Should Farmers Farm More? Comparing Marginal Products within Malawian Households*

Peter Brummund† **Joshua D Merfeld**‡

August 30, 2018

Abstract

According to standard economic theory, households should equate the marginal revenue product of an input across activities within the household. Using data on agricultural plots and non-farm enterprises in Malawi, we test whether the marginal revenue product of labor (MRPL) is equal across agricultural and non-farm production within a household. We control for many household and production characteristics, including household fixed effects, and find agricultural MRPL to be consistently higher than non-farm MRPL. These results hold when restricting estimation to periods of high and low non-farm labor allocation. Further analyses suggest both production and price risk may be partially responsible for these deviations from MRPL equality. We also show that comparing average revenue products of labor is not sufficient evidence for claiming allocative inefficiency. In our data, the average revenue products of labor lead to an erroneous conclusion, reiterating the need to compare marginal products when testing for allocative efficiency.

Keywords: Labor productivity, agriculture, non-farm production, risk, efficiency
JEL Codes: J24, J43, O13, R23

---

*This paper was previously circulated under the title "Allocative Efficiency of Labor in Agricultural Households: Evidence from Malawi." We thank Brian Dillon, Daniel LaFave, and Mark Long for comments on earlier drafts that greatly improved this paper. We appreciate helpful comments from participants at the 2015 Structural Transformation of African Agriculture and Rural Spaces (STAARS) Conference and the 2016 Northeast Universities Development Consortium (NEUDC) Conference. We gratefully acknowledge the Institute for the Study of Labor (IZA) and DFID's Growth and Labour Markets in Low Income Countries Programme (GLM | LIC) for funding the research leading to this paper. The views expressed herein and any remaining errors are our own. Declarations of interest: none
†Department of Economics, Finance and Legal Studies, University of Alabama; pbrummund@cba.ua.edu
‡Robert F. Wagner Graduate School of Public Service, New York University, 295 Lafayette Street, 3040B, New York, NY, 10012; Telephone: +1 (212) 998-7457; merfeld@nyu.edu

1 Introduction

A stylized fact of the development process is that agriculture's share of GDP decreases as a country develops (Lewis, 1954; Ranis and Fei, 1961). This relationship holds in the cross-section, with relatively more developed countries deriving a smaller percentage of GDP from agricultural sources (Chenery et al., 1975; Gollin et al., 2014). Moreover, even within countries, the non-farm sector tends to be more productive – as measured by the average revenue product of labor – than the agricultural sector (Gollin et al., 2014; McCullough, 2017; Young, 2013). Given these persistent empirical patterns, it is perhaps no surprise that many development policies focus on non-farm growth; the recent microfinance revolution is but one example of this (Armendáriz and Morduch, 2010).

However, a higher average revenue product in the non-farm sector does not necessarily imply a reallocation of labor is warranted, since theory predicts households should be equating marginal revenue products of labor (MRPL), not average revenue products. Whether the marginal revenue product suggests the same depends on the reasons that households operate non-farm and agricultural enterprises simultaneously. If households operate non-farm enterprises to protect themselves against agricultural production risk, for example, MRPL equality need not hold, even in rational households. A large body of research shows that agricultural households diversify into non-farm self-employment for a number of reasons, including to insure against production shocks or household shocks, or to accommodate agricultural seasonality or missing markets (Barrett et al., 2001; Haggblade et al., 2010; Lanjouw and Lanjouw, 2001; Merfeld, 2018b; Nagler and Naudé, 2014). In addition, diversification seems to be the norm, not the exception (Davis et al., 2017). Under these scenarios, households may not be moving into the non-farm sector chasing profits. Rather, they may instead be pushed into the sector due to a lack of more remunerative options and a desire to mitigate risk, leading to a lower marginal revenue product of labor in non-farm production. This paper revisits the question of allocative efficiency, focusing on the marginal revenue product, as opposed to the average revenue product. We show that the empirical choice between the two measures is a decisive one.

Regardless of the exact motivation for diversification, standard economic theory of profit maximization predicts the equality of the marginal revenue product of labor across productive activities – in the current context, agricultural and non-farm production – as well as the equality of MRPL with the market wage; if labor were allocated in any other way, it would be possible to increase profits by reallocating labor. However, this result generally relies on the assumption of complete and competitive markets, which recent work has shown to not hold across many developing countries (e.g. Dillon et al. (2017); LaFave and Thomas (2016)), as well as the assumption of a collective household model, which is also questionable (Merfeld, 2018a; Guirkinger et al., 2015; Udry, 1996; Walther, 2018). As such, whether households equate marginal revenue products across productive activities is an empirical question. Moreover, the answer to this question is not only interesting in its own right, but is integral to labor supply estimation (Abdulai and Regmi, 2000; Barrett et al., 2008; Jacoby, 1993; Seshan, 2014; Skoufias, 1994) and can even shed light on some of the underlying market conditions which characterize production in rural areas of developing countries. This, in turn, may help us better understand why households diversify into non-farm production and develop more appropriate development interventions. In this paper, we test this assumption of MRPL equality using household survey data from Malawi. To the best of our knowledge, this is the first test of MRPL equality across agricultural and non-farm production within households. At first glance, our results show that this common assumption fails for the median household. However, we go on to show that production risk and price risk can help explain why households are making the allocation decisions they do.

The assumption of market completeness is challenged by a vast literature. Under complete markets, agricultural households are able to treat the production and consumption decisions as recursive; households first maximize expected profit in production before making consumption decisions. This result, known as separation, suggests simple tests for market completeness. While early research was unable to reject this hypothesis (Benjamin, 1992), more recent research suggests otherwise (Dillon and Barrett, 2017; Dillon et al., 2017; LaFave and Thomas, 2016). This finding casts doubt on the assumption of complete markets that drives separation in the agricultural household model. As such, deviations from MRPL equality need not suggest irrational behavior on the part of households. In particular, differing risk profiles of production can lead to deviations from MRPL equality (Barrett et al., 2008; Stiglitz, 1974). If production risk differs across activities, households optimize by equalizing their expected marginal utilities across activities. Additionally, price risk – uncertainty over the market price for a good – can also affect MRPL equality. Barrett (1996) shows that price risk can affect households differently since it is likely to be correlated with production risk. In particular, households are predicted to behave differently depending on whether they are net buyers or net sellers of crops, with net sellers more likely to exhibit what we traditionally associate with risk: an underallocation of labor to the risky activity. The interplay of these two types of risk thus predicts substantial heterogeneity, a prediction confirmed in this paper.

To test this assumption of equality of the marginal revenue products of labor across household activities, we use three waves of the Malawi Integrated Household Survey (IHS). Without making further assumptions regarding market completeness, equality is only predicted for households that operate both agricultural and non-farm enterprises simultaneously. As such, we begin with summary statistics comparing households that operate both types of enterprises in the same wave to all households. The results do show the relevant subsample to be statistically different from the overall sample, limiting the external validity of the analysis.

We then examine whether the marginal revenue product of labor is equal across agricultural and non-agricultural production within a household. To the best of our knowledge, this is the first paper to implement such a test. The choice to compare marginal revenue products is important, as the comparison of average revenue products of labor leads to an erroneous conclusion. At first glance, our results show that agricultural MRPL is consistently higher than non-farm MRPL. This results holds under a variety of specification choices and sample restrictions.

However, important characteristics about the environment in which these households operate influence the allocation decisions of these households. We find evidence that price and production risk play an important role in household labor allocation. In particular, deviations from equality are much higher for plots planted with tobacco and cotton, two pure cash crops, than for plots planted with maize, a common subsistence crop in Malawi. This is consistent with households dealing with price risk by overallocating labor to crops of which they are net buyers and underallocating labor to crops of which they are net sellers.

We also examine the role of price risk in allocation decisions by using revenue, acreage, and crop sales as proxies for market access. For all three variables, households above the median value show substantially larger MRPL deviations than households below the median. For revenue, we are unable to reject the null hypothesis of MRPL equality for households with below median total revenue. Similarly, the MRPL difference is around twice as large for the upper half of the distribution (relative to the lower half) of both acreage and crop sales. Insofar as these variables are proxies for market access, this evidence is consistent with price risk influencing allocation decisions if access to markets is correlated with the probability of being net sellers of crops (Barrett, 1996).

Households which hire labor from the market are also more sensitive to fluctuations in output prices, and should therefore have higher MRPLs than households which don't hire labor from the market. We find exactly this pattern for households which hire for agricultural production, and for households which hire for non-farm production.

Consistent with production risk, we also find evidence that rainfall variability is associated with higher agricultural MRPL and suggestive evidence that it is also associated with deviations from equality. We only have rainfall variables for the first and second waves; higher rainfall variation (as proxied by the rainfall coefficient of variation, CV) is positively correlated with deviations from equality in these two waves. Overall, these results reinforce the commonly held belief that agriculture is risky and that farmers may deviate from profit-maximizing conditions to deal with this risk. In other words, reducing the risk faced by households could theoretically increase their expected (mean) incomes.

Our results are also relevant to the literature examining the "productivity gap" between the non-farm and agricultural sectors in developing countries (Gollin et al., 2014; Hicks et al., 2017; McCullough, 2017; Young, 2013). These studies all compare the average productivity of labor across non-farm and agricultural sectors, finding the non-farm sector to have consistently higher levels of average productivity (though to different degrees depending on the specific study). Our study compares both average productivity and marginal productivity, and finds the two measures lead to different conclusions. Even though we are analyzing allocative efficiency within households and not across sectors, our results show that it is imperative to compare marginal products when testing for allocative efficiency.

The rest of this paper is organized as follows. In section 2, we present the theory that informs our study. We discuss methodology and summary statistics in section 3. We present results of our analyses in section 4 before concluding in section 5.

