Labor Elasticities, Market Failures, and Misallocation: Evidence from Indian Agriculture*

Joshua D. Merfeld

November 4, 2021

Abstract

This paper presents evidence of misallocation across households in rural Indian agriculture. I show that household demographics predict own farm labor demand for smallholder farmers but not non-smallholder farmers. A simple model of labor allocation predicts a clear consequence of this duality: smallholder farmers will reallocate labor across plots less in response to price changes than non-smallholders. Detailed household panel data confirms this theoretical prediction. Three additional facts suggest that a lack of off-farm labor opportunities may be partly responsible for the behavior of smallholders, leading smallholders to overallocate labor to agricultural production. First, smallholders report fewer hours of involuntary unemployment when own crop prices increase. Second, yield is substantially higher for smallholders on plots of the same size. Finally, estimated marginal revenue products of labor are consistently lower for smallholders.

Keywords: misallocation, markets, market failures, agriculture, labor
JEL Codes: D24, J20, J43, O13, Q12, Q13, Q15, Q18, Q24

*I am indebted to Leah Bevis, Marc Bellemare, Rajeev Dehejia, Alex Eble, Jonathan Morduch, Martin Rotemberg, and Debraj Ray for discussions and comments. Seminar participants at NYU, Columbia, the KER conference, and OARES provided helpful comments. All remaining errors are my own.

1 Introduction

Agriculture in developing countries is much less productive than in developed countries (Gollin et al., 2014). The per worker GDP of the richest countries is 78 times higher than the same number in the poorest countries; outside of agriculture, on the other hand, the difference is just five (Restuccia et al., 2008), consistent with large differences in sectoral productivity even within countries (Gollin et al., 2014; McCullough, 2017). As such, agriculture appears to be one of the driving forces behind aggregate productivity differences across countries. A recent literature has argued that the allocation of land may be partly responsible for the low productivity of developing country agriculture. In particular, a clear conclusion that has emerged from much of this literature is that there are too many small farms (Adamopoulos and Restuccia, 2014; Adamopoulos et al., 2017; Adamopoulos and Restuccia, 2020; Foster and Rosenzweig, 2017).

An important feature of agricultural households is that they are both producers and consumers of the same good. This feature is described in the classical agricultural household model (Singh et al., 1986). In the canonical model under common assumptions, production and consumption decisions are separable. In other words, households are able to first make production decisions to maximize profits and then make consumption decisions. Importantly, this implies that production decisions are independent of consumption decisions and, thus, that household consumption preferences do not affect production decisions. Benjamin (1992) was the first to note that recursion implies production decisions should be independent of any household characteristics that affect only consumption, like demographic characteristics. As such, a straightforward test for market completeness is to regress total farm labor on household demographics; under the null hypothesis of complete markets, household demographics should not affect total farm labor demand and the vector of coefficients will be zero. Rejection of this hypothesis implies that markets are not complete. While Benjamin (1992) was unable to reject complete markets, more recent literature unequivocally challenges this finding; Dillon and Barrett (2017), Dillon et al. (2019), and LaFave and Thomas (2016) all strongly reject market completeness in multiple contexts.

However, incomplete markets have additional implications for agricultural production, as well. A simple model of labor allocation makes a clear prediction: households for which markets fail will reallocate labor less in response to crop price changes than will households for which markets are complete. To test this prediction, I implement Benjamin's basic test for market completeness in India and split the sample based on one variable possibly correlated with market completeness: landholdings. I find evidence of misallocation across landholding size, as I am unable to reject recursion for non-smallholders, but strongly reject recursion for smallholders. Additional results confirm the theoretical prediction of this differential behavior: smallholder farmers reallocate labor across plots in response to price changes less than do non-smallholder. In other words, non-smallholders appear to be able to take better advantage of new information – conveyed through local crop prices – than smallholders, leading non-smallholders to more efficiently allocate labor throughout the agricultural season. This relationship is driven by the fact that non-smallholders can treat individual plots separately, as if they were separate firms, while smallholders cannot due to the failure of recursion; they equate MRPLs across plots with one another, not with the market wage, which leads to reallocation of labor from one plot to another and blunts the labor reallocation effects of price changes. This is consistent with recent evidence of substantial differences in production responses across different household and firm types in developing countries (Hardy and Kagy, 2020).

An important remaining question is what the source of this misallocation is. Additional analyses present suggestive evidence that a lack of off-farm wage opportunities may be responsible. First, an (unexpected) increase in crop price induces smallholders to report lower levels of involuntary unemployment but does not affect their allocation to wage employment. This is consistent with a story in which a decrease in crop prices leads smallholders to reallocate time to (unsuccessfully) search for off-farm wage labor. Importantly, non-smallholders do not reallocate labor in a similar way in response to changes in crop prices; the coefficients are not only insignificant but also small in magnitude. Second, output per hectare is much higher on smallholder plots than non-smallholder plots, even for plots of the same size. In other words, it appears that smallholders are more intensively farming their plots than are non-smallholders, which is consistent with a lack of wage opportunities but inconsistent with a lack of credit preventing smallholders from hiring in additional labor.[1] Finally, I calculate MRPL from naïve production function estimates, identified with fixed effects, and find that MRPL estimates are much higher for non-smallholders than for smallholders. In particular, the median is 52 percent higher and the mean is 71 percent higher, indicating an overallocation of labor to agricultural production for smallholders. As prima facie evidence of face validity for these MRPL calculations, the median hourly MRPL for non-smallholders is around one-ninth the reported daily agricultural wage. Since non-smallholders hire in labor for agricultural production, the lack of off-farm wage opportunities does not appear to lead to substantial deviations from the predicted equality of MRPL and the market wage for this subsample of households.

This paper contributes to several lines of literature. First, this article is related to the literature on misallocation and productivity. Previous, more macro-oriented research (Restuccia and Rogerson, 2008; Hsieh and Klenow, 2009; Adamopoulos and Restuccia, 2020) has provided evidence substantial evidence of misallocation, across a range of different margins. Importantly, these effects are found not only in developing countries, but in developed countries, as well (Fajgelbaum et al., 2019; Hsieh and Moretti, 2019; Baqaee and Farhi, 2020). Much of this research has been able to quantify effects of misallocation through their use of structural modeling. However, the ability to quantify these effects comes at a cost, as it requires a number of assumptions. In this paper, I take a complementary approach, using fewer assumptions but, at the same time, being unable to

[1] This finding also touches on the vast literature on the inverse farm size-productivity relationship. Some recent examples include Carletto et al. (2013), Bevis and Barrett (2020), and Wineman and Jayne (2020).

quantify the total effect. Additionally, the results in this paper appear to corroborate recent evidence that the distribution of land may be an important source of misallocation in the aggregate (Adamopoulos and Restuccia, 2014; Adamopoulos et al., 2017; Adamopoulos and Restuccia, 2020; Foster and Rosenzweig, 2017).

This paper also extends the literature testing for market completeness using the agricultural household model (Benjamin, 1992; Dillon and Barrett, 2017; Dillon et al., 2019; LaFave and Thomas, 2016). In addition to providing evidence of market completeness in India in a more general sense, the results also underline the point that market failure is a household-specific, not market-specific, phenomenon. This is consistent with other contemporaneous evidence from Indonesia (LaFave et al., 2018). While this is not a new argument, this paper offers clear empirical evidence of some of the implications of household-specific market failures. Given recent evidence of household behavioral responses to change production environments – like climate change – the evidence presented in this article suggests smallholders may be less able to respond effectively (Jagnani et al., 2020).

The rest of this paper is organized as follows. The next section elaborates a theoretical model of how market completeness affects household production decisions. Section 3 discusses the data and methodology, including summary statistics. Section 4 presents the main results and Section 5 concludes.

2 Separation and Agricultural Production

Consider an agricultural household that maximizes its own utility, subject to agricultural production and a possible off-farm labor constraint. Consumption, c and leisure l, are the arguments in the household's utility function. The household operates $i > 0$ plots – with characteristics A_i – and allocates its own labor to any of those plots, L_i^F, and wage labor, L_w, subject to possible constraint on total wage labor supplied off-farm, $\bar{L}_w$. Thus, the household's total time endowment is $\bar{L} = l + \sum_i L_i^F + L_w$. The household can also hire in labor, L_i^H, where i again indexes different plots.

Thus, the household's problem is:

$$\max u(c, l), \text{ subject to :} \tag{1}$$

$$c \leq \sum_i p_i f_i(L_i^F + L_i^H; A_i) + w(L_w - \sum_i L_i^H) \tag{2}$$

$$\bar{L} \geq l + \sum_i L_i^F + L_w \tag{3}$$

$$\bar{L}_w \geq L_w \tag{4}$$

$$0 \leq L_i^H, L_i^F, L_w, l, \tag{5}$$

where p_i is the price of the crop grown on plot i, and w is the wage. Due to my identification strategy, I model this as a static problem. The strategy focuses on *intra*seasonal labor reallocations across plots and, thus, land is fixed during the relevant period.

As previously shown (Benjamin, 1992; Dillon et al., 2019; LaFave and Thomas, 2016), complete markets allow households to separate production decisions from consumption decisions. In effect, households first maximize farm profits before making consumption decisions. The household's production decisions can thus be modeled as:

$$\max \sum_i p_i f_i(L_i; A_i)\gamma_i + w(L_w - \sum_i L_i^H) \equiv \pi^*. \tag{6}$$

Powerfully, this means that total farm labor, L^{Total}, is only a function of prices:

$$L^{Total*} = L^{Total*}(p_i, w \mid \gamma_i) \tag{7}$$

This prediction forms the basis of most tests for complete markets. When markets are complete, total farm labor is uncorrelated with household consumption characteristics, such as household demographic variables. This is the basis for the first test in this paper, which follows LaFave and Thomas (2016) and which I describe in detail in section 3.

