The Gender Gap in Agricultural Productivity: The Role of Market Imperfections

Amparo Palacios-López (corresponding author)
Development Research Group, The World Bank
Phone: (202) 473 3263
apalacioslopez@worldbank.org

Ramón López
University of Maryland, College Park
Department of Economics, University of Chile, Santiago
rlopez@umd.edu

Acknowledgments

The authors would like to thank Vivian Hoffmann, Asif Islam, Talip Kilic, Patricio Korzeniewicz, Howard Leathers, Kenneth Leonard, Kabir Malik, Michael Settels, and the seminar participants at the Agricultural and Resource Economics Department of the University of Maryland for their comments on earlier versions of the paper.
This paper hypothesizes that labor and credit market imperfections - by discouraging off-farm income-generating activities and restricting access to inputs, respectively- affect female farm productivity more deeply than male productivity. The paper develops a theoretically model which decomposes the contribution of various market imperfections to the gender productivity gap. Empirically we show that agricultural labor productivity is on average 44 percent lower on female-headed plots than on those managed by male heads. Thirty four per cent of this gap is explained by differences in labor market access and 29% by differences in credit access.

**Keywords:** Gender, Agriculture, Productivity, Decomposition Methods, sub-Saharan Africa, Malawi.

**JEL Classification:** C21, J16, Q12.
1. INTRODUCTION

This paper hypothesizes that labor and credit market imperfections - by discouraging off-farm income-generating activities and restricting access to inputs, respectively - affect female farm productivity more deeply than male productivity, thus increasing the gender productivity gap. The paper theoretically models the relationship between gender differences in agricultural labor productivity and market imperfections and provides empirical evidence for Malawi consistent with the theoretical model. We find that agricultural labor productivity in Malawi is on average 44 percent lower on plots managed by female heads of household than on those managed by male heads of household; and that 34 percent of the observed agricultural labor productivity gap is explained by spillovers from labor market gender differences and 29 percent is explained by gender differences in the use of purchased inputs.

Two key features are prevalent in agriculture in Sub-Saharan Africa: the dominance of female labor in agriculture and the existence of a gender gap in agricultural productivity. In Sub-Saharan Africa women make up a higher proportion of agricultural labor than males, ranging from 30 to 60 percent of the total labor (UNECA 1982, FAO 1984, Doss 2011, Doss et al. 2011). In Malawi, in a 2004 survey, rural wages were 35% lower for females than males. Furthermore, around 89% of women working off-farm do so part time in contrast to only 67% of males. Women participating in rural wage employment tended to be concentrated in lower skill activities - about 61.4% - in contrast to the corresponding figure of 37% for males (Hertz et al. 2009).
Estimates of gender differentials in agricultural productivity range from 4 to 40 percent depending on the country, the representativeness of the data, the empirical approach used, the type of crop, among other variables (Akresh, 2005; Alene et al., 2008; Gilbert et al., 2002; Goldstein and Udry, 2008; Moock 1976; Peterman et al., 2011; Oladeebo and Fajuyigbe, 2007; Quisumbing et al., 2001; Saito et al., 1994; Tiruneh et al., 2001; Udry, 1996; Vargas Hill and Vigneri, 2011). The large participation of women in agriculture and the significant gender productivity gap are prime motivations for this study. Both of these factors influence the larger issue at hand which is the gender disparity of household income.

Several reasons for the gender productivity gap have been identified in the literature: differences in (i) access and use of agricultural inputs, (ii) tenure security and related investments in land and improved technologies, (iii) market and credit access, (iv) human and physical capital, and (v) informal and institutional constraints affecting farm/plot management and marketing of agricultural produce (Peterman et al. 2011). However, the relationship between labor market discrimination and the observed gender gap in agricultural productivity has received little attention in the literature.

The literature has acknowledged the presence of multiple market failures in agriculture, especially in the labor market. López and Romano (2000) conclude that rural income may be improved by developing labor and credit markets. Several studies have focused on the channels by which rural household income diversification can alleviate rural poverty (FAO, 1998; Lanjouw and Lanjouw, 2001; Haggblade et al., 2007; Winters et al., 2009). The general theme is that labor supply behavior is affected by risk, search and transaction costs, location preferences, gender preferences, and gender discrimination (Barret, 1996; Biswanger and Rosenzweig, 1986).
This drives a wedge between the marginal product of labor and the prevailing market wage rate for the same type of labor (Barret et al. 2008). An important contribution of this paper is to show theoretically and empirically that the wedge between the marginal product of labor and the prevailing market wage rate varies by gender and is generally larger for women.

The literature on gender discrimination has typically followed one of two avenues. One explores labor market discrimination in terms of wages in off-farm activities (O’Neill & O’Neill, 2006, Fortin, 2006). The other explores gender differences in the agricultural sector with a focus on issues related to inputs, credit access, market access, and cultural constraints (Peterman et al. 2011). Our study contributes to the literature by exploring the spillover effects of gender discrimination in the labor market on the gender gap in on-farm productivity.