2 Theory

We begin by elaborating a simple model, assuming no risk. We use Barrett (1996) as the starting point for our model. The household's problem is:

$$\max_{N,S,h,L_f,L_n,L_M,z_f,z_n} \quad u(N,S,h) \tag{1}$$

subject to

$$\begin{aligned} N + p_S S \leq & Y^* \\ Y^* = & p_S f(L_f, z_f; k_f) - p_{z_f} z_f + p_n n(L_n, z_n; k_n) - p_{z_n} z_n + w L_m \\ \bar{T} \geq & h + L_f + L_n + L_m, \end{aligned}$$

where Y^* is endogenous income; N is a non-staple good, available only through market purchase, with price normalized to one; p_S is the price of a staple good, S, available either through the market or through home production, with production function $f(\cdot)$, which in satisfies the common assumptions of differentiability and concavity; L_f is days of agricultural labor; h is leisure; z_f is other agricultural variable inputs; k_f is a vector of exogenous characteristics affecting agricultural output, including soil quality and weather; p_n is the price of the non-farm enterprise output, whose production function is given by $n(\cdot)$, which also satisfies the common assumptions; L_n is non-farm enterprise labor; k_n is exogenous characteristics affecting non-farm output; w is the market wage; L_m is days worked on the market; and $\bar{T}$ is the total time endowment. Note that we omit hired labor for simplicity.[1]

At an interior solution – that is, for households that operate both types of enterprises and

[1] In our data, only about 13 percent of agricultural plots and only 11 percent of non-farm enterprises reported hiring any labor. In the empirical specifications, we pool hired and family labor.

work on the market – both constraints will bind. The first order conditions are:

$$p_S \frac{\partial f}{\partial L_f} = p_n \frac{\partial n}{\partial L_n} = w \tag{2}$$

Focusing on labor, this is the standard treatment in the literature: at the optimum, the marginal revenue product of labor is equated across all activities and is equal to the market wage, assuming the individual is active on the market (or has the option to be).

However, if an individual is limited in their ability to work on the market – perhaps due to labor market frictions – then the above equalities will no longer include the wage rate, but rather the shadow wage,

$$w^* = w + \frac{\lambda_B}{\lambda_T}, \tag{3}$$

where λ_B is the Lagrange multiplier on the budget constraint and λ_T is the multiplier on the time constraint (Barrett et al., 2008). This again assumes the household operates at least one non-farm and one agricultural enterprise. In this scenario, the marginal revenue products of labor across activities within the household are still equated under the most common assumptions. That is,

$$p_f \frac{\partial f}{\partial L_f} = p_n \frac{\partial n}{\partial L_n}. \tag{4}$$

It is this prediction of equality of marginal revenue products of labor across activities within the household that we test in this paper.

This result need not hold in the presence of risk. Here we discuss two types of risk that affect input allocation: production risk and price risk. Production risk is the risk associated with uncertainty in production. While farmers, for example, make agricultural decisions –

planting, applying fertilizer, allocating labor, etc. – based on what they expect to happen, idiosyncratic shocks can cause actual outcomes to deviate from expected outcomes. In the classic treatment of production risk, households will underallocate labor to the relatively riskier activity, causing MRPL in that activity to be higher than MRPL in the less-risky activity. This drives a wedge between MRPL across the two different activities but, importantly, is a rational response to risk. While households may not be maximizing expected profit, they are nonetheless maximizing expected utility.

Price risk, on the other hand, relates to uncertainty regarding the price of agricultural outputs following harvest. Under these circumstances, the above model needs to be amended. To see this, we extend the model from Barrett (1996). In a two-period model, households make labor allocation decisions before post-harvest prices are revealed. The household's problem is

$$\max_{h,L_f,L_n,L_M,z_f,z_n} \quad E\left(\max_{N,S} u(N,S,h)\right)$$

subject to

$$\begin{aligned} N + p_S S \leq& Y^* \\ Y^* =& p_S f(L_f, z_f; k_f) - p_{z_f} z_f + p_n n(L_n, z_n; k_n) - p_{z_n} z_n + wL_m \\ \bar{T} \geq& h + L_f + L_n + L_m, \end{aligned}$$

where all variables are defined as in Equation 1. In this model, households must make all labor allocation decisions prior to the realization of prices. By duality, we can instead work with the indirect utility function, $V(h, Y^*, P)$ – where P is a vector of all prices – which is homogeneous of degree zero in income and prices. The household's problem is now,

$$\max_{h,L_f,L_n,L_M,z_f,z_n} E\left(V(h,Y^*,P)\right) \tag{5}$$

subject to

$$Y^* = p_S f(L_f,z_f;k_f) - p_{z_f} z_f + p_n n(L_n,z_n;k_n) - p_{z_n} z_n + w(\bar{T} - L_f - L_n - h)$$

Here, we present only the main result. Interested readers are directed to Barrett (1996) for a full exposition.

Focusing on farm labor, Barrett (1996) shows that $cov(V_Y, P_S)$ is crucial to predicting labor allocation. In pure producer theory, Arrow-Pratt income risk aversion implies $cov(V_Y, P_S) < 0$ (Sandmo, 1971) and, thus, that a household under uncertainty produces less – and thus applies less labor and has a higher MRPL – than would be the case without uncertainty (Barrett, 1996). However, this result focuses only on pure producers and does not take into account a household model of production, in which households can sell and *consume* the agricultural goods they produce. The insight of Barrett (1996) is that labor allocation hinges on the importance of the crop to households as a consumption good relative to the importance of the crop as a source of income (cash). In particular, he shows that if a household is a net *seller* of a crop, then it will underemploy labor: $MRPL_{ag} > w$. On the other hand, if a household is a net *buyer* of a good, then it will overemploy labor in the presence of risk, so that $MRPL_{ag} < w$.

Note that, by assumption, households do not consume their non-farm good. Thus, non-farm production will follow the familiar result if non-farm production is risky: $MRPL_{nf} > w$. However, if non-farm production is a relatively risk-free activity, then $MRPL_{nf} = w$. Putting together these predictions, we can say, for net buyers of crops, that: $MRPL_{ag} <$

$MRPL_{nf}$. However, for net sellers of crops, the direction is ambiguous and depends on the relative risk and importance of each productive activity.

3 Methods and Data

The basic steps involved in testing MRPL equality are as follows, and are similar to those in Linde-Rahr (2005), who studies allocative efficiency across plots planted with different crops. First, we estimate production functions for both agricultural and non-farm enterprises. Second, we compute the marginal revenue products of labor across different activities within the household. With these MRPL estimates, we then explicitly test whether marginal revenue products of labor – or shadow wages – are equal across activities within the household.

We present production function results using both a Cobb-Douglas production function and a translog production function. However, given that we reject the nested Cobb-Douglas within the translog production function, we present MRPL results only for translog production function estimates. The translog specification we estimate is:

$$\begin{aligned}\ln R_{iht} = \alpha_h + I(\mathit{Agriculture} = 1) \times (\sum_j \beta_j \ln \gamma_{jiht} + \frac{1}{2}\sum_j\sum_k \beta_{jk} \ln \gamma_{jiht} \ln \gamma_{kiht} + \delta C_{iht} \\ + D_{dt} + \eta_m) + \sum_j \beta_j \ln \gamma_{jiht} + \frac{1}{2}\sum_j\sum_k \beta_{jk} \ln \gamma_{jiht} \ln \gamma_{kiht} + \delta C_{iht} + D_{dt} + \eta_m \\ + I(\mathit{Agriculture} = 1) + \varepsilon_{iht},\end{aligned} \tag{6}$$

where α_h is an intercept that varies by household (household fixed effects); $I(\mathit{Agriculture} = 1)$ is a dummy variable equal to one if the observation is an agricultural plot and equal to zero if the observation is a non-farm enterprise; γ_{jiht} and γ_{kiht} are inputs j and k for enterprise *i* in household *h* in wave *t*; C_{iht} is a vector of controls that may affect revenue and which differ depending on whether the enterprise is (non-)agricultural; D_{dt} is district/wave fixed effects; η_m is a set of dummy variables indicating the month of interview; and ε_{iht} is a conditional mean-zero error term. We include labor (log of days), acres (log), and fertilizer (log of kg) as productive inputs in the agricultural production functions and labor (log of days) and total costs (log of MWK) as productive inputs in the non-farm production functions. We set land and fertilizer to zero for all non-farm enterprises and non-farm costs to zero for all agricultural plots.

For agricultural plots, we restrict attention to plots planted with a select number of crops. Part of the process of constructing the marginal revenue products involves finding prices for the agricultural output. We construct these prices by taking medians at the lowest administrative level of aggregation with a sufficient number of observations. The crops we use have a sufficient number of observations with which to create median prices. In contrast, we do not need to restrict estimation in a similar way for non-farm enterprises; entrepreneurs are directly asked about their total revenue. Finally, we trim the top one percent of agricultural and non-farm revenue and labor before estimating production functions.[2]

For both the Cobb-Douglas and translog specifications, we pool the data and estimate a single production function for both agricultural and non-farm enterprises, but we allow the effect of all variables – other than the fixed effect – to vary by type of enterprise. We employ household fixed effects to identify the production functions. Since the identifying assumptions for a pooled sample are more restrictive, we also estimate separate agricultural and non-farm production functions.

After estimating the production functions, we then calculate the marginal revenue product of labor for each input. Using the translog specification, we construct our MRPL estimates

[2]We report non-trimmed results in Appendix Table A4. Results are unchanged.

for agricultural plots as

$$\frac{\partial R}{\partial L} = \frac{\hat{R}_{iht}}{L_{iht}} \left[\beta_L + \beta_{LL} \log L_{iht} + \beta_{LA} \log A_{iht} + \beta_{LF} \log F_{iht}\right], \tag{7}$$

where $\hat{R}_{iht}$ is predicted revenue, β_{LL} is the coefficient on the labor squared term, β_{LA} is the coefficient on the interaction between labor and acreage, and β_{LF} is the coefficient on the interaction between labor and fertilizer use. We use predicted revenue as it is the best estimate we have of the farmer's expected revenue. In other words, predicted revenue removes idiosyncratic risk and is the target farmers use when making labor allocation decisions.[3] We calculate the MRPL for non-farm enterprises as

$$\frac{\partial R}{\partial L} = \frac{\hat{R}_{iht}}{L_{iht}} \left[\beta_L + \beta_{LL} \log L_{iht} + \beta_{LC} \log C_{iht}\right], \tag{8}$$

where β_{LC} is the coefficient on the interaction between labor and costs.