Next, consider the first-order conditions for total labor demand:

$$p_i \frac{\partial f_i}{\partial L_i} = w = p_j \frac{\partial f_j}{\partial L_j}, \tag{8}$$

Importantly, conditional on the price of crop i and the market wage, the price of crop $j \neq i$ does not affect labor demanded on plot i. Intuitively, this is driven by the fact that each plot's labor choice is driven by equality of the marginal revenue product of labor (MRPL) with the market wage. A change in the price of crop j will affect the MRPL of labor on plot j, but will not affect the MRPL of labor on plot i. Consider now a world in which the off-farm labor constraint (equation 4) is binding.[2] The first-order conditions are now,

$$p_i \frac{\partial f_i}{\partial L_i} = w^* = p_j \frac{\partial f_j}{\partial L_j}. \tag{9}$$

Importantly, the shadow wage, w^*, is defined by the household's MRPL and, as such, the shadow wage generally does not equal the market wage. Suppose recursion fails for all households, such that equation (9) is true. This does not, in and of itself, provide evidence of misallocation *across agricultural households*, because all households could face the same friction. However, consider a situation in which recursion fails for some subgroup of households but holds for some other group of households. Since the former are equating MRPLs with the shadow wage but the latter are equating MRPLs with the market wage,

[2]More generally, at least two markets are required to fail if we are to observe non-separation in the data.

this is consistent with the two groups of households not equating MRPLs with one another; in other words, it is prima facie evidence of misallocation.

We can see one clear implication of this differential behavior in labor allocation decisions. Suppose p_j increases, increasing the MRPL of L_j. To bring the household back to an optimal allocation of labor, the household must increase L_j in order to bring $MRPL_j$ back to equality with other labor uses. At the same time, given the market constraint – and the fact that households for whom $\bar{L}_w$ binds do not simultaneously hire labor – households must allocate labor *away* from other activities. In the current model, this would be l and L_i. Thus, an increase in p_j affects both L_i and l, whereas neither is affected when markets are complete.

Figure 1 presents a graphical explanation of this. Panel A shows the change in L_i when p_i changes and markets are complete. The increase in price induces an increase in MRPL, which in turn induces the household to increase L_i through either an increase in L_i^F – if the household supplies labor to the market – or by hiring in labor and increasing L_i^H – if L_w was already zero at the previous optimum. Again, however, labor allocated to other plots (or to household activities) should be unaffected. Panel B shows the same situation when markets are incomplete. A rise in p_i induces an increase in $MRPL_i$, which then causes the household to increase labor allocated to plot i. However, given the labor constraint, this reallocation must come from a decrease in labor allocated to other activities, such as plot j. This reduction in labor allocated to plot j causes a simultaneous increase in $MRPL_j$. This adjustment results in a smaller increase in L_i in the incomplete markets case than in the complete markets case. In other words, the elasticity of labor with respect to crop price is lower when markets are incomplete. This is the key prediction tested in this paper.

2.1 Productivity Spillovers

One alternative explanation for the predictions above is productivity spillovers. For example, the total amount of labor a farmer allocates to her plots may directly affect productivity on plots, even conditional on total labor allocated to each plot individually. One possible explanation for such a relationship would be if input (or output) prices depend on the quantity purchased (sold). Importantly, any effects on output either directly or indirectly through labor productivity can impact the predictions made above.

As way of example, suppose the total amount of labor applied across all plots affects the marginal product on a plot, even conditional on that plot's labor allocation. Specifically, assume increasing total labor applied across plots also increases the marginal (revenue) product of labor on plot i. Then, an increase in the price of the crops planted on other plots will lead to an increase in labor applied to those other plots. In turn, this increase in labor application leads to an increase in the marginal revenue product of labor on plot i. In the complete market case, the household would then *increase* labor applied to plot i to re-equate MRPL on that plot with the market wage. In the incomplete market case, on the other hand, this would lead to less of a reallocation of labor away from plot i. To test for these possibilities, I test for spillovers of this type in the results section below.

2.2 Efficiency of Family and Hired Labor

Another alternative, explicated in Benjamin (1992), is that family and hired labor have different prices. This could be driven, for example, by differing efficiencies of family and hired labor. There are two possibilities. First, consider a situation in which family labor is less efficient than hired labor. In this case, the household's profit-making labor allocation is to allocate family labor completely to the market, until the point that the marginal utility of an additional hour of work is equated with the marginal utility of an additional hour of leisure, with consumption being the relevant trade-off between the two. Importantly, the predictions above are unchanged, as separation still occurs: the household maximizes on-farm profits by allocating only hired labor to agricultural production, up to the point that the MRPL of that hired labor equals their wage.

The second case, in which family labor is more efficient than hired labor, is different.

Following Benjamin (1992), assume family and hired labor are perfectly substitutable, but one hour of hired labor is equal to α hours of family labor. Thus, we can write total family labor-efficient units as:

$$L_e^F = L^F + \alpha L^H. \tag{10}$$

Importantly, the exact mix of family and hired labor will depend on household preferences; separation does not occur. However, the *total efficiency units of labor do not depend on household preferences*. Rather, they are determined by the first-order conditions (Benjamin, 1992). Thus, for a given α and a given L_e^F, the total amount of labor applied will be unchanged across plots, within the household.

As we will see, we are able to reject recursion for smallholders but not non-smallholders, even for those in the same village. As such, for differences in the efficiency of labor to drive findings, efficiency would have to be different across categories. Given that I find MRPL for smallholders is lower than MRPL for non-smallholders, efficiency differences seem unlikely to drive the results.

3 Data and Empirical Strategy

This paper uses ICRISAT's Village Dynamics in South Asia (VDSA) data.[3] ICRISAT has been collecting longitudinal data in India for several decades, but I use the most recent longitudinal data, which spans the years 2010 to 2014. My final sample, which I describe in more detail below, comprises 1,089 different households across 17 districts in 8 different states. Importantly, the data contains monthly-level information on labor and resource allocation across agricultural plots for the entire five years of the panel. Data is collected monthly, so recall is minimized. In addition, the village data collects information on individual crop prices relevant for local farmers, also at monthly intervals, which plays

[3]http://vdsa.icrisat.ac.in/vdsa-index.htm

an important role in the empirical strategy I employ. Finally, five separate years of data remove some concerns regarding the heterogeneity of effects when populations are subject to aggregate shocks (Rosenzweig and Udry, 2020).

3.1 Empirical Strategy and Identification

This paper approaches the question of separation and complete markets in several ways, making use of the rich panel data. Since I use household-level fixed effects, all regressions cluster standard errors at the household level unless otherwise reported. First, I borrow specifications from prior literature and analyze whether household demographics predict farm-level labor demand (Benjamin, 1992; Dillon and Barrett, 2017; Dillon et al., 2019; LaFave and Thomas, 2016). I diverge from the prior literature in two key ways. First, five years of panel data allow me to employ fixed effects at much lower levels of aggregation than other literature. In particular, I am able to estimate regressions using household-plot-crop fixed effects, which restricts attention only to plots planted with the same crop in multiple years. Second, much of the previous literature has used data from Africa (Dillon and Barrett, 2017; Dillon et al., 2019) or Indonesia (Benjamin, 1992; LaFave and Thomas, 2016), whereas the ICRISAT VDSA data was collected in India.

I first explore the relationship between household demographics and plot-level labor demand:

$$logL_{ikt} = \alpha_{cik} + \gamma_{vtdc} + X_{ikt} + rain_{vt} + \beta(\Sigma_{g=1}^{G}\delta_{gkt}) + \varepsilon_{ikt}, \tag{11}$$

where $logL_{ikt}$ is log of total labor applied to plot i in household k in season t, α_{cik} is household-plot-crop fixed effects, γ_{vtdc} is village-wave-season-crop fixed effects, X_{ikt} is a vector of time-variant plot characteristics – area planted and area irrigated – $rain_{vt}$ is total rainfall in village v in season t, δ_{gkt} is a group of variables indicating the number of household members that reside in the household in each demographic group g, and ε_{ikt} is a mean-zero error term. Following previous literature, the assumption of complete markets implies that β is a vector of zeros, that is, that demographic variables do not belong in the labor-demand equation. Thus, the null hypothesis is that $\beta = 0$, and F is the appropriate test statistic. I split household members into five separate demographic groups: prime-age males (15-59), prime-age females, elderly males (60+), elderly females, and children (<15). The main specifications include log of household size along with shares of four of the five demographic groups, with children being the omitted category. I also test robustness to alternative demographic definitions.

Identification here relies on there being no unobserved time-variant variables correlated with the error term and the demographic variables. The village-wave-season-crop fixed effects help alleviate any concerns that crop-specific aggregate village shocks in a given season are correlated with household size and total labor allocation (see for example, the shock studied in Dower and Markevich (2018)). This could be the case if, for example, shocks lead to changes in migration. Household-plot-crop fixed effects alleviate additional concerns related to the endogeneity of household size, crop choice, and area planted. For example, if changes in household size are correlated with planting decisions – perhaps a larger household will decide to plant a larger area – then controlling for area planted may actually lead to biased demographic coefficients. While this would be a clear rejection of the separation hypothesis, if the effect of demographics on labor demand operates completely through area planted, then the bias might lead to a failure to reject the null hypothesis if we control for area planted. The fixed effects help alleviate this concern.

While the previous literature on separation and market failures has pooled all households, such a specification does not provide evidence of misallocation. As such, I estimate Equation 11 separately for smallholders and non-smallholders, as defined by the survey. Smallholders are defined as landholders in the lowest two brackets of the landholding distribution, which includes the landless.[4] Table 1 presents summary statistics of some of the differences between smallholders and non-smallholders in the VDSA data. I estimate fully interacted models, allowing the effects of all variables – including fixed effects – to vary by smallholder status. I also vary the definition of smallholder status with the main results, which I discuss below.

