pr(T > t) = 0.950	

T-test of differences for female managers					
Variable	Obs	Mean	Std. Err.	[95% Conf.	Interval]
Manager's report	36	55.565	9.906	35.454	75.676
Worker's report	45	25.576	5.847	13.793	37.359
Difference		29.99	11.503	6.964	53.015
H0 : Difference = 0 t=2.607					
Ha : Difference < 0		Ha : Difference! = 0		Ha : Difference > 0	
Pr(T < t) = 0.994		Pr(|T|>|t|) = 0.012		Pr(T > t) = 0.006	

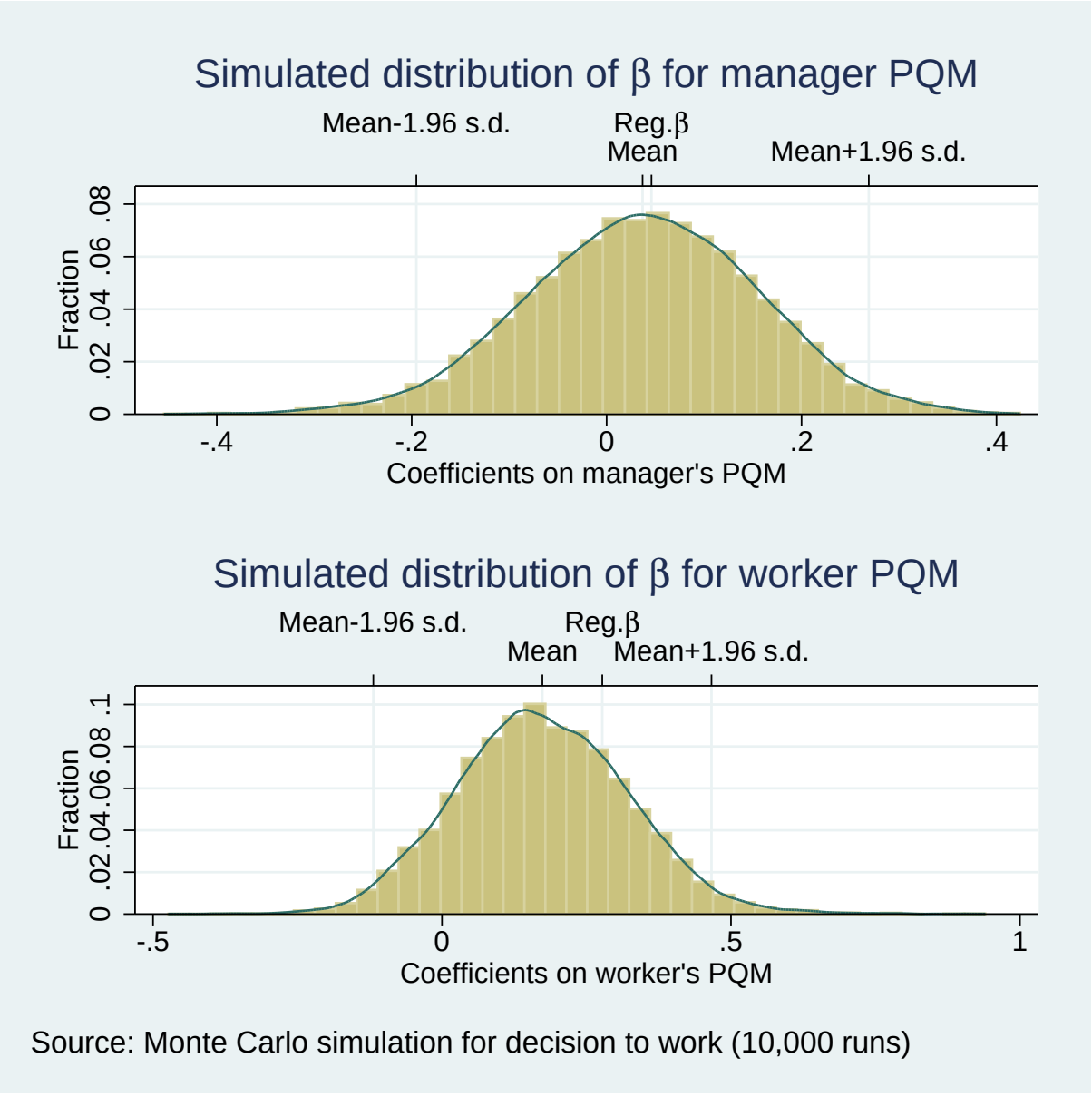
T-test of differences for male managers					
Variable	Obs	Mean	Std. Err.	[95% Conf.	Interval]
Manager's report	105	77.588	7.297	63.117	92.058
Worker's report	115	113.615	10.41	92.994	134.237
Difference		-36.028	12.713	-61.095	-10.96
H0 : Difference = 0 t=-2.834					
Ha : Difference < 0		Ha : Difference! = 0		Ha : Difference > 0	
Pr(T < t) = 0.002		Pr(|T|>|t|) = 0.005		Pr(T > t) = 0.998	

Figure A1: Distributions of Coefficients on Manager's Plots and Female Manager, for Decision to Work Regression (Column 1 in Tables 13 and 14).