The production function in Equation 6 is at the plot/enterprise level. However, after estimating MRPLs, we need to aggregate to the household level in order to compare MRPLs across productive activities. This raises questions about the best way to aggregate multiple MRPLs, since many households operate more than one plot or more than one enterprise in the same wave. We aggregate MRPLs to the household level in two ways. First, we take the simple median across plots within each household. If a household operates an even number of plots, we use the mean of the two middle plots to construct a household average MRPL. We construct the household average MRPL similarly for non-farm enterprises, though most households operate only one. Second, we compute the household's weighted average MRPL, weighting by labor allocation across plots or enterprises.

Aggregating over plots and enterprises is an important part of MRPL estimation. How-

[3] As a robustness check, we use actual revenue in the calculation of MRPLs and do not find signifiicantly different results. These results are shown in Table A5 of the Appendix.

ever, it is not clear what the "correct" aggregation method is. To bypass this problem completely, and due to issues with recall bias raised by Arthi et al. (2018),[4] we also estimate production functions in which we aggregate (separately) all plots and all enterprises to the household level. We sum revenue, labor, acreage, fertilizer, and non-farm costs to the household level. For crop dummies, we collapse the data and leave these as indicator variables for whether the household grew that crop in that wave. For the plot characteristics, we include continuous variables for the percent of total household land with each characteristic. We then estimate production functions at the household level, leading to one MRPL for each household enterprise.

MRPL equality holds only for households that operate both types of enterprises. Therefore, we only use households that operate both non-farm and agricultural enterprises in the same wave to construct MRPL estimates. However, we present estimates of production functions using these households as well as separate estimates using all households in order to examine the external validity of our final sample.

Finally, in order to conduct inference over a multi-step estimator, we bootstrap the standard errors with 1,000 replications. Since we employ household fixed effects, we set up the bootstrap to draw households, including all non-farm and agricultural enterprises operated by that household, across all waves.

3.1 Data

We use the Malawi Integrated Household Survey (IHS) in this paper. The IHS is part of the World Bank's Living Standards Measurement Study (LSMS) program. The IHS data consists of three waves: the first wave was collected from 2010-2011, the second in 2013, and the third in 2016-2017. There is a common panel sample in all three waves, such that

[4] Arthi et al. (2018) find that reported labor allocation is mismeasured at the plot level. However, when labor is aggregated to the household, this mismeasurement largely disappears.

we are able to follow some households across three separate years. However, the first and third waves also had large cross-section components, which dwarf the panel component in size. After restricting our sample and looking only at households that operate both types of enterprises in the same wave, we are left with 3,786 unique households. Of these, 3,506 households appear in one wave, 482 households appear in two waves, and 117 households appear in all three waves. These numbers correspond to 5,873 plots and 3,750 enterprises in the one-wave households, 881 plots and 547 enterprises in the two-wave households, and 250 plots and 129 enterprises in the three-wave households. Thus, our total sample with which we estimate production functions consists of 7,004 plots and 4,426 enterprises.

Many of the variables used in estimation come from different modules in the IHS. As such, certain variables are at different levels of aggregation (i.e. household, household-plot, or household-plot-crop). For example, although crop output is reported at the plot-crop level, labor is only reported at the plot level. Thus, missing plot-crop observations require that we drop the entire plot from the sample. Additionally, we drop any non-farm enterprises that were in operation for less than six months in the previous year and any non-farm enterprises that were not in operation in the month prior to data collection. In the following sub-sections, we document some of these idiosyncracies and the resulting decisions.

3.2 Revenue

The key dependent variable in the production functions is revenue. The Malawi IHS, like many household surveys, asks farmers for agricultural output in weight, not in value. Since our methodology is designed to estimate the marginal revenue product of labor, we must construct crop prices to value output. We construct prices from sales information collected from households which sold crops.

However, most households do not report selling crops. As such, we impute prices based on the most local median price possible. We construct aggregate prices at four separate levels – the enumeration area, the traditional authority, the district, and the region – by taking the median crop price at each level of aggregation separately. We then assign prices to households using the lowest level of aggregation at which there are at least five valid price observations.[5] If any regions have less than five prices observations for any given price, we assign a missing price value for all observations of that crop in that district. In this way, we are able to construct prices for 13 crops: maize, tobacco, groundnut, rice, sweet potato, potato, beans, soya, pigeon peas, cotton, sunflower, nkhwani, and tomato.

3.3 Labor and Selection of Plots

We use days of labor as the independent variable of interest. This variable includes both family and hired labor, which we aggregate into the single variable. However, family labor generally predominates, as is clear in the summary statistics below. A major issue is that labor is reported at the plot level, while crop output is reported at the plot-crop level. This means that the labor inputs we observe on a given plot are applied to all crops on that plot, and we are not able to disaggregate that labor by crop. As such, after constructing price observations, we also drop any plots that are planted with at least one crop with a missing price value. In this way, the entirety of labor allocated to the plot is applicable to the entirety of the output value we construct. This restriction drops less than ten percent of plots in each wave.

[5] In the appendix, we report results requiring at least ten valid price observations and without any price requirements. Our conclusions are unchanged.

3.4 Other Controls

In addition to the productive inputs, we also control for a number of agriculture variables, including plot quality (as reported by the farmer), plot type, plot erosion, plot slope, and whether the plot is located in a swamp or wetland. In agricultural regressions, in addition to days of labor, we include the size of the plot (log of acres) and amount of fertilizer applied to the plot (log of kilograms) as productive inputs. In non-farm regressions, we include one separate productive input in addition to labor: log of total costs. This variable is directly asked of all households with a non-farm enterprise. The use of value for a productive input could bias the coefficients in our production functions if input prices vary by region and/or household (Jacoby, 1993). We include district/wave fixed effects to help alleviate both regional and temporal differences in input prices. Additional industry controls include indicator variables for industry and a single indicator variable for whether the non-farm enterprise has electricity.

Finally, the dependent variable for non-farm production is constructed using a single survey question, which asks respondents for total revenue in the 30 days prior to the interview. We include month of interview fixed effects in order to control for any seasonality in labor allocation across the year. However, we also estimate production functions on separate subsamples, depending on month of interview, and present these results below.

3.5 Summary Statistics and Sample Restrictions

Summary statistics at the household/wave level are shown in Table 1. To calculate these statistics, we collapse all agricultural and non-farm statistics to the household level before taking logs. The first column presents statistics for all households that are in the Malawi IHS data and meet our criteria for calculating plot revenue, but are not in the final sample because they do not operate both types of enterprises in the same wave. The second column includes all households that are in the final sample. The third column presents the p-value of tests for equality across the two samples.[6] Of the households with valid revenue observations but that do not operate both types of enterprises, approximately 86 percent of them have plots but no non-farm enterprise while 14 percent operated non-farm enterprises but had no valid plot observation.[7] Overall, the statistics suggest there are some differences between the two samples, which may limit the external validity of our results.

One concern with our estimation method is that labor is seasonal, and thus we may not be making the proper temporal comparison across productive activities. This concern is compounded by the fact that the survey captures agricultural labor for the whole season whereas non-farm labor is only reported for the last 30 days. To help examine whether this is affecting our results, we present a graph of median non-farm revenue and labor by month of interview in Figure 1. In Malawi, there appears to be clear seasonality in labor allocation to non-farm production. In particular, both labor allocation and revenue appear to be lowest from February to around July, and then increase between August and January. Given that there does appear to be some seasonality in labor allocation and revenue, we will explore this further below, though we control for month of interview in all production function estimates.

4 Results

We begin our discussion of the results with the pooled production function estimates in Table 2. We present results for both the Cobb-Douglas and translog specifications as well

[6] We construct these p-values by regressing each variable on a dummy variable indicating whether a household/wave observation is in the final sample, clustering the standard errors at the household level.

[7] Since non-farm enterprises are never dropped due to restrictions on revenue – unlike agricultural plots – a household only appearing in the non-farm statistics does not imply the household does not operate any plots. Rather, it is possible that the household operates plots that get dropped in our price-creation procedure explained above.

as for all households (first two columns) and for households that operate both enterprises in the same wave (last two columns). Comparing the Cobb-Douglas results for all households (column 1) to final sample households (column 3) suggests the production technology is relatively similar. While labor appears slightly more productive – for both agricultural and non-farm production – for the select subsample of households, the opposite is true for land. The coefficients on fertilizer and non-farm costs, however, are relatively similar.

The translog results are harder to interpret given the large number of interaction terms. Moreover, it is clear moving from column two to column four that we lose a substantial amount of precision in our estimates. This is especially true for agriculture. However, the translog coefficients for non-farm production are remarkably similar for both groups of households, with the largest difference being just 0.006 (costs). A glance at the interaction terms suggests that the Cobb-Douglas specification might not be rejected for agriculture; there is only one significant interaction term (including the squared terms) in each column. The same cannot be said for non-farm production, however. In both columns two and four, the squared costs term and interaction term are significant, with the former having a t-statistic of more than 20 in each column. We formally test for the nested Cobb-Douglas in the translog specifications and present these results at the bottom of the table. Consistent with the above interpretation of the coefficients, we reject the Cobb-Douglas specification for agriculture for all households but fail to reject for our final sample. However, the F-test strongly rejects the Cobb-Douglas specification for non-farm production. Given these results, we use the translog specification in all results that follow.

The theoretical analysis above predicts that households should equate the marginal revenue products across each activity the household is engaged in. However, some previous studies have used average products when analyzing allocative efficiency across sectors. Before presenting our MRPL estimates, we first show our estimates of Average Revenue

Product of Labor (ARPL) to show the importance of the choice between marginal and average products. Figure 2 presents kernel density estimates of ARPL using actual revenue (Panel A) and predicted revenue from the production functions (Panel B). In both cases, it appears that non-farm ARPL stochastically dominates agricultural ARPL. Table 3 presents the mean and median ARPLs, again using both actual and predicted revenue. We find the average revenue product in non-farm production is substantially higher than the average revenue product in agricultural production. Higher non-farm average revenue products, however, does not imply that a reallocation of labor towards non-farm production would bring about welfare gains.