The predictions of the model relate to how households respond to changes in prices. It is difficult to find exogenous variation in prices with respect to household labor allocation and output at an aggregate level. As such, I focus on individual households and how they reallocate labor across plots *within* the agricultural season. In particular, I focus on monocropped plots – plots planted with just a single crop – and examine how households reallocate labor when the price of one crop changes relative to the price of another crop on that household's own plots. Before testing the predictions of the model, I first verify that households do indeed reallocate labor across plots in response to changes in crop prices. I estimate:

$$logL_{ikmy} = \alpha_{cik} + \gamma_{dymc} + X_{ikmy} + Z_{ikt} + rain_{vmy} + \beta logP_{ymvc} + \varepsilon_{ikmy}, \tag{12}$$

where $logL_{ikmy}$ is log of non-planting and non-harvest labor allocated to plot i in household k in month m in year y, α_{cik} is household-plot-cropped fixed effects, γ_{dymc} is district-year-month-crop fixed effects, X_{ikmy} is a vector of characteristics that vary by month (specifically, the amount of labor and materials that had been allocated to that plot in that season up to that point), Z_{ikt} is a vector of planting hours and planting materials in that season, and $logP_{ymvc}$ is the monthly price of the crop planted on that plot, which is defined at the village level. In some specifications, I allow the effects of X, Y, and $rain$ to vary by the month of the year. In this specification, the coefficient of interest is β; it shows

[4]Some landless households rent in land for agricultural production at some point in the survey.

the effect of a change in the crop price on labor allocation at the plot level. For labor, note that this includes both family labor and hired labor. Perhaps unsurprisingly, hired labor is much more common for non-smallholder households.

The main hypotheses relate to how households respond to changes in crop price, similar to the specification in Equation 12. To this end, many of the specifications are simple variations on Equation 12. For prediction one, I restrict attention to households with just two separate crops. I add an additional variable to the specification, which is the price of the second crop grown by the household (that is, the price of the crop that is not grown on plot i). For the other two predictions, I interact the crop price dummy ($logP_{cvmy}$) with a dummy for smallholder (prediction two) or with both a dummy for smallholder and the number of crops grown by the household (prediction three).

3.2 Identification

This paper explores how households respond to change in crop prices *within* the agricultural season. One advantage of this strategy is that area planted is necessarily fixed after the planting season, avoiding one complication. However, the key drawback is that household labor allocation and crop prices may be endogenous. Most obviously, they may both be responding to (expected) temporal changes in the agricultural season or a shared cause, like rainfall. What my empirical strategy needs to accomplish is to purge any expected changes in crop prices, as well as any spurious causation caused by other variables. In essence, I need the crop price variable to represent *unexpected* changes in the crop price for any given household.

To accomplish this, the identification strategy relies heavily on fixed effects. Within variation comes from district-year-month-crop fixed effects. Since all households in a district will be similarly affected by aggregate shocks for a given crop, identification comes from unexpected differences in the price for a single crop across villages within a district. This helps purge any possibility that households change plots based on preplanting signals – like weather forecasts (Rosenzweig and Udry, 2014, 2019). Also note that permanent differences in prices for a given crop in different villages within the same district are swept out by the household-plot-crop fixed effects.[5]

Since identification comes from within-season variation in crop prices, I control for all previous plot-level decisions. This includes the number of hours and materials used during planting as well as the sum of all previous hours and materials allocated to the plot between planting and the month of observation. Controlling for previous decisions should help alleviate any concerns that cyclical patterns are driving decisions, in addition to the fixed effects. In most specifications, the effect of previous hours and materials are allowed to vary by the month of the year. I also include monthly rainfall totals, which are also allowed to vary by the month of the year, since the timing of rainfall is especially important in rain-fed agriculture.

Since crop prices vary at the village-crop level and both smallholders and non-smallholders reside in each village, it is unlikely that differences in the predictive power of crop prices alone can explain the results. Nonetheless, if there is heterogeneity in the make-up of each village, this is possible. Appendix Table A2 shows that, at the plot level, lagged crop prices are equally predictive of current crop prices for both smallholders and non-smallholders. In other words, any differential reactions to price changes are not driven by differences in the predictive power of prices for different households.

A short discussion of what actually drives the (unexpected) crop price changes is warranted. At first glance, one might wonder whether this is simply noise. However, certain households in the sample, non-smallholders, respond very strongly to these signals, suggesting noise cannot alone explain the variation. Within-country price variation persists in developing countries (Osborne, 2004; Chatterjee and Kapur, 2016; Zant, 2018), even

[5]The district-year-month-crop fixed effects are also important to sweep out aggregate correlations in prices, output, and labor allocation driven by weather. Table A1 shows that in a simple cross-section regression total output, at the plot level, is highly negatively correlated with harvest price. This is likely driven by the fact that poor output driven by weather shocks often leads to higher crop prices. However, once we include the district-year-month-crop fixed effects, this negative correlation disappears.

in countries like India, which has invested significant amounts of money in improving infrastructure (Bellemare et al., 2013; Chatterjee and Kapur, 2016). In India, specifically, much of this variation may be driven by strict laws governing where farmers are able to market their agricultural output; farmers are generally only allowed to market output in the state in which they live, leading to cross-border discontinuities in prices, despite geographic proximity (Chatterjee, 2019). This does not, however, explain intra-state variation in prices. Instead, a look towards infrastructure may provide a partial answer. Though I control for village-specific rainfall, I do not control for rainfall and general agricultural productivity conditions in areas to which a given village is connected. Idiosyncratic changes in connected villages may drive similarly idiosyncratic changes in a given village. In fact, previous research has shown that just how one village is connected to other areas plays an important roll in price variation (Zant, 2018).

3.3 Summary Statistics

Table 1 presents summary statistics for a number of different variables, broken down by smallholder status. Note that observations are at the household-plot level, the same level at which Equation 11 is estimated. First, note that non-smallholders actually have larger households, on average, than smallholders. The demographic breakdown of the households are somewhat similar, though it appears that non-smallholders have slightly more prime-age female and smallholders have slightly more children.

On average, actual plots are larger, in terms of area planted, for non-smallholders, though they do not appear more likely to be irrigated, at least as a percentage of the plot. Consistent with this, non-smallholders also allocate more hours to these large plots. However, hours do not increase as much as area when comparing non-smallholders and smallholders, suggesting smallholder plots may be cultivated more intensively than non-smallholder plots, consistent with previous research. While individual plots are approximately 65 percent larger for non-smallholders, overall area planted is even higher, approximately three times larger than smallholder area planted.

Crop choice is somewhat similar across household types, though there are some differences. The highest average price per kg is highest for green gram, and non-smallholder plots are slightly more likely to be planted with green gram. Chickpea and pigeonpea are also higher-priced crops, and non-smallholder plots are more likely to grow the former, but not the latter. However, the most commonly grown crops, paddy and wheat – which together make up more than half of all plots – are equally likely to be grown on a plot across the two household types.

Table 2 presents summary statistics, at the plot level, for five different "months" of the year: one month prior to harvest, two months prior to harvest, three months prior to harvest, four months prior to harvest, and five months prior to harvest. The first thing to note is that total non-planting hours show modest differences by month, with total hours increasing by approximately 30 percent from one month prior to harvest to three months prior to harvest. Though the overall pattern for hired labor is similar, the magnitude of the change is much greater for hired labor than for total labor. Total materials used (in rupees), on the other hand, does not show similar patterns. Rather, materials appear to be increasing up until one month prior to harvest, at which point they decrease markedly. Crop prices appear to be increasing as we approach harvest – consistent with the months just before harvest being the leanest time of the year – other than five months prior to harvest. However, there are relatively few observations five months prior, so it is difficult to draw any firm conclusions.

4 Results

4.1 Testing for Separation Using Household Demographics

The results begin with the household separation regression in Equation 11. Table 3 presents these results. The first column is a cross-section regression of total plot-level labor on household demographics. The F-test at the bottom of the table fails to reject the null hypothesis for the first column. Column two – which adds household-plot fixed effects – and column three – which adds household-plot-crop fixed effects – suggest similar conclusions. In all three case, we are unable to reject the null hypothesis of separation. In fact, we are never close to rejection, with the smallest p-value being just 0.440 (column two).

However, the first three columns assume all households face similar conditions. Columns four through six allow differential effects for non-smallholders and smallholders, across all variables and fixed effects in columns five and six.[6] It appears that the results do not reject separation for non-smallholder households, but do reject separation for smallholder households. This is consistent across the three columns, including when we fully interact smallholder and when we remove area planted from the regression.[7] In other words, it appears that smallholder households do not act as if markets are complete, as consumption characteristics are predictors of production decisions. We are unable to reject no correlation for non-smallholders, however.[8] This combination fo results is inconsistent with an efficient allocation of production factors across the two subgroups, as the results indicate that smallholders and non-smallholders are not both equating MRPL with the market wage.

[6]They are estimated in a single regression, however.

[7]Since household-plot fixed effects are included in all columns, area planted is not necessarily a required covariate. Moreover, as shown in Table A3, household size is strongly correlated with area planted on individual plots. Nonetheless, column six removes area planted and area irrigated as additional covariates. Qualitative conclusions are unchanged.

[8]Table A4 presents robustness checks of varying demographic definitions. We consistently reject the null for smallholders but not for non-smallholders.

The correlation between demographics and labor demand for smallholders appears to be driven by prime females. Since the four demographic groups are share variables and household size is included as a covariate, the share variables are interpreted as changing the make-up of the household, but not changing the household total size. The omitted category is children, so the coefficients are interpreted as increasing each group relative to (i.e. decreasing) children. For smallholders, apparently having more women and fewer children leads to an increase in labor demand. One possibility is that women have more trouble finding outside work than men and, as such, the excess labor is applied to household production, in this case, to household agricultural production.

4.2 Labor Allocation and Crop Prices

Having provided evidence of misallocation, the rest of this paper explores a clear empirical prediction of this misallocation and then provides evidence that a lack of off-farm wage opportunities may be responsible. Before digging into the key predictions of differences in price-labor elasticities, I first present evidence that mid-season hours are productive and that there are no obvious cross-plot spillovers due to total labor allocation. I then show that households respond to changes in crop prices by reallocating labor across plots based on these price changes.