Source: Monte Carlo simulation for decision to work (10,000 runs)

Figure A2: Distributions of Coefficients on Manager's and Worker's PQM, for Decision to Work Regression (Column 1 in Tables 13 and 14).



Source: Monte Carlo simulation for decision to work (10,000 runs)

Figure A3: Distributions of Coefficients on Female Manager and PQM, for Manager Decision Regression (Column 2 in Tables 13 and 14).

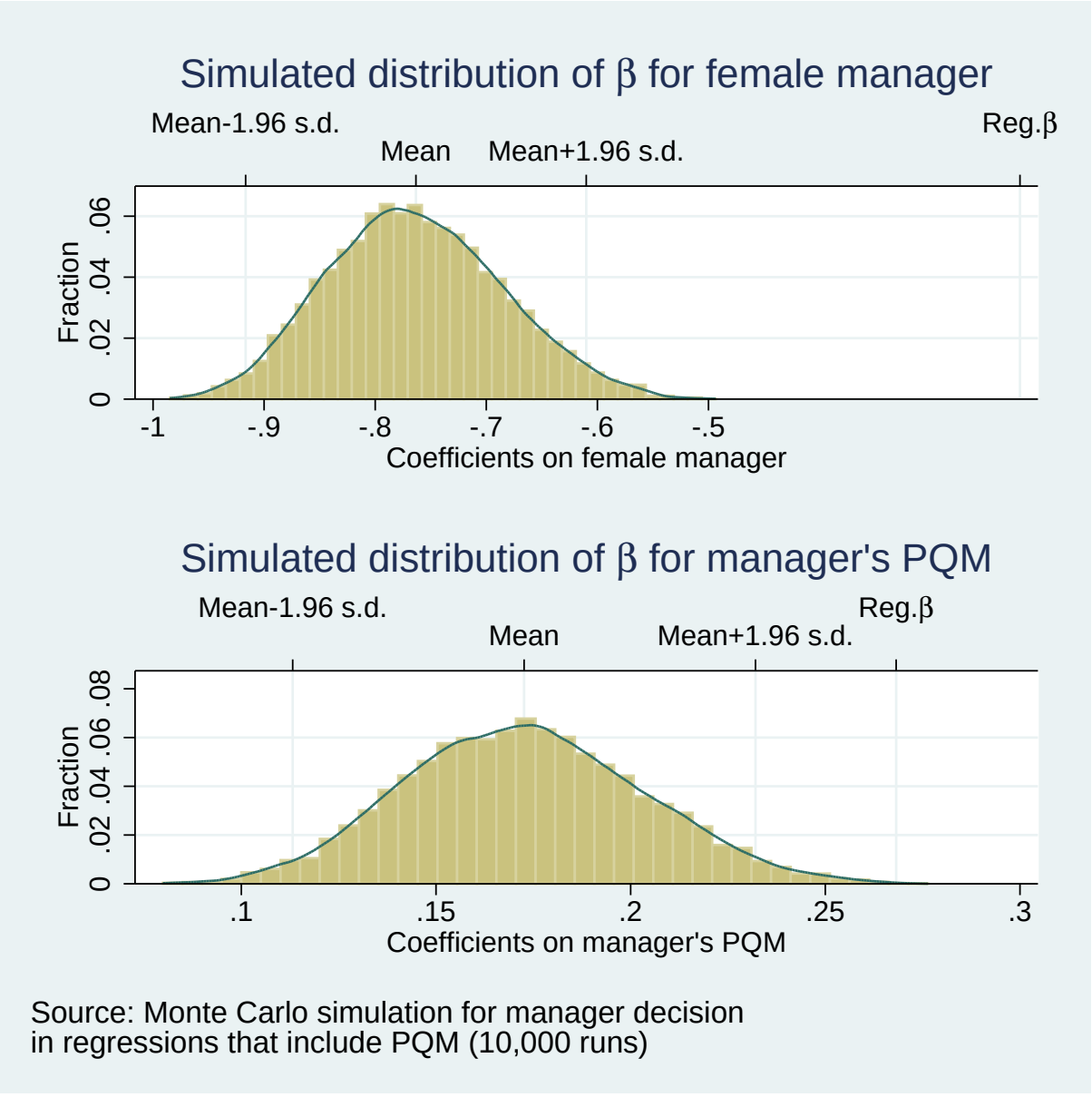