When we use the marginal revenue product of labor to analyze allocative efficiency, we find the opposite result. We present these MRPL estimates in Table 4. Since MRPL equality is only theoretically predicted for households that operate both types of activities, we present MRPL results for only this subset of households. All MRPL estimates are in (March) 2010 Malawi Kwacha (MWK). We construct MRPL estimates using the production function in column four of Table 2 and present these in the first two columns of Table 4. Recall that we construct household MRPL estimates in two ways: taking the simple median across plots/enterprises and weighting MRPLs across plots/enterprises based on labor allocation. The simple median suggests an average agricultural MRPL of about 127 MWK. In March of 2010, the exchange rate was approximately 150 MWK to USD, so this MRPL translates to slightly less than one US dollar per worker-day. The average non-farm MRPL, on the other hand, is significantly smaller, at just 23 MWK. Finally, the difference, approximately 101 MWK, is highly significant.

The second column presents results using the same base production function as the first column, but aggregates MRPLs across plots/enterprises within a household using labor allocation weights. Two patterns emerge. First, all three MRPL estimates are substantially more precisely estimated than the simple medians. Second, the average agricultural MRPL is smaller than in column 1, possibly suggesting some plots with low labor allocation and high MRPL were inflating the estimated MRPL in the first column.[8] The end result is that the estimated MRPL difference is less than sixty percent as large in column 2. Nonetheless, the increase in precision results in a highly significant difference of 56 MWK, or around 0.35 USD, per worker-day.

Taken at face value, these numbers point to possible harm from encouraging a reallocation of labor away from agricultural production and towards non-farm production. At the margin, reallocating a single day of labor in this direction is estimated to *decrease* income for the median household by around 56 Kwacha, which is a substantial percentage of average income in Malawi. If it is the case that the non-agricultural sector is indeed more productive than the agricultural sector, then our results suggest that the margin of adjustment is not within the household, but instead made by people completely switching sectors.

Overall, the first two columns suggest that households could increase their income by reallocating some labor away from non-farm production and towards agricultural production. However, the estimates in the first two columns also rely on pooled production functions, in which we estimate agricultural and non-farm production functions in a single regression. This increases our statistical power, but at the cost of a more restrictive estimation strategy. We relax the restrictions on the fixed effects in columns three and four by estimating the agricultural and non-farm production functions separately, allowing the fixed effect to affect agricultural and non-farm production differently. Reassuringly, the patterns and general conclusions are identical.

Given the evidence that disaggregated labor statistics are more prone to bias than are

[8] This is not unexpected. The MRPL formula requires dividing by total labor allocation on a plot. Very small reported labor allocation can thus result in very large estimated MRPL values.

aggregated labor statistics, as households tend to forget plots and distribute labor from those plots when reporting on other plots (Arthi et al., 2018).[9] We present another set of results in column 5 in which we aggregate all productive inputs to the household level *prior* to estimating production functions. We still pool these results and we present these MRPL estimates in column 5. Again, the results are generally consistent with the previous four columns. Taking these results together, we interpret this as evidence that agricultural MRPL tends to be higher than non-farm MRPL within the median household.

4.1 Possible Explanations

While the above results show agricultural MRPL to be consistently higher than non-farm MRPL, there may be important factors of the environment in which these households operate that influence their allocation decisions. In this section, we first show that substantial heterogeneity in MRPLs exists, and then examine the role of price and production risk in explaining the allocation decisions we observe. Given the relative consistency across specifications and the fact that weighting by labor (column two of Table 4) is not only the most precisely estimated but also the most conservative estimate of MRPL difference, we use this specification as the basis for all analyses that follow. We begin with a simple histogram of MRPL difference in Figure 3. For ease of presentation, we trim the top 5 percent of the sample. Consistent with our empirical results, the vast majority of the distribution lies to the right of zero, with the highest density around 50 MWK. While there is a wide distribution of MRPL differences, fully 86 percent of household/wave observations have a positive difference, which underscores the empirical results.

We next examine whether price and production risk help explain the allocation decisions we have observed. Table 5 and Table 6 examine the role of price risk in allocation decisions.

[9] We note, however, that it is not clear how this would affect our results, as we do not use all plots in estimation.

In Table 5, we split the sample by crop choice and hiring. In Panel A, we look at households that grow maize (column one) and households that grow tobacco and/or cotton (column two). Maize is a common subsistence crop and tobacco and cotton are common cash crops. As shown by Barrett (1996), we expect larger deviations from equality for cash crops than subsistence crops when price risk affects decision-making. Consistent with this hypothesis, MRPL difference is more than twice as large for households that grow tobacco and/or cotton than for households that grow maize. However, given the small sample size for cash crops, there is substantial imprecision; the standard error is around 40 percent larger than the *point estimate* for the maize subsample. Nonetheless, the pattern supports the hypothesis.

In Panel B and Panel C, we split the sample by households that hire for agriculture (Panel B) and for non-farm (Panel C) and estimate MRPL separately for each group. Households which hire labor from the market may have higher productivity for the activity in which they hire for two reasons. First, the households may have higher productivity levels in order to afford hiring the outside labor. Secondly, households which hire labor are more exposed to price risk, and therefore will underallocate labor to that activity and have a higher MRPL. Barrett (1996) shows that households that hire labor for agricultural production are more likely to be net sellers. Since all non-farm production is sold on the market, households which hire for non-farm production are more exposed to price risk for that activity. The results in Panels B and C of Table 5 support these predictions, showing higher MRPLs in the activities for which households hire outside labor.

We also study MRPL heterogeneity using proxies for market access in order to test the importance of price risk. We now split households based on the median of three separate variables: total revenue, total acreage, and total crop sales in Panel A, Panel B, and Panel C, respectively. We assume households in the upper half of the distribution of these three variables are have better access to markets. Column one presents MRPL estimates for all households above the median of the respective variable while column two presents estimates for all households below the median. If these variables are correlated with marketable surplus, we expect larger MRPL differences for households above the median than below if price risk appreciably affects labor allocation decisions. In fact, this is exactly what we see. When splitting the sample by revenue, households above the median show a median MRPL difference of more than 120 MWK, while households below the median show a difference of just 6 MWK. We see a similar, though less pronounced, pattern for both acreage and crop sales. This is again consistent with the idea that price risk is an important component of agricultural decision-making.

Having established that price risk appears to be an important predictor of deviations from MRPL equality, we now move to rainfall. Figure 4 presents locally weighted regressions of MRPL on rainfall in waves one and two. Panel A presents results using rainfall in levels while Panel B presents results using the rainfall coefficient of variation. It is difficult to come to any conclusions regarding the direction of the correlation, other than that rainfall appears to have no significant effect on non-farm MRPL. To examine this relationship more formally, we estimate MRPL using our main results in Table 4. We then estimate quantile (median) regressions of MRPL on both current rainfall and the rainy season precipitation coefficient of variation.[10] We do not to split the sample as in Table 5 and Table 6 in order to be able to include both rainfall variables simultaneously. We present these results in Table 7. In the table, there are three separate regressions. We only have geovariables for waves one and two, so results are restricted to these waves. When including both rainfall variables in each regression, clear patterns emerge. Intuitively, higher rainfall is associated with higher agricultural MRPL and a higher difference. Rainfall CV – which we use as

[10]The rainy season (November to April) precipitation coefficient of variation is defined using the previous 15 years (McCarthy and Kilic, 2015).

a proxy of production risk – is also positively correlated with both agricultural MRPL and MRPL difference. The results indicate that production risk affects labor allocation significantly.

4.2 Robustness Checks

One concern already documented is seasonality. Our specific concern is that although we restrict estimation to households that operate both non-farm enterprises and agricultural plots in the same year, it may still be the case that households allocate labor seasonally within the year, such that marginal revenue products in non-farm production come only from labor allocation in the agricultural slack season (either between planting and harvest or in the off-season). To check this possibility, we can estimate production functions and construct MRPL estimates based on month of interview. If seasonality is affecting our results, then we are likely to see substantially different MRPL differences during planting and harvest seasons, especially since these two seasons apparently show large differences in non-farm labor allocation. Households that are interviewed during this times are simultaneously allocating labor to both types of production. [11]

Fortunately, the Malawi IHS allows us to examine whether seasonality may be affecting our results. Panel households were enumerated twice in every wave for the Malawi IHS. In general, approximately half of households responded to the non-farm module during the first visit – which tended to be during or just after the planting season – while approximately half of panel households responded to the non-farm module during the second visit, just after harvest.[12] Importantly, only panel households are visited twice. Moreover, panel households are the minority of our sample, as there are large cross-section components of waves one and three. As such, a large portion of households responded to the agricultural

[11]Table A1 in the appendix presents summary statistics for month of interview across all three waves.
[12]All households should have reported agricultural output during the harvest visit.

and non-farm modules during the same visit, during or immediately following harvest.

We estimate production functions separately for the subsample of households that were interviewed during the "low non-farm revenue" season, which we define as February to July[13], and the subsample of households that were interviewed during times of higher non-farm labor allocation and revenue, from August to January. We present these results in Table 8. The first two columns restrict estimation to only non-farm enterprises that were interviewed from August to January, while the last two columns restrict estimation to only households interviewed from February to July. It appears that agricultural MRPL is slightly higher for those households interviewed from February to July. However, non-farm MRPL is also slightly higher, resulting in remarkably consistent MRPL differences across time periods. We interpret this as suggestive evidence that temporal variations in labor allocation and survey timing are unlikely to be driving our results.

Another possible explanation is that rainfall happened to be very good in the three survey years, which might induce an increase in MRPL over what was expected by farmers.[14] Figure 5 presents kernel density estimates of rainfall in waves one and two (the only waves for which we have rainfall data). The figure shows that rainfall was relatively higher than the ten-year average in wave one but relatively lower in wave two. While we do not have rainfall for wave three, we can look at other proxies for rainfall. Whenever farmers report area harvested lower than area planted, they are asked for the cause. In wave three, around 60 percent of plot-crop observations report lower area harvest than area planted. Of these, between 80 and 90 percent (depending on whether we use the cross-section or panel subsample) blame drought or irregular rains. In other words, it appears that wave three was actually a relatively poor rainfall year in Malawi. Based on these facts, we conclude high

[13]This only refers to the non-farm module. The agricultural statistics still come from the post-harvest questionnaire.