First, Table 4 explores the productivity of mid-season hours – those hours between planting and harvest. The table presents results from a simple Cobb-Douglas production function, with log of output, in rupees, as the dependent variable and total mid-season labor hours (log plus one) as the key independent variable. Additional covariates also include planting decisions, which are likely correlated with both output and mid-season hours (Kochar, 1999). Columns one through four show that mid-season hours are significant predictors of total output at the end of the season. One percent higher mid-season labor hours is associated with an increase in output of somewhere between 0.04 and 0.06 percent. In other words, these hours are indeed productive and, as such, farmers may reallocate their hours in response to changes in expected revenue driven by changes in crop prices.

Since the key prediction this article tests is that market failures lead to a linkage across plots, we must first rule out spillovers across plots due to other reasons. One possibility is that there are productivity spillovers due to labor allocation. Perhaps an increase in hours allocated to plots increases productivity due to bulk purchase discounts, for example. Columns five and six test this possibility. Column five adds as a covariate total mid-season hours on other plots. The coefficient is small – less than one-third the size of the mid-season hours coefficient – and not significantly different from zero. In column six, I include an additional interaction between mid-season hours on the plot and total mid-season hours on other plots. Again, there do not appear to be any spillovers related to mid-season labor allocation, at least not conditional on the other covariates included in the model.

Recall that when households act as if markets are complete – i.e. separation/recursion holds – the relevant labor trade-off for each plot is between the productivity on that plot and wage employment. If a household is working both for a wage and on one's own farm, then when a crop price increases, the household will shift family labor away from wage work and towards that crop. On the other hand, if the crop price decreases, the household will shift labor away from that crop and towards wage employment. Importantly, since households are price takers, this additional labor allocation to wage employment does not affect the wage. If markets are not complete, however, this is not the case. For a household that faces a wage labor constraint, they are not able to shift additional labor towards wage employment. This means any additional labor allocated to or from a plot must be from/to other places: leisure, domestic production, or other plots. Unlike with wage employment, this shift in labor also affects marginal productivity of labor in said tasks. This leads to a smaller labor reallocation response for these households than for households that are able to reallocate labor towards or away from wage employment.

Table 5 tests this prediction with four different dependent variables: total mid-season hours, family mid-season hours, hired mid-season hours, and mid-season materials. Column one presents the results for all labor. Consistent with economic theory, the elasticity of labor (re)allocation with respect to the crop price is significantly lower for smallholders than non-smallholders. For non-smallholders, a one-percent change in the crop price leads to a change in labor allocation of approximately 0.4 percent. For smallholders, on the other hand, it is just 0.15 percent.

Columns two and three present results for family and hired labor, respectively. While smallholders and non-smallholders reallocate family labor similarly, non-smallholders respond to crop price changes with hired labor much more than do smallholders. This is consistent with a world in which non-smallholder households need to hire in additional wage labor to meet their labor requirements, but smallholder households have what amounts to excess labor. Excess labor in this sense refers to smallholders having MRPLs on their plots lower than the market wage if they were to allocate all of their available labor to own agricultural production. Non-smallholders, on the other hand, would have MRPLs *higher* than the market wage, leading to them hiring additional labor. If true, smallholders would, on average, hire much less labor than non-smallholders, which is exactly what we see in the data. Moreover, the average seasonal total planted area per person – across the crops used in this study – is 2.27 acres per person for non-smallholders but just 0.77 acres per person for smallholders. We also see a difference in elasticities for materials, though the estimate for non-smallholders is quite imprecisely estimated.

One issue with the data is that smallholder status is not explicitly defined. As a robustness check, I create a new variable based on the percentiles of total land area. I do this using the first time a household is observed in the data so that smallholder status is identical across all waves for a given household. I match the average breakdown of smallholder status (around 34 percent of household-month observations are smallholders) using total land area. This new smallholder variable is equivalent to defining smallholder as any household with two or less acres of land.[9] I re-estimate Table 5 using this new variable and present the results in Table A5. All conclusions are identical. As an additional robustness check, I redefine smallholder again, this time using four acres. Table A6 present these results. Conclusions are identical, with the exception of materials.

The appendix presents several different robustness checks. First, Table A7 relaxes identifying assumptions by including village-year-month-crop fixed effects instead of district-year-month-crop fixed effects. The level effect of crop price is no longer identified, but the difference in its effect across household types is. All four specifications yield identical conclusions. Additionally, column five, including village-year-month-crop-fixed effects, also adds next month's crop price and its interaction with smallholder as an additional covariate. This is included to insure predicted price changes are not driving results. Column six instead includes district-year-month-crop fixed effects, the same specification as those in Table 5, as well as the following month's crop price, for a more apples-to-apples comparison to the results listed here in the main text. Qualitative conclusions are unchanged. In other words, it does not appear to be expected seasonal patterns driving the reallocation of labor.

Finally, all specifications in Table 5 include planting variables and the sum of all previous input choices as covariates. While these are not lagged variables in the traditional sense, it nonetheless raises some concerns regarding serial correlation and panel data. Table A8 presents results removing all planting and previous input allocation decisions from the regression. Conclusions are again unchanged.

4.3 Which markets fail?

A key question is which markets fail, leading to the misallocation observed across households. In this section, I present three pieces of evidence that a lack of off-farm employment opportunities may be responsible: individual time allocation, yields, and direct estimates

[9] This definition also leads to 34 percent of household-month observations being smallholders.

of MRPLs.

The dataset also collects information on monthly labor allocation of individual household members. We can use this information to further explore some aspects of the theory and perhaps better understand the nature of agricultural production in rural India. To do this, I first look at changes in individual-level time allocation to different productive activities based on changes in crop prices. One big issue with this is that most households grow more than one crop. As such, I create a price variable that is a weighted average of all crops grown by a household, weighted by the percentage of area in a given season allocated to each crop. In other words, for each household, I look at their total acreage in a season. I then take the percentage of that total acreage devoted to each individual crop and use that percentage to weight crop prices. Since I include fixed effects for the combination of crops grown – not each individual crop separately – variation comes from households that planted the exact same combination of crops but in different proportions.

The first set of results are in Table 6. I look at five separate types of activities: own farm days, wage days, other (productive) days, involuntary unemployment days, and non-farm work days (which are defined as the sum of wage and other days). I include all individuals in the dataset and also include age and age squared as covariates. Since there are a lot of zeros – only 25 percent of individual-month observations have non-zero wage days, for example – I transform all of the time-use variables using an inverse hyperbolic sine transformation (Bellemare and Wichman, 2020).[10] The first column is own farm days. As we would expect, we see a positive relationship between monthly (weighted) crop price and own farm days for both smallholder and non-smallholder households.

Column two presents result for wage labor. There appears to be no relationship between individual wage employment and crop price. However, this is additional evidence that

[10]Using logs is preferred because it allows for a straightforward elasticity interpretation. As such, in the main results presented above, I use logs as there are many more non-zero observations. Crop price, for example, is never zero, while mid-season hours has many more non-zero observations. Table A9 presents these same time-use results in levels. Qualitative conclusions are unchanged.

many smallholders may face a wage labor constraint; if that constraint is already binding, then smallholders will not be reallocating any labor to or away from wage employment unless the MRPL on that household's plots is only slightly below the market wage, such that an increase in price would lead them to again equate MRPLs and the market wage. The negative coefficient is consistent with some households facing such a situation, but the coefficient is small in magnitude, so caution is warranted interpreting even the sign. For non-smallholders, on the other hand, if they need to hire in labor for agricultural production, we would not expect them to also work on the market. A null effect is consistent with this.

We see no movement of other days, a relatively ill-defined category. The most interesting results are for involuntary unemployment days. There is a strong negative correlation between weighted crop price and involuntary unemployment days for smallholders. One possible interpretation for involuntary unemployment is when an individual searches for wage employment but fails to find any for a day, resulting in a day of (involuntary) unemployment. While an increase in one's own crop price(s) should not affect the probability of finding outside employment conditional on searching, it could affect the probability an individual searches for outside unemployment. A negative coefficient for smallholders suggests higher own crop prices makes individuals less likely to search for wage employment. Again, we see no relationship for non-smallholders, as we would expect.

While results in previous sections of the paper are consistent with this, the results are merely suggestive evidence of a failure in the off-farm labor market. Here I present two more direct pieces of evidence. First, consider another possible explanation: a failure of credit markets. Perhaps smallholders would like to hire in additional labor to help farm their plots, but are unable to find financing to do so. While this is hard to reconcile with the involuntary unemployment days finding, it is by no means conclusive. However, if credit markets are responsible, we would likely see lower yields for smallholders than non-smallholders, since non-smallholders would be able to hire additional labor, while smallholders would not. I construct residuals from regressions of yield on the independent variables in column two of Table 4, including household-plot-crop fixed effects. Figure 2 presents the relationship between these residuals and plot size. Since the distribution of plot sizes is quite different for smallholders and non-smallholders, I restrict the figure to just areas of common support, defined as between the 5th and 95th percentile of smallholder plot sizes. Across all plot sizes, yield is substantially higher for smallholders than for non-smallholders. This is inconsistent with smallholders lacking additional productive inputs, like labor, due to credit market failures. It is consistent, however, with smallholders allocating all their excess labor to own agricultural production.

The last piece of evidence is more direct, but also requires more assumptions. If smallholders are overallocating to own production, smallholder MRPL should be lower than non-smallholder MRPL. I estimate simple production functions, relying on household-plot-crop fixed effects for identification. These estimates are susceptible to the usual concerns regarding endogeneity and, as such, should be treated with a grain of salt. Nonetheless, it is the most direct way to test for a possible cause of misallocation.

Figure 3 presents MRPLs by percentile, separately for smallholders and non-smallholder. MRPLs are constructed from translog production functions, with materials, land, and labor as the productive inputs. Across all percentiles, non-smallholder MRPL is higher than smallholder MRPL. At the mean, the difference is around 71 percent. At the median, which is less influenced by some of the higher estimates in the upper end of the distribution, MRPLs are 52 percent higher for non-smallholders than smallholders. These MRPLs are hourly estimates. As a small test of the face validity of the estimates, non-smallholder median MRPL is around one-ninth the median daily wage reported in the data. In other words, non-smallholder MRPL is relatively close to the market wage, as we would expect from previous results for separation.