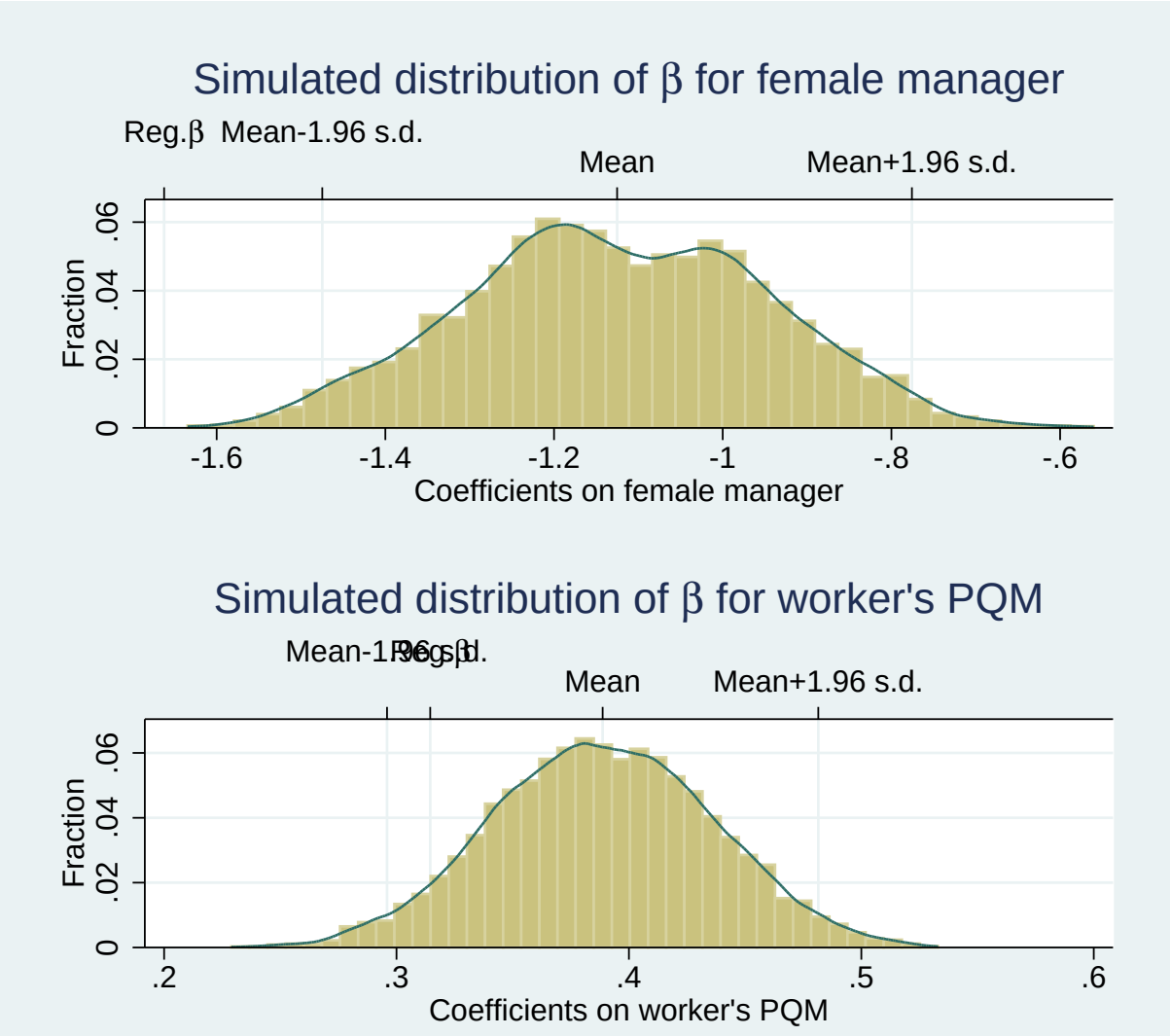


Figure A4: Distributions of Coefficients on Female Manager and PQM, for Worker Decision Regression (Column 4 in Tables 13 and 14).



Source: Monte Carlo simulations for worker decision in regressions that include PQM (10,000 runs)

Figure A5: Distributions of Coefficients on Female Manager, for Regressions of Manager and Worker Decisions to Report (Columns 3 and 5 in Tables 13 and 14).