[14]Recall that we use predicted revenue to construct our main MRPL estimates, which is likely a better predictor of households' ex ante expectations. However, results in the appendix also show that the use of actual revenue instead of predicted revenue does not affect our results.

rainfall is unlikely to be driving our high agricultural MRPL estimates.

In the appendix, we also explore the robustness of MRPL estimates to several variations in specifications. We estimate MRPL when using a minimum of 10 prices observations instead of five to construct aggregate crop prices (Table A4) as well as no minimum number of price observations (Table A5). In Table A2, we present MRPL estimates when not trimming the top one percent of revenue and labor. Finally, in Table A3, we estimate MRPL using actual revenue instead of predicted revenue. Our conclusions are unchanged and agricultural MRPL remains higher than non-farm MRPL in all models/specifications.

5 Conclusion

In this paper, we examine allocative efficiency across agriculture and non-farm production in rural households. To the best of our knowledge, this is the first paper to test allocative efficiency using the marginal revenue product of labor across agricultural and non-farm activities within the household. At first glance, we present evidence of labor misallocation in Malawi, though in a surprising direction; across a number of specifications, agricultural MRPL is consistently *higher* than non-farm MRPL. Horeover, we show that risk plays an important role in labor allocation, suggesting that deviations from profit-maximizing conditions are not necessarily inconsistent with rationality. Our results suggest that removing risk – or, similarly, insuring farmers against risk – might result in higher incomes.

These results suggest that the median household could increase its mean income by reallocating labor out of non-farm production and into agricultural production. In other words, the results seemingly point to a misallocation of labor. However, failure of MRPL equality does not imply households can increase their expected *utility* by reallocating labor. The fact that risk appears to play an important role in this apparent misallocation suggests that households may indeed by making rational labor allocation decisions. These results show that risk leads households to protect themselves ex ante, but at the cost of a lower ex post income.

Moreover, our findings highlight the importance of using marginal revenue products, instead of average revenue products, when examining allocative efficiency. Our results show that the use of average revenue products would lead to an erroneous conclusion regarding allocative effiiciency within the household.

There are some important caveats to our findings. First, our sample is a relatively select group of households from just a single country. Previous research has shown that research findings may differ substantially across countries in sub-Saharan Africa (Dillon et al., 2017). Second, our estimation strategy requires a specific set of identification assumptions and, as such, omitted variables are a very real possibility. As with all such assumptions, additional studies that rely on different identification assumptions are required. Given the policy implications of our results, testing their robustness across different identification assumptions and country contexts is a ripe area for future research.

Abdulai, A. and Regmi, P. P. (2000). Estimating Labor Supply of Farm Households under Nonseparability: Empirical Evidence from Nepal. *Agricultural Economics*, 22(3):309–320.

Armendáriz, B. and Morduch, J. (2010). *The economics of microfinance*. MIT press.

Arthi, V., Beegle, K., De Weerdt, J., and Palacios-López, A. (2018). Not your average job: Measuring farm labor in Tanzania. *Journal of Development Economics*, 130:160–172.

Barrett, C. B. (1996). On Price Risk and the Inverse Farm Size-Productivity Relationship. *Journal of Development Economics*, 51(2):193–215.

Barrett, C. B., Reardon, T., and Webb, P. (2001). Nonfarm income diversification and household livelihood strategies in rural Africa: concepts, dynamics, and policy implications. *Food policy*, 26(4):315–331.

Barrett, C. B., Sherlund, S. M., and Adesina, A. A. (2008). Shadow Wages, Allocative Inefficiency, and Labor Supply in Smallholder Agriculture. *Agricultural Economics*, 38(1):21–34.

Benjamin, D. (1992). Household Composition, Labor Markets, and Labor Demand: Testing for Separation in Agricultural Household Models. *Econometrica*, 60(2):287–322.

Chenery, H. B., Syrquin, M., and Elkington, H. (1975). *Patterns of development, 1950-1970*, volume 75. Oxford University Press London.

Davis, B., Di Giuseppe, S., and Zezza, A. (2017). Are African households (not) leaving agriculture? Patterns of householdsâĂŹ income sources in rural Sub-Saharan Africa. *Food Policy*, 67:153–174.

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. (2017). Asymmetric Non-Separation and Rural Labor Markets. *Working Paper*.

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

Guirkinger, C., Platteau, J.-P., and Goetghebuer, T. (2015). Productive inefficiency in extended agricultural households: Evidence from Mali. *Journal of Development Economics*, 116:17–27.

Haggblade, S., Hazell, P., and Reardon, T. (2010). The rural non-farm economy: Prospects for growth and poverty reduction. *World Development*, 38(10):1429–1441.

Hicks, J. H., Kleemans, M., Li, N. Y., and Miguel, E. (2017). Reevaluating agricultural productivity gaps with longitudinal microdata. Technical report, National Bureau of Economic Research Working Paper 23253.

Jacoby, H. G. (1993). Shadow Wages and Peasant Family Labour Supply: An Econometric Application to the Peruvian Sierra. *The Review of Economic Studies*, 60(4):903–921.

LaFave, D. and Thomas, D. (2016). Farms, Families, and Markets: New Evidence on Completeness of Markets in Agricultural Settings. *Econometrica*, 84(5):1917–1960.

Lanjouw, J. O. and Lanjouw, P. (2001). The rural non-farm sector: issues and evidence from developing countries. *Agricultural economics*, 26(1):1–23.

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

Linde-Rahr, M. (2005). Differences in agricultural returns: an empirical test of efficiency in factor input allocation using vietnamese data. *Agricultural Economics*, 32(1):35–45.

McCarthy, N. and Kilic, T. (2015). The nexus between gender, collective action for public goods and agriculture: Evidence from Malawi. *Agricultural Economics*, 46(3):375–402.

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

Merfeld, J. D. (2018a). Hiding with your Feet: Agricultural Plots, Labor Allocation, and Income Hiding in Ethiopia. *Working Paper*.

Merfeld, J. D. (2018b). Moving up or Just Surviving? Non-Farm Self-Employment in India. *Working Paper*.

Nagler, P. and Naudé, W. (2014). Non-farm enterprises in rural Africa: New empirical evidence. *World Bank Policy Research Working Paper*, (7066).

Ranis, G. and Fei, J. C. (1961). A theory of economic development. *The American economic review*, pages 533–565.

Sandmo, A. (1971). On the theory of the competitive firm under price uncertainty. *The American Economic Review*, 61(1):65–73.

Seshan, G. K. (2014). The Impact of Trade Liberalisation on Household Welfare in a Developing Country With Imperfect Labour Markets. *Journal of Development Studies*, 50(2):226–243.

Skoufias, E. (1994). Using Shadow Wages to Estimate Labor Supply of Agricultural Households. *American Journal of Agricultural Economics*, 76(2):215–227.

Stiglitz, J. E. (1974). Incentives and Risk Sharing in Sharecropping. *The Review of Economic Studies*, 41(2):219–255.

Udry, C. (1996). Gender, Agricultural Production, and the Theory of the Household. *Journal of Political Economy*, 104(5):1010–1046.

Walther, S. (2018). Noncooperative decision making in the household: Evidence from Malawi. *Journal of Development Economics*.

Young, A. (2013). Inequality, the urban-rural gap, and migration. *The Quarterly Journal of Economics*, 128(4):1727–1785.

Tables

Table 1: Summary Statistics

	(1) Dropped Households	(2) Final Sample	(3) Diff (p-value)
Agricultural Production Statistics			
Household has plot in sample	0.856	1.000	
	(0.351)		
Ag revenue (2010 MWK - log + 1)	10.005	10.186	0.000
	(1.261)	(1.233)	
Total labor (log)	4.348	4.268	0.000
	(0.831)	(0.900)	
Total family labor (log + 1)	4.254	4.086	0.000
	(0.972)	(1.137)	
Total hired labor (log + 1)	0.422	0.771	0.000
	(0.996)	(1.261)	
Household hired for ag production (yes = 1)	0.179	0.316	0.000
	(0.383)	(0.465)	
Fertilizer (kg - log + 1)	1.377	1.499	0.000
	(1.594)	(1.609)	
Acres in sample	1.738	2.178	0.310
	(18.724)	(26.109)	
Maize in sample	0.930	0.923	0.160
	(0.256)	(0.267)	
Tobacco in sample	0.101	0.077	0.000
	(0.301)	(0.266)	
Non-FarmProduction Statistics			
Household has non-farm ent. in sample	0.144	1.000	
	(0.351)		
NF revenue (2010 MWK - log + 1)	9.613	9.094	0.000
	(1.522)	(1.391)	
Total labor (log)	3.085	2.868	0.000
	(0.809)	(0.852)	
Total family labor (log + 1)	2.954	2.790	0.000
	(0.803)	(0.845)	
Total hired labor (log + 1)	0.400	0.243	0.000
	(1.114)	(0.856)	
Household hired for NF production (yes = 1)	0.124	0.083	0.000
	(0.330)	(0.275)	
Last monthly costs (2010 MWK - log + 1)	8.500	7.808	0.000
	(2.768)	(2.825)	
Household Characteristics			
Male household head (yes = 1)	0.735	0.803	0.000
	(0.441)	(0.398)	
Household size	4.667	5.144	0.000
	(2.141)	(2.127)	
Total acres owned	3.621	4.941	0.159
	(65.536)	(51.902)	
N	19154	4105	

Standard deviations are in parentheses. Statistics are at the household/wave level. Agricultural and non-farm statistics were collapsed to the household/wave level before taking logs. The "Dropped Households" columns include all households that meet our criteria for valuing agricultural output (on at least one plot) or which operate a non-farm enterprise but do not operate both in the same wave. The "Final Sample" column includes the sample of households we use to calculate MRPL: those that operate at least one plot *and* at least one enterprise in the same wave. To calculate the p-value for their difference, we regress each variable in the far left column on a single dummy variable for whether the household is in our final sample or not. We cluster the standard errors at the household level.