5 Conclusion

This paper presents an omnibus test for misallocation across households in agriculture. I find evidence of misallocation, specifically across categories of landsize: smallholders vs. non-smallholders. Specifically, I show that a simple test for market failures comes to different results across the two groups, which is prima facie evidence for misallocation. I also find support for theoretical predictions of these results, with smallholders reallocating labor in response to crop price changes less than non-smallholders. In other words, non-smallholders appear to be better able to react to new information – in the form of crop prices – than smallholders.

The overall thrust of the evidence is that a lack of off-farm wage employment is at least partly responsible for this misallocation. It appears that smallholders are overallocating labor to agricultural production, with higher yields on similar-sized plots and lower MRPLs across the distribution. As such, while the results are consistent with recent evidence that there may be too many small farms in developing countries, the fact that the misallocation across farm sizes may be partly driven by a lack of employment opportunities suggests caution when interpreting these results as evidence for a reallocation of land.

In addition, the overall evidence suggests that smallholders and non-smallholders respond much differently to price changes. These results are especially pertinent in a country like India, where the government is heavily involved in setting agricultural prices. The Indian government directly affects prices on both the consumption and production sides, through its public distribution system (PDS) and minimum support price (MSP) policies, respectively. Indirectly, other policies can also impact prices. For example, legislation that restricts farmers' sales to only markets within that farmer's state may artificially depress prices received by farmers (Chatterjee, 2019). However, any policy changes that affect agricultural prices – like the PDS and MSP – are likely to have different impacts on agricultural output in different areas. In other words, different households will respond differently to the same change in price driven by an agricultural policy. Production may remain relatively constant for a given crop in areas with more smallholders but may change significantly in areas with more large landholders. Similarly, the results on number of crops grown and the *increase* in elasticity for smallholders suggests districts with environments more amenable to diverse production may see larger changes in crop production than districts that are less suitable for diverse production if price of that crop changes.

References

Adamopoulos, T., Brandt, L., Leight, J., and Restuccia, D. (2017). Misallocation, selection and productivity: A quantitative analysis with panel data from China. Technical report, National Bureau of Economic Research.

Adamopoulos, T. and Restuccia, D. (2014). The size distribution of farms and international productivity differences. *American Economic Review*, 104(6):1667–97.

Adamopoulos, T. and Restuccia, D. (2020). Land reform and productivity: A quantitative analysis with micro data. *American Economic Journal: Macroeconomics*, 12(3):1–39.

Baqaee, D. R. and Farhi, E. (2020). Productivity and misallocation in general equilibrium. *The Quarterly Journal of Economics*, 135(1):105–163.

Bellemare, M. F., Barrett, C. B., and Just, D. R. (2013). The welfare impacts of commodity price volatility: evidence from rural Ethiopia. *American Journal of Agricultural Economics*, 95(4):877–899.

Bellemare, M. F. and Wichman, C. J. (2020). Elasticities and the inverse hyperbolic sine transformation. *Oxford Bulletin of Economics and Statistics*, 82(1):50–61.

Benjamin, D. (1992). Household composition, labor markets, and labor demand: testing for separation in agricultural household models. *Econometrica*, pages 287–322.

Bevis, L. E. and Barrett, C. B. (2020). Close to the edge: High productivity at plot peripheries and the inverse size-productivity relationship. *Journal of Development Economics*, 143:102377.

Bils, M., Klenow, P. J., and Ruane, C. (2020). Misallocation and mismeasurement? Technical report, National Bureau of Economic Research.

Carletto, C., Savastano, S., and Zezza, A. (2013). Fact or artifact: The impact of measurement errors on the farm size–productivity relationship. *Journal of Development Economics*, 103:254–261.

Chatterjee, S. (2019). Market power and spatial competition in rural india. *Working paper*.

Chatterjee, S. and Kapur, D. (2016). Understanding Price Variation in Agricultural Commodities in India: MSP, Government Procurement, and Agriculture Markets. In *India Policy Forum July*, volume 12, page 2016.

Dillon, B. and Barrett, C. B. (2017). Agricultural factor markets in Sub-Saharan Africa: An updated view with formal tests for market failure. *Food Policy*, 67:64–77.

Dillon, B., Brummund, P., and Mwabu, G. (2019). Asymmetric non-separation and rural labor markets. *Journal of Development Economics*, 139:78–96.

Dower, P. C. and Markevich, A. (2018). Labor misallocation and mass mobilization: Russian agriculture during the Great War. *Review of Economics and Statistics*, 100(2):245–259.

Fajgelbaum, P. D., Morales, E., Suárez Serrato, J. C., and Zidar, O. (2019). State taxes and spatial misallocation. *The Review of Economic Studies*, 86(1):333–376.

Foster, A. D. and Rosenzweig, M. R. (2017). Are there too many farms in the world? Labor-market transaction costs, machine capacities and optimal farm size. Technical report, National Bureau of Economic Research No. w23909.

Gollin, D., Lagakos, D., and Waugh, M. E. (2014). The agricultural productivity gap. *The Quarterly Journal of Economics*, 129(2):939–993.

Hardy, M. and Kagy, G. (2020). It?s Getting Crowded in Here: Experimental Evidence of Demand Constraints in the Gender Profit Gap. *The Economic Journal*, 130(631):2272–2290.

Hsieh, C.-T. and Klenow, P. J. (2009). Misallocation and manufacturing TFP in China and India. *The Quarterly Journal of Economics*, 124(4):1403–1448.

Hsieh, C.-T. and Moretti, E. (2019). Housing constraints and spatial misallocation. *American Economic Journal: Macroeconomics*, 11(2):1–39.

Jagnani, M., Barrett, C. B., Liu, Y., and You, L. (2020). Within-Season Producer Response to Warmer Temperatures: Defensive Investments by Kenyan Farmers. *The Economic Journal, forthcoming.*

Kochar, A. (1999). Smoothing consumption by smoothing income: hours-of-work responses to idiosyncratic agricultural shocks in rural India. *Review of Economics and Statistics*, 81(1):50–61.

LaFave, D., Peet, E., and Thomas, D. (2018). Who behaves as if rural markets are complete? *Colby College Working Paper*.

LaFave, D. and Thomas, D. (2016). Farms, families, and markets: New evidence on completeness of markets in agricultural settings. *Econometrica*, 84(5):1917–1960.

McCullough, E. B. (2017). Labor productivity and employment gaps in Sub-Saharan Africa. *Food Policy*, 67:133–152.

Osborne, T. (2004). Market news in commodity price theory: Application to the Ethiopian grain market. *The Review of Economic Studies*, 71(1):133–164.

Restuccia, D. and Rogerson, R. (2008). Policy distortions and aggregate productivity with heterogeneous establishments. *Review of Economic Dynamics*, 11(4):707–720.

Restuccia, D. and Rogerson, R. (2017). The causes and costs of misallocation. *Journal of Economic Perspectives*, 31(3):151–74.

Restuccia, D., Yang, D. T., and Zhu, X. (2008). Agriculture and aggregate productivity: A quantitative cross-country analysis. *Journal of Monetary Economics*, 55(2):234–250.

Rosenzweig, M. R. and Udry, C. (2014). Rainfall forecasts, weather, and wages over the agricultural production cycle. *American Economic Review*, 104(5):278–83.

Rosenzweig, M. R. and Udry, C. (2020). External validity in a stochastic world: Evidence from low-income countries. *The Review of Economic Studies*, 87(1):343–381.

Rosenzweig, M. R. and Udry, C. R. (2019). Assessing the Benefits of Long-Run Weather Forecasting for the Rural Poor: Farmer Investments and Worker Migration in a Dynamic Equilibrium Model. *National Bureau of Economic Research No. w25894*.

Singh, I., Squire, L., and Strauss, J. (1986). *The basic model: theory, empirical results, and policy conclusions*, pages 17–47.

Wineman, A. and Jayne, T. S. (2020). Factor market activity and the inverse farm size-productivity relationship in Tanzania. *The Journal of Development Studies*, pages 1–22.

Zant, W. (2018). Trains, trade, and transaction costs: How does domestic trade by rail affect market prices of Malawi agricultural commodities? *The World Bank Economic Review*, 32(2):334–356.