Table 2: Production Function Estimates

	All households		Final sample	
	(1)	(2)	(3)	(4)
	C-D	Translog	C-D	Translog
Ag Labor (L_a)	0.255***	0.149*	0.284***	0.027
	(0.017)	(0.088)	(0.027)	(0.125)
Acres (A)	0.365***	0.450***	0.336***	0.440***
	(0.016)	(0.088)	(0.028)	(0.145)
Fertilizer (F)	0.084***	0.250***	0.085***	0.216**
	(0.008)	(0.050)	(0.014)	(0.085)
$L_a \times L_a$		0.012		0.026
		(0.011)		(0.016)
$A \times A$		0.008		−0.005
		(0.015)		(0.022)
$F \times F$		−0.026***		−0.026**
		(0.008)		(0.013)
$L_a \times A$		−0.017		−0.020
		(0.017)		(0.029)
$L_a \times F$		−0.011		0.003
		(0.009)		(0.014)
$F \times A$		0.001		0.000
		(0.010)		(0.017)
NF Labor (L_n)	0.188***	0.303***	0.201***	0.305***
	(0.027)	(0.103)	(0.027)	(0.109)
Costs (C)	0.236***	−0.289***	0.232***	−0.295***
	(0.012)	(0.033)	(0.011)	(0.032)
$L_n \times L_n$		0.012		0.009
		(0.018)		(0.019)
$C \times C$		0.057***		0.057***
		(0.002)		(0.002)
$L_n \times C$		−0.037***		−0.036***
		(0.010)		(0.010)
Test for Nested Cobb-Douglas (p-value)				
Agriculture		0.007		0.192
Non-Farm		0.000		0.000

Standard errors clustered at the household level are in parentheses. Household fixed effects are included in all regressions. Also included are month of interview fixed effects and wave/district fixed effects. Month of interview and wave/district fixed effects are allowed to vary by type of production. In addition, we include crop dummies, plot quality variables, non-farm industry dummies, and a dummy indicating whether the non-farm industry has access to electricity. The F-tests present tests for a nested Cobb-Douglas production function in each translog; the p-value is constructed by testing whether all squared and interaction terms are simultaneously zero. The "All Households" columns include all households that operate at least one plot or at least one enterprise across all three waves of the Malawi LSMS. The "Final Sample" column includes the sample of households we use to calculate MRPL: those that operate at least one plot *and* at least one enterprise in the same wave. Revenue and non-farm costs are in (March) 2010 MWK.

* $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$

Table 3: Mean and Median of Average Revenue Product of Labor

	(1) Actual Revenue	(2) Predicted Revenue
Panel A: Mean		
Agriculture	875	681
Non-Farm	1584	1567
Panel B: Median		
Agriculture	355	337
Non-Farm	541	495

All statistics are calculated as revenue in MWK over number of days worked. Actual revenue uses reported revenue (non-farm) or constructed revenue using aggregate prices and reported harvest (agriculture). Predicted revenue uses revenue predicted from the production function estimates used to estimate the marginal revenue product of labor in Table 2.

Table 4: MRPL Estimates

	Pooled Production Function		Separate Production Functions		Collapsed
	(1)	(2)	(3)	(4)	(5)
	Simple Median	Weighted by Labor	Simple Median	Weighted by Labor	Collapsed
Agriculture	127.013***	81.819***	170.996***	111.566***	99.882***
	(17.944)	(8.375)	(30.461)	(14.242)	(10.105)
Non-Farm	22.607	21.461**	19.839	18.418	30.522***
	(15.303)	(10.385)	(29.702)	(19.783)	(11.443)
Difference	100.606***	56.227***	134.491**	79.468***	68.121***
	(28.246)	(14.974)	(54.096)	(28.667)	(17.581)
N (Household/Wave)	4097	4097	4097	4097	4097

Standard errors are constructed through bootstrapping the MRPL construction 1,000 times. The bootstrap is set to draw households. MPRL estimates are in 2010 MWK. MRPL difference is constructed as agricultural MRPL minus non-farm MRPL. The MRPLs in the first and second columns are constructed by estimating a single production function but allowing the effects of all variables (except household fixed effects) to vary by enterprise type. MRPLs in the third and fourth columns are constructed by estimating separate production functions for agriculture and non-farm production. In the fifth column, MRPL is constructed by collapsing across plots and enterprises to the household/wave level and estimating a single production function, again allowing the effects of all variables (except household fixed effects) to vary by enterprise type. * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$

Table 5: MRPL, Crop Choice, and Hiring

	(1)	(2)
Panel A: Crop Choice	**Subsistence (Maize)**	**Cash (Tobacco/Cotton)**
Agriculture	73.552***	109.278**
	(11.398)	(52.374)
Non-Farm	13.775	9.198
	(13.688)	(29.418)
Difference	47.183**	113.605
	(19.350)	(70.341)
N (Household/Wave)	2422	405
Panel B: Hires for Agriculture	**Yes**	**No**
Agriculture	144.275***	71.485***
	(26.984)	(7.827)
Non-Farm	16.110	30.255***
	(23.307)	(30.255)
Difference	119.240***	40.046***
	(42.953)	(14.896)
N (Household/Wave)	1295	2800
Panel C: Hires for Non-Farm	**Yes**	**No**
Agriculture	104.212**	79.106***
	(52.327)	(8.330)
Non-Farm	110.335	13.268
	(73.866)	(8.874)
Difference	3.625	59.456***
	(88.822)	(14.462)
N (Household/Wave)	337	3758

A separate pooled production function is estimated for each column in each panel. Panel A splits the sample by crop choice, Panel B splits the sample by whether the household hires for agricultural production, and Panel C splits the sample by whether the household hires for non-farm production. Standard errors are constructed through bootstrapping the process 1,000 times. The bootstrap is set to draw households. * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$

Table 6: MRPL and Production Characteristics

	(1) Above Median	(2) Below Median
Panel A: Revenue		
Agriculture	141.216***	40.476***
	(23.912)	(6.218)
Non-Farm	30.684	29.679**
	(19.583)	(11.689)
Difference	122.875***	6.492
	(36.921)	(10.793)
N (Household/Wave)	2044	2053
Panel B: Acreage		
Agriculture	82.795***	72.194***
	(11.362)	(13.184)
Non-Farm	13.063	25.517*
	(12.999)	(14.783)
Difference	67.382***	39.648**
	(22.566)	(19.481)
N (Household/Wave)	2045	2052
Panel C: Crop Sales		
Agriculture	85.513***	71.023***
	(13.528)	(10.785)
Non-Farm	7.403	35.247**
	(10.148)	(15.270)
Difference	85.039***	31.222*
	(21.821)	(18.454)
N (Household/Wave)	1973	2124

A separate pooled production function is estimated for each column in each panel. Panel A splits the sample by median of total revenue, Panel B splits the sample by median total acreage, and Panel C splits the sample by median of crop sales (gross). Standard errors are constructed through bootstrapping the process 1,000 times. The bootstrap is set to draw households. * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$

Table 7: Median Regression - MRPL and Rainfall

	(1)	(2)	(3)
MRPL:	Agriculture	Non-Farm	Difference
Current rainfall (mm)	0.023***	0.001	0.027*
	(0.007)	(0.008)	(0.015)
Rainfall CV	0.346**	0.099	0.354*
	(0.166)	(0.107)	(0.213)

The results are from three separate quantile (median) regressions using the different MRPL estimates as dependent variables. MRPL estimates come from column two of Table 4. Only waves one and two are included. Standard errors are constructed through bootstrapping the process 1,000 times. The bootstrap is set to draw households. * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$

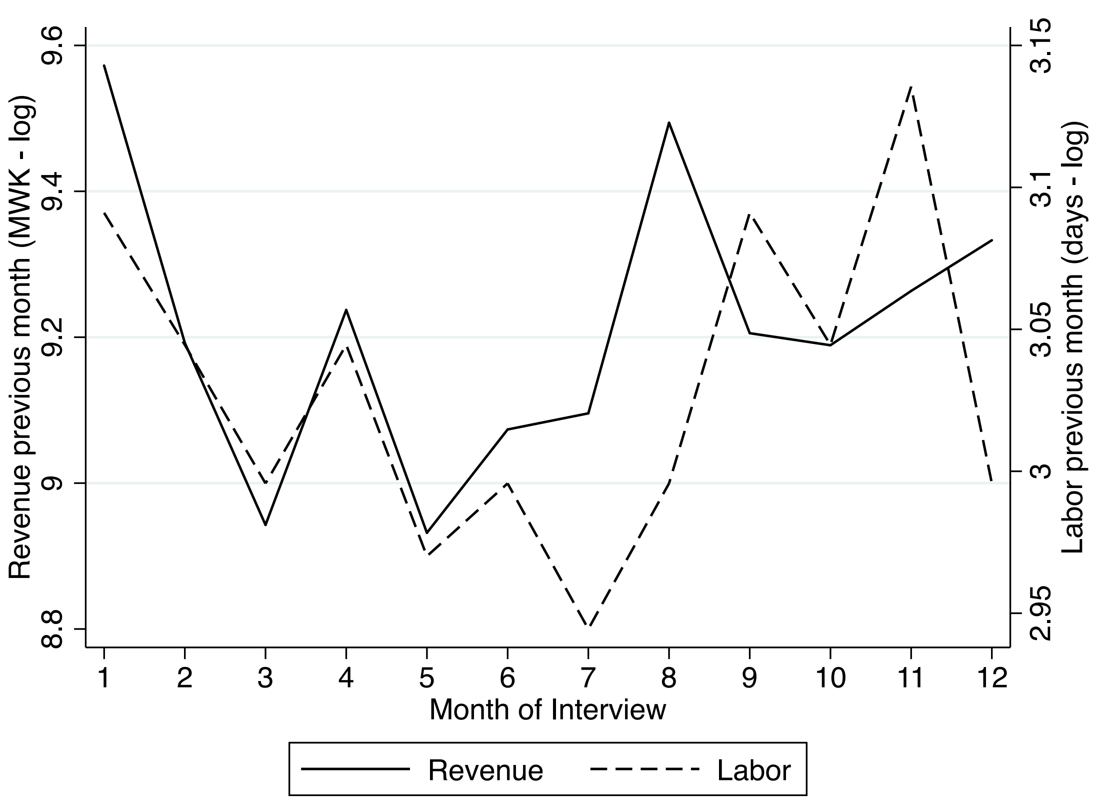
Table 8: MRPL by Months of Interview

	Aug to Jan		Feb to July	
	(1)	(2)	(3)	(4)
	Simple Median	Weighted by Labor	Simple Median	Weighted by Labor
Ag MRPL	72.219***	71.955***	87.723***	89.321***
	(12.675)	(13.031)	(11.444)	(16.039)
NF MRPL	18.883	18.772	34.801**	34.715**
	(12.974)	(12.906)	(14.193)	(15.956)
Difference	46.508**	45.952**	48.235***	49.265*
	(20.385)	(20.530)	(18.183)	(29.401)
N (Household/Wave)	1899	1899	2200	2200