Tables

Figure 1: Agricultural Production and Market Failures

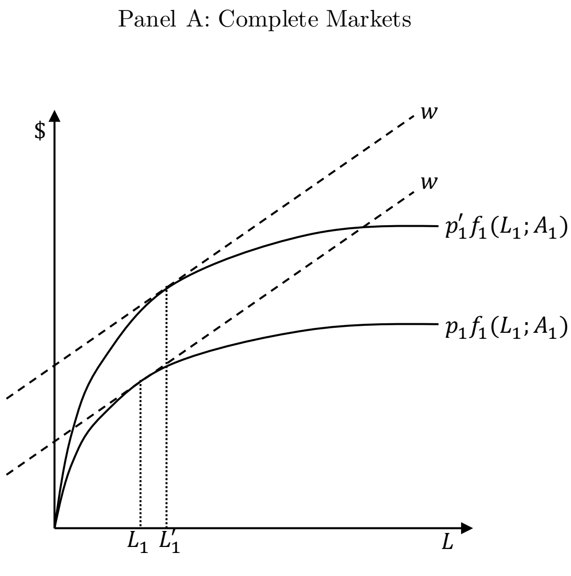


Table 1: Demography and Plot Summary Statistics

	(1) Non-smallholders	(2) Smallholders	(3) Diff (1-2)	(4) p-value
Household size	6.983	5.698	1.285	0.000
	(3.488)	(2.260)		
Elderly male percent	0.110	0.102	0.008	0.439
	(0.145)	(0.158)		
Elderly female percent	0.027	0.019	0.008	0.092
	(0.079)	(0.065)		
Prime male percent	0.493	0.473	0.020	0.287
	(0.241)	(0.240)		
Prime female percent	0.146	0.148	−0.001	0.924
	(0.179)	(0.176)		
Child percent	0.223	0.258	−0.034	0.025
	(0.187)	(0.199)		
Plot area planted (log acres)	−0.393	−0.893	0.501	0.000
	(1.219)	(1.148)		
Plot area irrigated (percent)	0.581	0.540	0.041	0.218
	(0.492)	(0.497)		
Total plot non-harvest labor (log hours)	3.390	3.073	0.317	0.005
	(1.847)	(1.766)		
Total area, all plots (log acres)	1.802	0.674	1.128	0.000
	(1.201)	(1.043)		
Total non-harvest labor, all plots	3.186	2.165	1.021	0.000
(log hours)	(2.331)	(2.139)		
Blackgram	0.023	0.020	0.004	0.467
	(0.151)	(0.140)		
Chickpea	0.085	0.054	0.031	0.003
	(0.278)	(0.225)		
Greengram	0.011	0.006	0.005	0.036
	(0.103)	(0.078)		
Groundnut	0.045	0.037	0.008	0.424
	(0.207)	(0.188)		
Lentil	0.034	0.026	0.008	0.196
	(0.180)	(0.158)		
Maize	0.050	0.073	−0.023	0.025
	(0.219)	(0.261)		
Paddy	0.303	0.296	0.007	0.806
	(0.460)	(0.457)		
Pigeonpea	0.101	0.103	−0.002	0.908
	(0.302)	(0.304)		
Sorghum	0.069	0.120	−0.051	0.001
	(0.253)	(0.325)		
Wheat	0.279	0.265	0.014	0.532
	(0.448)	(0.442)		
Observations	6,957	3,577		

Statistics are at the household-year-season-plot level, the same level and sample used in Table 3. The first column presents means and standard deviations for non-smallholders, as defined by the survey, while the second column presents the same statistics for smallholders. The third column presents the difference between (1) and (2) and the fourth column presents the p-value for that difference, calculated using a regression and clustering standard errors at the household level.

Table 2: Monthly Summary Statistics

	Months until harvest				
	(1)	(2)	(3)	(4)	(5)
	One	Two	Three	Four	Five
Non-planting hours (log)	2.619	2.697	2.892	2.807	2.719
	(1.222)	(1.320)	(1.262)	(1.091)	(1.266)
Non-planting hired hours (log)	0.888	1.183	1.528	1.254	1.258
	(1.463)	(1.618)	(1.696)	(1.590)	(1.604)
Materials used (log Rs)	4.574	4.864	4.349	4.087	3.273
	(2.586)	(2.452)	(2.760)	(2.547)	(3.201)
Monthly crop price (log Rs)	3.250	3.154	3.050	2.908	3.341
	(0.701)	(0.718)	(0.705)	(0.627)	(0.971)
Total previous non-planting hours (log)	4.289	3.436	2.643	2.298	1.795
	(1.505)	(1.777)	(1.793)	(1.793)	(1.737)
Total previous materials (log Rs)	6.832	5.580	4.626	4.565	3.003
	(2.347)	(3.048)	(3.264)	(3.173)	(3.426)
Observations	3,534	5,178	4,483	2,037	326

Statistics are at the household-year-month-plot level, the same level and sample used in Table 5. The statistics are broken down by months until harvest at the plot level, with five including any months greater than five, as well.

Table 3: Labor Allocation and Household Demographics

	(1) All labor	(2) All labor	(3) All labor	(4) All labor	(5) All labor	(6) All labor
					Non-smallholders	
Household size	0.005	−0.010	−0.006	−0.016	−0.014	−0.006
	(0.005)	(0.014)	(0.014)	(0.017)	(0.018)	(0.018)
Prime male percent	0.087	−0.118	−0.012	−0.018	0.004	−0.041
	(0.057)	(0.157)	(0.172)	(0.234)	(0.231)	(0.255)
Prime female percent	0.065	0.403*	0.407	−0.063	−0.016	−0.167
	(0.126)	(0.239)	(0.275)	(0.367)	(0.361)	(0.381)
Elderly male percent	0.031	−0.045	0.087	−0.043	−0.006	−0.193
	(0.073)	(0.167)	(0.188)	(0.285)	(0.280)	(0.313)
Elderly female percent	0.177	0.310	0.104	0.162	0.165	0.416
	(0.180)	(0.501)	(0.572)	(0.588)	(0.591)	(0.631)
					Smallholders	
Household size				0.007	0.005	0.007
				(0.015)	(0.015)	(0.015)
Prime male percent				−0.165	−0.097	−0.128
				(0.240)	(0.238)	(0.242)
Prime female percent				1.233***	1.346***	1.340***
				(0.445)	(0.430)	(0.446)
Elderly male percent				0.125	0.170	0.143
				(0.234)	(0.229)	(0.242)
Elderly female percent				1.200*	0.888	0.833
				(0.728)	(0.836)	(0.850)
Fully interacted model?	No	No	No	No	Yes	Yes
Area planted?	Yes	Yes	Yes	Yes	Yes	No
Fixed Effects:						
Village-wave-season-crop FE	X	X	X	X	X	X
Household-Plot FE		X				
Household-Plot-Crop FE			X	X	X	X
F-tests (all demographics = 0):					Non-smallholders	
F	0.642	0.962	0.557	0.280	0.225	0.370
p-value	(0.668)	(0.440)	(0.733)	(0.924)	(0.952)	(0.869)
					Smallholders	
F				2.581	2.575	2.422
p-value				(0.025)	(0.025)	(0.034)
Observations	10,524	10,524	10,524	10,524	10,524	10,524

In all columns, the dependent variable is the total number of hours allocated to agricultural production, including hired and family labor but excluding harvest labor. Observations are at the household-year-season-plot level. Regressions also control for total rainfall and rainfall squared. Standard errors are clustered at the household level.

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

Table 4: Productivity of Mid-Season Hours

	(1) All	(2) All	(3) All	(4) By month	(5) Spillovers	(6) Spillovers
Mid-season hours (log)	0.062***	0.048**	0.042**		0.046**	0.038*
	(0.020)	(0.020)	(0.020)		(0.021)	(0.022)
Total harvest labor (log)			0.210***			
			(0.038)			
Hours one month before harvest (log)				0.015		
				(0.026)		
Hours two months before harvest (log)				0.043**		
				(0.017)		
Hours three months before harvest (log)				0.036**		
				(0.017)		
Hours four months before harvest (log)				−0.008		
				(0.012)		
Hours five+ months before harvest (log)				0.037**		
				(0.018)		
Mid-season hours on other plots					0.014	0.007
					(0.009)	(0.014)
Mid-season hours times hours on other plots						0.002
						(0.005)
Planting hours (log)	−0.008	−0.013	−0.015	−0.016	−0.013	−0.013
	(0.031)	(0.031)	(0.028)	(0.031)	(0.031)	(0.031)
Planting mats (log)	0.019	0.009	0.017	0.007	0.009	0.009
	(0.013)	(0.013)	(0.014)	(0.014)	(0.013)	(0.013)
Fixed Effects:						
Village-Wave-Season-Crop	X	X	X	X	X	X
Household-Plot	X					
Household-Plot-Crop		X	X	X	X	X
Observations	10,534	10,534	10,534	10,534	10,534	10,534

Observations are at the household-year-season-plot level. The dependent variable in all columns is the log of total output (Rs). Mid-season labor is defined as any labor between planting and harvest. Regressions also control for total rainfall, rainfall squared, area planted, area irrigated, total planting materials used, and total planting hours.

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

Table 5: Input Allocation and Crop Price Elasticities

	(1) All	(2) Family	(3) Hired	(4) Materials
Monthly crop price (log R)	0.408***	0.311**	0.450***	0.580
	(0.117)	(0.131)	(0.137)	(0.468)
Monthly crop price times Smallholder	−0.250**	−0.020	−0.294**	−0.659***
	(0.104)	(0.102)	(0.123)	(0.184)
Fixed Effects:				
District-Year-Month-Crop	X	X	X	X
Household-Plot-Crop	X	X	X	X
Previous labor, previous materials, and rainfall:				
By month	X	X	X	X
Observations	24,318	24,318	24,318	24,318

Observations are at the household-year-month-plot level. In column one, the dependent variable is log (plus one) of all mid-season labor hours, defined as any labor between planting and harvest in each month. In column two, the dependent variable is log (plus one) of mid-season family hours. In column three, the dependent variable is log (plus one) of mid-season hired hours. In column four, the dependent variable is log (plus one) of total materials (Rs) allocated to the plot in that month. Regressions also control for total rainfall, rainfall squared, area planted, area irrigated, total planting materials used, total planting hours, total materials used in all months prior, and total labor allocated in all months prior. In all columns, the effects of these variables are allowed to vary by month of the year.