Standard errors are constructed through bootstrapping the process 1,000 times. The bootstrap is set to draw households. The results re-estimate the results from columns one and two of Table 4, restricting estimation only to households surveyed during the "high" non-farm season of August to January (columns one and two) or the "low" non-farm season of February to July (columns three and four). * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$

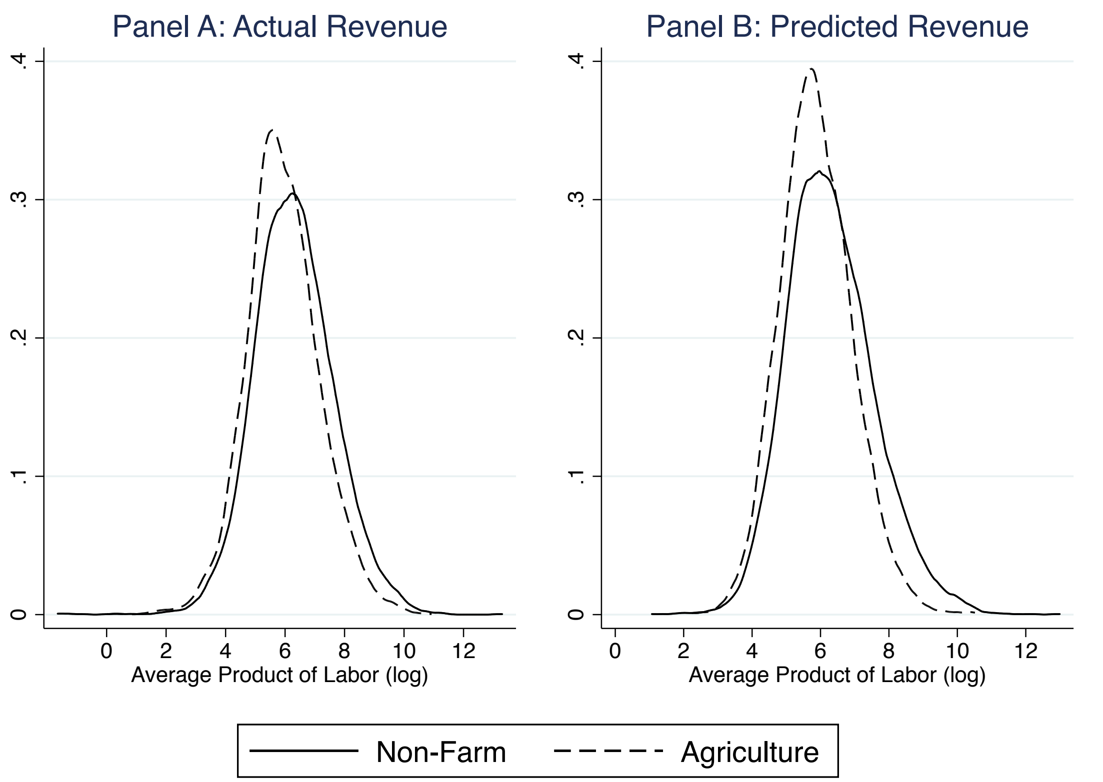
Figures

Figure 1: Month of Interview and Non-Farm Production Characteristics



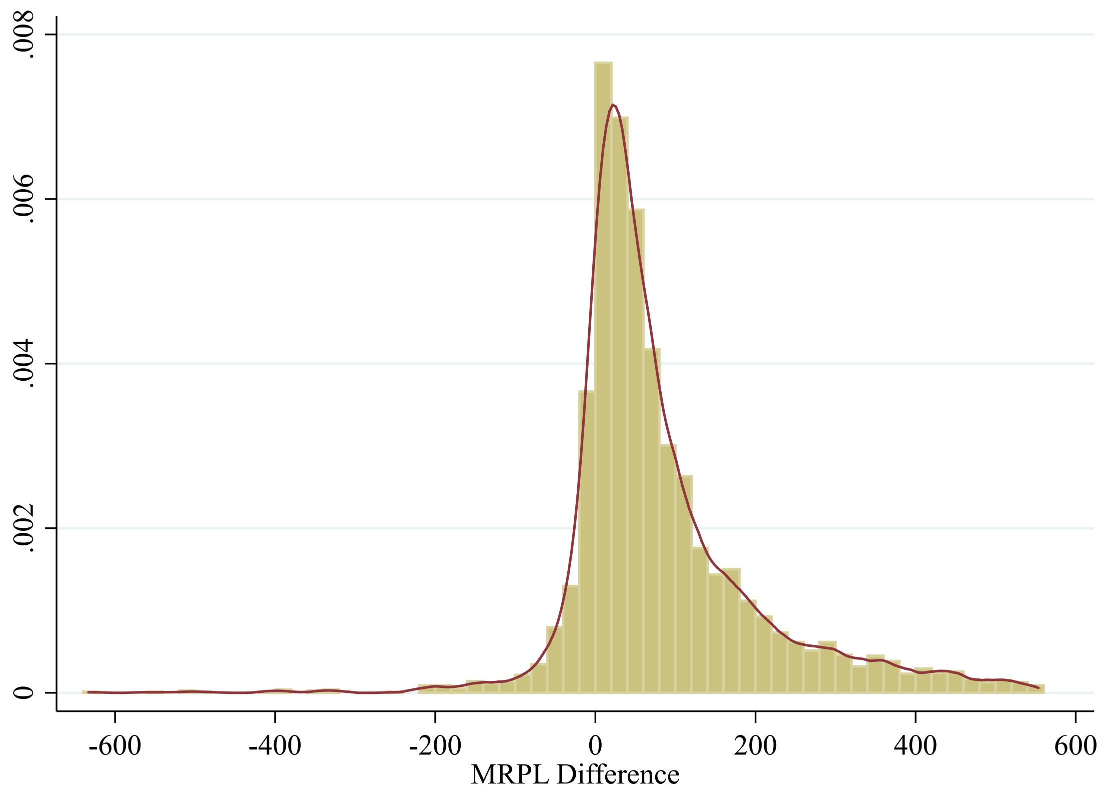
Revenue and labor are for non-farm production only.

Figure 2: Average Revenue Product of Labor



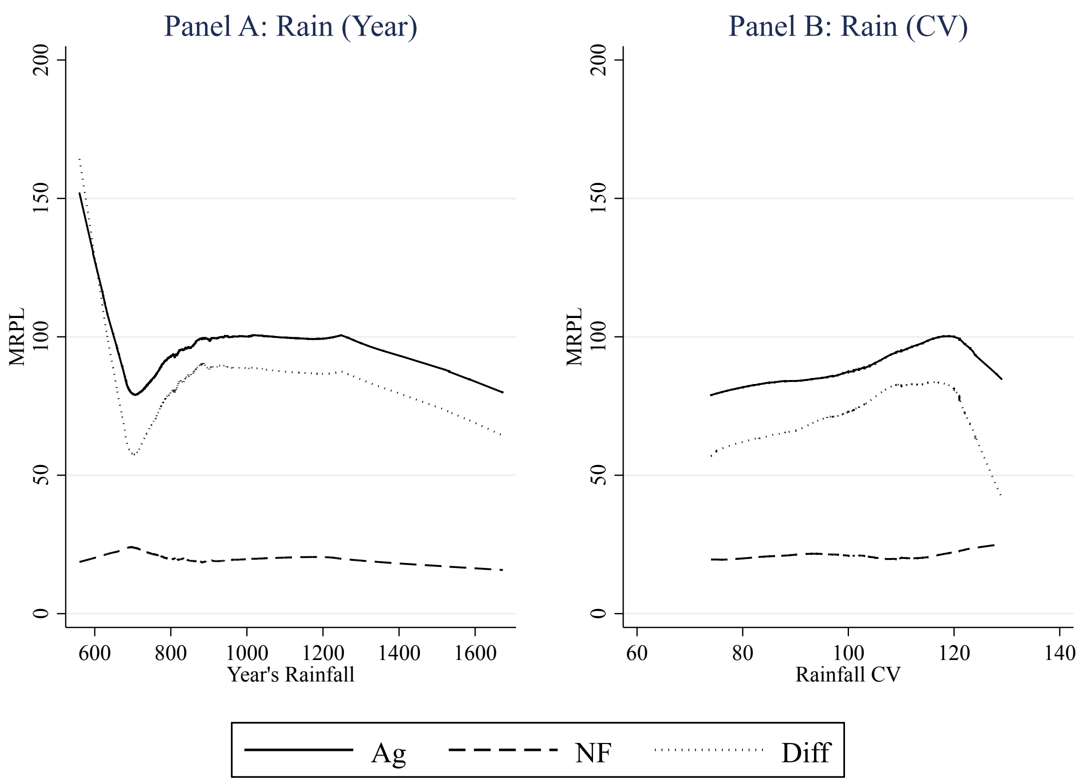
All statistics are calculated as revenue in MWK over number of days worked. Actual revenue uses reported revenue (non-farm) or constructed revenue using aggregate prices and reported harvest (agriculture). Predicted revenue uses revenue predicted from the production function estimates used to estimate the marginal revenue product of labor in Table 2.