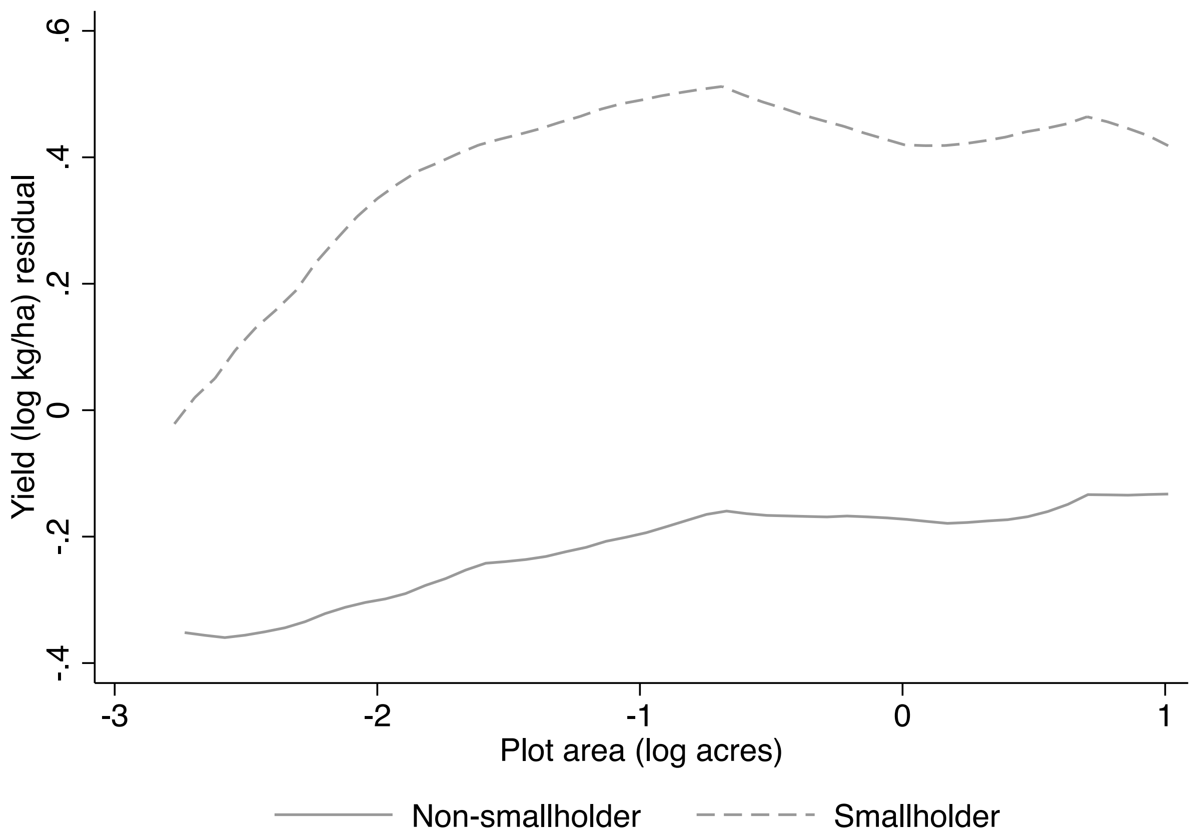
* p<0.1 ** p<0.05 *** p<0.01

Table 6: Accounting for Changes in Time Use of Household Members

	Own farm days		Wage days		Other days		Inv. Unemp days		NF work days (wage + other)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	S	N	S	N	S	N	S	N	S	N
Monthly weighted crop price (log R)	0.028***	0.006*	−0.006	0.003	0.001	0.002	−0.012**	0.003	−0.006	0.004
	(0.010)	(0.003)	(0.009)	(0.003)	(0.003)	(0.001)	(0.006)	(0.002)	(0.009)	(0.003)
Age	0.102***	0.054**	0.078***	0.047***	0.014	0.015***	0.055***	0.010	0.088***	0.061***
	(0.024)	(0.022)	(0.030)	(0.015)	(0.012)	(0.004)	(0.020)	(0.007)	(0.031)	(0.015)
Age squared	−0.001***	−0.001***	−0.001***	0.000***	0.000***	0.000***	−0.001**	0.000	−0.001***	−0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Fixed Effects:										
District-year-month-crop	X	X	X	X	X	X	X	X	X	X
Individual-crop	X	X	X	X	X	X	X	X	X	X
Previous labor, previous materials, and rainfall:										
By month	X	X	X	X	X	X	X	X	X	X
Observations	14,128	30,746	14,128	30,746	14,128	30,746	14,128	30,746	14,128	30,746

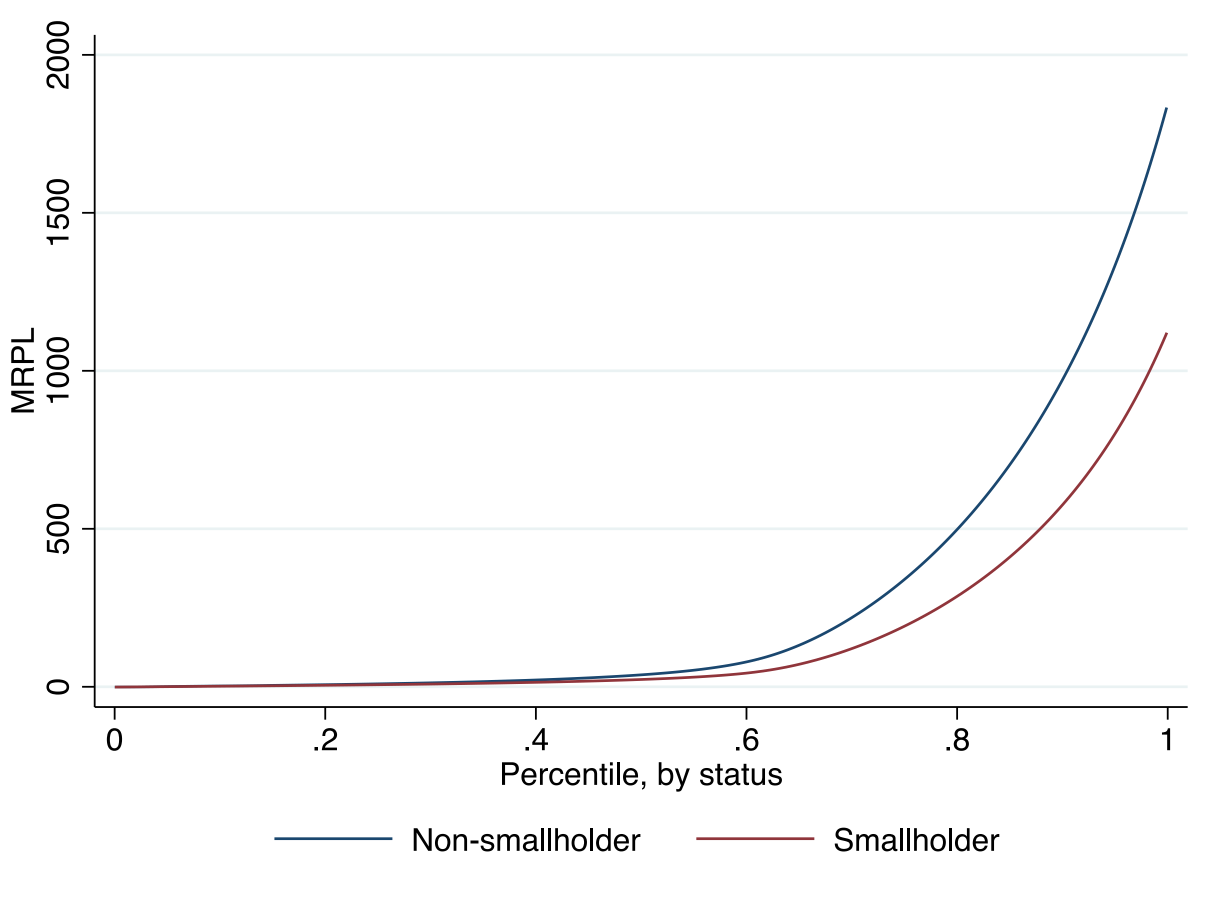
Observations are at the individual-month level and standard errors are clustered at the individual level. The dependent variable in each column is transformed using an inverse hyperbolic sine transformation. Odd-numbered columns are individuals in smallholder households and even-numbered columns are individuals in non-smallholder households. The dependent variable in each column is indicated in the column title. The main dependent variable, *monthly weighted crop price*, is average crop price faced by the household, weighted by the amount of land allocated to each crop. Regressions also control for total rainfall, rainfall squared, area planted, area irrigated, total planting materials used, total planting hours, total materials used in all months prior, and total labor allocated in all months prior. In all columns, the effects of these variables are allowed to vary by month of the year. In all columns, age and age squared are additional covariates.

Figure 2: Output per hectare and landholdings



The figure shows the relationship between the residual of yield and plot size, with size restricted to the middle 90 percent of smallholder plot sizes in order to compare in areas of common support.

Figure 3: Estimated Marginal Revenue Product of Labor, by percentiles



The figure shows estimated MRPL by percentile, separately for smallholders and non-smallholders.

Appendix A

Table A1: Cross-Section vs. Within Variation - Prices and Output

	(1)	(2)	(3)	(4)
Harvest price (log Rs)	−0.429***	−0.390***	−0.250***	−0.042
	(0.070)	(0.071)	(0.069)	(0.080)
Planting price (log Rs)	−0.026	−0.024	0.009	−0.004
	(0.072)	(0.071)	(0.073)	(0.095)
Wave-season FE		X		
District-wave-season FE			X	
District-wave-season-crop FE				X
Observations	10,534	10,534	10,534	10,534

Observations are at the household-plot-crop-season level. Regressions also control for total rainfall, rainfall squared, area planted, area irrigated, total planting materials used, and total planting hours.

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

Table A5: Input Allocation and Crop Price Elasticities, Smallholder Defined at 2 Acres

	(1) All	(2) Family	(3) Hired	(4) Materials
Monthly crop price (log R)	0.406***	0.259**	0.539***	0.576
	(0.124)	(0.129)	(0.139)	(0.480)
Monthly crop price times Smallholder	−0.230**	0.175	−0.604***	−0.605**
	(0.105)	(0.123)	(0.153)	(0.240)
Fixed Effects:				
District-Year-Month-Crop	X	X	X	X
Household-Plot-Crop	X	X	X	X
Previous labor, previous materials, and rainfall:				
By month	X	X	X	X
Observations	24,318	24,318	24,318	24,318

Observations are at the household-year-month-plot level. In column one, the dependent variable is log (plus one) of all mid-season labor hours, defined as any labor between planting and harvest in each month. In column two, the dependent variable is log (plus one) of mid-season family hours. In column three, the dependent variable is log (plus one) of mid-season hired hours. In column four, the dependent variable is log (plus one) of total materials (Rs) allocated to the plot in that month. Regressions also control for total rainfall, rainfall squared, area planted, area irrigated, total planting materials used, total planting hours, total materials used in all months prior, and total labor allocated in all months prior. In all columns, the effects of these variables are allowed to vary by month of the year.