Figure 3: MRPL Distribution - Pooled Production Function



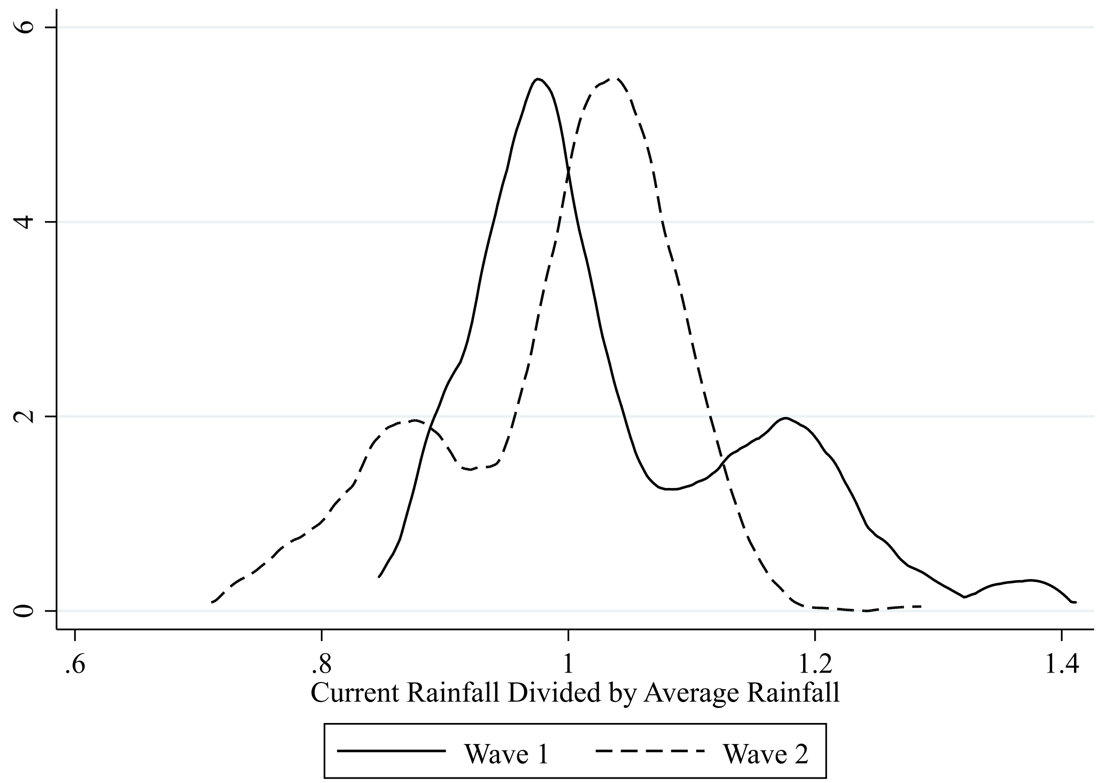
MRPL estimates are constructed as in column two of Table 4. MRPL difference is constructed as agricultural MRPL minus non-farm MRPL. The top five percent are trimmed from the figure for ease of presentation.

Figure 4: MRPL and Rainfall



All figures are locally weighted regressions of MRPL on rainfall. In Panel A, the rainfall variable is total rainfall. In Panel B, the rainfall variable is the coefficient of variation of rainfall.

Figure 5: Yearly Rainfall in Waves One and Two



Appendix

Table A1: Month of Interview by Type of Enterprise

	(1) Agriculture	(2) Non-Farm
January	371	221
February	409	278
March	434	334
April	363	463
May	392	480
June	409	435
July	458	410
August	908	283
September	1442	593
October	1038	482
November	425	261
December	338	179

Counts are the number of observations in the restricted sample (households that operate both types of enterprises in the same wave) and when they responded to the agricultural module (first column) and non-farm module (second column). Cross-section households responded to both modules in the same sitting. Approximately half of panel households responded in the same sitting while the other half responded at different times.

Table A2: MRPL Estimates without Trimming

	Pooled Production Function		Separate Production Functions		Collapsed
	(1)	(2)	(3)	(4)	(5)
	Simple Median	Weighted by Labor	Simple Median	Weighted by Labor	Collapsed
Ag MRPL (Median)	150.591***	99.145***	197.038***	134.347***	112.322***
	(25.295)	(11.808)	(47.375)	(22.149)	(14.244)
NF MRPL (Median)	32.378	31.392**	40.383	37.789	35.553**
	(20.203)	(13.454)	(38.192)	(25.424)	(15.576)
Difference (Median)	114.029***	65.439***	137.176**	81.493**	75.230***
	(39.011)	(20.672)	(69.210)	(36.798)	(25.128)
N (Household/Wave)	4180	418	4180	4180	4180

Standard errors are constructed through bootstrapping the MRPL construction 1,000 times. The bootstrap is set to draw households. MPRL estimates are in 2010 MWK. MRPL difference is constructed as agricultural MRPL minus non-farm MRPL. The MRPLs in the first and second columns are constructed by estimating a single production function but allowing the effects of all variables (except household fixed effects) to vary by enterprise type. MRPLs in the third and fourth columns are constructed by estimating separate production functions for agriculture and non-farm production. In the fifth column, MRPL is constructed by collapsing across plots and enterprises to the household/wave level and estimating a single production function, again allowing the effects of all variables (except household fixed effects) to vary by enterprise type. Unlike our main estimates, we do not trim the top one percent of labor and revenue prior to estimating production functions. * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$

Table A3: MRPL Estimates Using Actual Revenue Instead of Predicted Revenue

	Pooled Production Function		Separate Production Functions		Collapsed
	(1)	(2)	(3)	(4)	(5)
	Simple Median	Weighted by Labor	Simple Median	Weighted by Labor	Collapsed
Ag MRPL (Median)	133.007***	86.289***	180.967***	116.759***	98.435***
	(19.472)	(8.699)	(32.658)	(14.675)	(9.877)
NF MRPL (Median)	21.931	20.840**	19.888	18.744	31.826***
	(15.547)	(10.586)	(29.825)	(19.841)	(12.064)
Difference (Median)	110.104***	63.762***	142.224**	83.901***	65.848***
	(31.569)	(16.234)	(56.740)	(29.172)	(18.024)
N (Household/Wave)	4097	4097	4097	4097	4097

Standard errors are constructed through bootstrapping the MRPL construction 1,000 times. The bootstrap is set to draw households. MPRL estimates are in 2010 MWK. MRPL difference is constructed as agricultural MRPL minus non-farm MRPL. The MRPLs in the first and second columns are constructed by estimating a single production function but allowing the effects of all variables (except household fixed effects) to vary by enterprise type. MRPLs in the third and fourth columns are constructed by estimating separate production functions for agriculture and non-farm production. In the fifth column, MRPL is constructed by collapsing across plots and enterprises to the household/wave level and estimating a single production function, again allowing the effects of all variables (except household fixed effects) to vary by enterprise type. Unlike our main estimates, we use actual revenue instead of predicted revenue for these estimates. * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$

Table A4: MRPL Estimates with 10 Prices Observations

	Pooled Production Function		Separate Production Functions		Collapsed
	(1)	(2)	(3)	(4)	(5)
	Simple Median	Weighted by Labor	Simple Median	Weighted by Labor	Collapsed
Ag MRPL (Median)	121.732***	78.485***	166.973***	112.578***	94.436***
	(17.591)	(8.301)	(30.494)	(14.341)	(10.290)
NF MRPL (Median)	23.407	22.421**	21.691	20.539	31.940**
	(16.299)	(10.908)	(28.891)	(19.104)	(12.796)
Difference (Median)	90.160***	50.041***	129.881**	76.199***	59.107***
	(29.395)	(15.301)	(51.857)	(27.753)	(18.700)
N (Household/Wave)	3985	3985	3985	3985	3985

Standard errors are constructed through bootstrapping the MRPL construction 1,000 times. The bootstrap is set to draw households. MPRL estimates are in 2010 MWK. MRPL difference is constructed as agricultural MRPL minus non-farm MRPL. The MRPLs in the first and second columns are constructed by estimating a single production function but allowing the effects of all variables (except household fixed effects) to vary by enterprise type. MRPLs in the third and fourth columns are constructed by estimating separate production functions for agriculture and non-farm production. In the fifth column, MRPL is constructed by collapsing across plots and enterprises to the household/wave level and estimating a single production function, again allowing the effects of all variables (except household fixed effects) to vary by enterprise type. Unlike our main estimates, we require a minimum of ten price observations for the estimates in this table. * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$

Table A5: MRPL Estimates with no Minimum Price Observations

	Pooled Production Function		Separate Production Functions		Collapsed
	(1)	(2)	(3)	(4)	(5)
	Simple Median	Weighted by Labor	Simple Median	Weighted by Labor	Collapsed
Ag MRPL (Median)	144.143***	94.083***	201.662***	134.372***	101.685***
	(21.437)	(9.789)	(37.520)	(16.937)	(12.685)
NF MRPL (Median)	26.975	25.716**	27.823	26.133	32.661**
	(17.082)	(11.549)	(29.991)	(19.986)	(12.836)
Difference (Median)	114.588***	65.085***	158.979***	93.264***	69.563***
	(33.120)	(17.264)	(57.004)	(29.630)	(20.663)
N (Household/Wave)	4160	4160	4160	4160	4160

Standard errors are constructed through bootstrapping the MRPL construction 1,000 times. The bootstrap is set to draw households. MPRL estimates are in 2010 MWK. MRPL difference is constructed as agricultural MRPL minus non-farm MRPL. The MRPLs in the first and second columns are constructed by estimating a single production function but allowing the effects of all variables (except household fixed effects) to vary by enterprise type. MRPLs in the third and fourth columns are constructed by estimating separate production functions for agriculture and non-farm production. In the fifth column, MRPL is constructed by collapsing across plots and enterprises to the household/wave level and estimating a single production function, again allowing the effects of all variables (except household fixed effects) to vary by enterprise type. Unlike our main estimates, we do not require any minimum number of price observations for these estimates. * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$