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

Table A2: Household Type and Predictive Power of Lagged Prices

	(1) Non Smallholders	(2) Smallholders	(3) All	(4) All
			Non-smallholders	
Lagged (x1) monthly crop price (log R)	0.280***	0.342***	0.317***	0.352***
	(0.037)	(0.033)	(0.028)	(0.022)
Lagged (x2) monthly crop price (log R)				0.103***
				(0.020)
			Smallholders	
Lagged (x1) monthly crop price (log R)			0.325***	0.361***
			(0.028)	(0.024)
Lagged (x2) monthly crop price (log R)				0.098***
				(0.020)
Fixed Effects:				
District-year-month-crop	X	X	X	X
Household-Plot-Crop	X	X	X	X
Inputs and rain by month	X	X	X	X
F-test (non-smallholder = lag smallholder):			Lag (x1)	
F			0.740	1.100
(p-value)			(0.390)	(0.295)
F-test (non-smallholder = lag smallholder):			Lag (x2)	
F				0.528
(p-value)				(0.468)
Observations	13,324	6,743	20,067	19,320

In all columns, the dependent variable is the current crop price of the crop planted on the plot. "Lagged (x1)" indicates the price of that same crop in the previous month and "Lagged (x2)" indicates the price of that same crop two months prior. The F-tests test the null hypothesis that lagged crop prices predict current crop prices the same for both smallholders and non-smallholders.

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

Table A3: Household Size and Planted Plot Area

	(1)	(2)	(3)	(4)
Household size	0.038***	0.016*	0.018***	0.013*
	(0.010)	(0.009)	(0.007)	(0.006)
Fixed Effects:				
Wave-Crop-Season-Village FE	X	X	X	X
Household FE		X		
Household-Plot FE			X	
Household-Plot-Crop FE				X
Observations	10,534	10,534	10,534	10,534

The dependent variable in all columns is the log of planted area on the individual plot.

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

Table A4: Labor Allocation and Household Demographics

	(1) Log hhsize (with pcts)	(2) Logs (All)	(3) IHS (All)
		Non-smallholders	
Household size/child	−0.005	−0.039	−0.027
	(0.111)	(0.066)	(0.051)
Prime male	0.058	0.151	0.115
	(0.252)	(0.117)	(0.093)
Prime female	−0.278	−0.340**	−0.252**
	(0.383)	(0.147)	(0.111)
Elderly male	−0.090	−0.004	−0.002
	(0.310)	(0.070)	(0.054)
Elderly female	0.403	−0.021	−0.017
	(0.579)	(0.158)	(0.124)
		Smallholders	
Household size/child	0.027	0.067	0.052
	(0.132)	(0.072)	(0.056)
Prime male	−0.227	0.040	0.033
	(0.246)	(0.127)	(0.098)
Prime female	1.144***	0.531***	0.395***
	(0.414)	(0.191)	(0.144)
Elderly male	0.014	0.094	0.074
	(0.238)	(0.072)	(0.057)
Elderly female	−0.557	−0.134	−0.105
	(1.317)	(0.426)	(0.336)
F-tests (all demographics = 0):		Non-smallholders	
F	0.580	1.372	1.300
p-value	(0.715)	(0.232)	(0.262)
		Smallholders	
F	2.161*	1.999*	1.942*
p-value	(0.056)	(0.076)	(0.085)
Observations	10,524	10,524	10,524

In all columns, the dependent variable is the total number of hours allocated to agricultural production, including hired and family labor but excluding harvest labor. Observations are at the household-year-season-plot level. In column one, household size is defined as the log of total household size and each demographic variable is defined as a percentage of the household, with children being the omitted category. In columns two and three, the household size coefficient refers to the coefficient on children only. In column two, all demographic variables are defined as the natural log of the number of persons plus one. In column three, all demographic variables are defined as the inverse hyperbolic sine transformation ($ln(x + \sqrt{x^2 + 1})$) of persons.

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

Table A6: Input Allocation and Crop Price Elasticities, Smallholder Defined at 4 Acres

	(1) All	(2) Family	(3) Hired	(4) Materials
Monthly crop price (log R)	0.404***	0.298**	0.546***	0.341
	(0.124)	(0.137)	(0.141)	(0.559)
Monthly crop price times Smallholder	−0.148*	0.021	−0.425***	0.166
	(0.088)	(0.104)	(0.121)	(0.269)
Fixed Effects:				
District-Year-Month-Crop	X	X	X	X
Household-Plot-Crop	X	X	X	X
Previous labor, previous materials, and rainfall:				
By month	X	X	X	X
Observations	24,318	24,318	24,318	24,318

Observations are at the household-year-month-plot level. In column one, the dependent variable is log (plus one) of all mid-season labor hours, defined as any labor between planting and harvest in each month. In column two, the dependent variable is log (plus one) of mid-season family hours. In column three, the dependent variable is log (plus one) of mid-season hired hours. In column four, the dependent variable is log (plus one) of total materials (Rs) allocated to the plot in that month. Regressions also control for total rainfall, rainfall squared, area planted, area irrigated, total planting materials used, total planting hours, total materials used in all months prior, and total labor allocated in all months prior. In all columns, the effects of these variables are allowed to vary by month of the year.

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

Table A7: Plot-Level Monthly Labor Allocation and Village Fixed Effects

	(1) All	(2) Family	(3) Hired	(4) Materials	(5) All	(6) All
Monthly crop price times Smallholder	−0.168*	0.086	−0.248**	−0.451***	−0.246	−0.357*
	(0.098)	(0.096)	(0.125)	(0.162)	(0.177)	(0.190)
Next month's crop price times Smallholder					0.057	0.072
					(0.106)	(0.116)
Fixed Effects:						
Village-year-month-crop	X	X	X	X	X	
District-year-month-crop						X
Household-Plot-Crop	X	X	X	X	X	X
Inputs and rain by month	X	X	X	X	X	X
Previous labor, previous materials, and rainfall:						
By month	X	X	X	X	X	X
Observations	24,318	24,318	24,318	24,318	18,297	18,297

The table presents additional specifications for Table 5. Instead of district-year-month-crop fixed effects, however, the specifications include village-year-month-crop fixed effects in columns one through five. The level effect of month crop price is thus not identified, but the difference across household types is. In columns five and six, the specification also includes next month's crop price.

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

Table A8: Plot-Level Monthly Labor Allocation - No Lagged/Planting Variables

	(1) All	(2) Family	(3) Hired	(4) Materials
Monthly crop price (log R)	0.320***	0.233**	0.496***	1.041***
	(0.100)	(0.115)	(0.149)	(0.335)
Monthly crop price times Smallholder	−0.210**	0.014	−0.262**	−0.682***
	(0.104)	(0.102)	(0.127)	(0.185)
Fixed Effects:				
District-Year-Month-Crop	X	X	X	X
Household-Plot-Crop	X	X	X	X
Previous labor, previous materials, and rainfall:				
By month	X	X	X	X
Observations	24,318	24,318	24,318	24,318

The table presents additional specifications for Table 5. All planting and lagged input variables are removed in all columns.
* p<0.1 ** p<0.05 *** p<0.01

Table A9: Accounting for Changes in Time Use of Household Members - Levels

	Own farm days		Wage days		Other days		Inv. Unemp days		NF work days (wage + other)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	S	N	S	N	S	N	S	N	S	N
Monthly weighted crop price (log R)	0.131***	0.023	−0.039	0.021	−0.002	0.004	−0.045**	0.014	−0.041	0.025
	(0.041)	(0.014)	(0.057)	(0.019)	(0.009)	(0.003)	(0.022)	(0.010)	(0.056)	(0.020)
Age	0.513***	0.313***	0.465**	0.324***	0.020	0.034***	0.207**	0.035	0.485**	0.358***
	(0.118)	(0.105)	(0.201)	(0.097)	(0.038)	(0.010)	(0.083)	(0.029)	(0.202)	(0.098)
Age squared	−0.007***	−0.006***	−0.007***	−0.003***	−0.001**	−0.001***	−0.002**	0.000	−0.008***	−0.003***
	(0.001)	(0.001)	(0.002)	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)	(0.002)	(0.001)
Fixed Effects:										
District-year-month-crop	X	X	X	X	X	X	X	X	X	X
Individual-crop	X	X	X	X	X	X	X	X	X	X
Previous labor, previous materials, and rainfall:										
By month	X	X	X	X	X	X	X	X	X	X
Observations	14,128	30,746	14,128	30,746	14,128	30,746	14,128	30,746	14,128	30,746

Observations are at the individual-month level and standard errors are clustered at the individual level. Odd-numbered columns are individuals in smallholder households and even-numbered columns are individuals in non-smallholder households. The dependent variable in each column is indicated in the column title. The main dependent variable, *monthly weighted crop price*, is average crop price faced by the household, weighted by the amount of land allocated to each crop. Regressions also control for total rainfall, rainfall squared, area planted, area irrigated, total planting materials used, total planting hours, total materials used in all months prior, and total labor allocated in all months prior. In all columns, the effects of these variables are allowed to vary by month of the year. In all columns, age and age squared are additional covariates.

Table A10: Season Price Changes, Output, and Labor Allocation

	Non-smallholder			Smallholder		
	(1)	(2)	(3)	(4)	(5)	(6)
	Labor	Labor	Output	Labor	Labor	Output
Harvest price (log Rs)	0.272*	0.209		0.062	0.019	
	(0.159)	(0.171)		(0.133)	(0.144)	
Planting price (log Rs)		0.226			0.236	
		(0.188)			(0.166)	
Mid-season hours (log)			0.030*			0.070
			(0.018)			(0.050)
District-wave-season-crop FE	X	X		X	X	
Village-wave-season-crop FE			X			X
Household-Plot-Crop FE	X	X	X	X	X	X
Observations	6,957	6,957	6,957	3,577	3,577	3,577

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