This is one of the first studies to both theoretically and empirically decompose the sources of gender differences in agricultural labor productivity relating them to credit and labor market imperfections. Credit markets may treat women and men differently in a discriminatory fashion which causes women to have less access to purchased inputs (FAO 2011b). Labor market imperfections may impact women and male wages differently resulting in women receiving lower wages from off-farm activities than men (FAO 2011b, Hertz et al. 2009, Doss 2011, Doss et al. 2011). As a consequence, women work less in off-farm activities and more on the farm than men therefore creating farm productivity differentials. Also, the typical requirements of women’s domestic activities may mean that off-farm work implies greater sacrifices for them than for men. This may cause women to allocate more of their time to on-farm work causing their agricultural labor productivity to be lower than men’s.
The theoretical model of household farm productivity in this study deviates from the literature in two ways. It examines agricultural labor productivity subject to the head-of-household’s gender and traces the effects of labor market imperfections and credit market constraints on productivity. The model’s inclusion of “gender-specific effective off-farm work time” (GSEOW) borrows from studies by López (1984, 1986), and is defined as the total time required to perform off-farm work which includes not only the actual working time but also commuting time faced by men and women. López (1984, 1986) examines how time allocation between on-farm and off-farm work has different connotations on utility as a consequence of commuting time required by off-farm work. He shows how the consideration of commuting time leads to a model in which the household acts as having different preferences between on-farm and off-farm work, even if preferences are defined purely in terms of leisure. In this setting the optimization problem of the household becomes non-separable and thus the production decisions would be directly linked to the household’s characteristics and consumption decisions. Our model predicts that GSEOW leads to labor productivity differences between men and women, with women as heads of household devoting a higher proportion of their time to agricultural activities.

Predictions derived from the model are empirically illustrated using nationally representative data from Malawi. We explore gender differences in productivity at the plot level. We specifically compare plots managed by female heads of household with plots managed by male heads of household. Our examination of gender differences mostly pertains to female-headed households in which a male spouse is absent.
The econometric approach adapts a decomposition method from labour economics, most notably in the analyses of the gender wage gap, union wage gap, and growing wage inequality (Oaxaca, 1973 and Blinder, 1973). Our study specifically decomposes the average differences in agricultural labour productivity between plots managed by male-headed households and those managed by female-headed households into four effects: (i) labour market effect, (ii) purchased inputs effect, (iii) endowment effect, and (iv) pure marginal productivity effect which is the gender differences in coefficients of the various factors of production and household characteristics.

2. THEORETICAL MODEL

Farmers in developing countries face budget and working capital constraints. Given that they experience limited access to credit, they can alleviate credit constraints by generating income from off-farm labor activities. The labor allocation across on-farm and off-farm work, as well as the type and quantity of non-labor agricultural inputs chosen, are central to the productivity of the farm.

2.1 HOUSEHOLD WELFARE MAXIMIZATION

We posit that all households have the same preferences. We assume that the utility function is additive and increasing in the present value of earnings \(Y\) and leisure \(l\) (Eswaran and Kotwal, 1986).

\[
U(Y, l; Z_{hh}) = Y + u(l)
\]  

(1)

where \(Z_{hh}\) is a vector of exogenous household characteristics.
The household allocates its time endowment \((H)\) between leisure \((l)\), on-farm labor \((L_f)\), and off-farm labor \((L_o)\). We model the existence of a GSEOW (López, 1984, 1986) which is determined by the structure of the household. GSEOW as defined here includes commuting time as well as actual off-farm work time. The intuition is as follows. An increase in off-farm time is not simply an equivalent subtraction of hours spent on farm work or leisure. There is an additional time cost incurred due to several reasons. For instance households may have to alter their schedule to accommodate off-farm work. This cost of re-organization of activities may increase with every increase in time spent in off-farm activities. Furthermore, there may be synergies between on-farm work and household work including child bearing. Thus a reduction in time allocated to the farm may incur an additional cost increasing with every additional unit of time spent in off-farm work. Importantly such effective off-farm work time may vary by gender. Female headed-households may face higher commuting time burden due to household care responsibilities that are culturally assigned to women (child care, cooking, getting water, etc.).

Given the above considerations, and contrary to the standard practice, we may regard the time constraint faced by the households as non-additive as follows:

\[
H = l + L_f + g(L_o)
\]

where \(g(L_o)\) is “effective off-farm work time” and \(g(L_o) = \alpha_o + \alpha_i L_o\); \(\alpha_o\) is fixed time incurred when the household participates in off-farm activities, which is the same for both male and female headed households, and \(\alpha_i \geq 1\) represents the portion that is determined by the degree of household care activities; \(\alpha_i = 1\) implies that the household head has the lowest household care responsibilities. We assume that the gender-specific effective off-farm work time is higher for female than male heads of household, \(\alpha_i^F > \alpha_i^M\).
On-farm production requires the use of two variable inputs: labor ($L_f$) and non-labor inputs ($X$). Non-labor inputs include inorganic fertilizer, improved seeds, and traditional seeds. For simplicity we assume that the farmer uses only household labor. The production function is presented as follows: $f(X, L_f; Z_{prod})$ where $Z_{prod}$ is a vector of exogenous farm characteristics.

Net income is equal to the sum of revenues from all sources minus household expenditures:

$$Y = w^*L_o + pf(X, L_f; Z_{prod}) - r^*X$$ \hspace{1cm} (3)

where $w^*$ is the wage rate received in off-farm activities, $r$ is the price of non-labor input $X$, and $p$ is the price of output. All prices are exogenous.

In addition, the household faces a working capital (liquidity) constraint where the inputs purchased are less than or equal to the amount borrowed plus the income from off-farm activities. All expenses are incurred at the beginning of the production period. Hence,

$$r^*X \leq B + wL_o$$ \hspace{1cm} (4)

For simplicity, the model assumes that male and female-headed households have the same demographic composition.

The household maximizes its utility by allocating labor between on-farm, off-farm work, and deciding how much non-labor input they will use by solving the following maximization problem:

$$\max_{L_o,L_f,X} V = w^*(L_o) + pf(X, L_f; Z_{prod}) - X^*r + u(H - L_f - (\alpha_o + \alpha_iL_o))$$ \hspace{1cm} (5)

Subject to

$$r^*X \leq B + wL_o$$
The following Lagrangean follows from (5):

\[
\max_{L_o, L_f, X} V = w^*(L_o) + pf(X, L_f; Z_{prod}) - X^* r + u(H - L_f - (\alpha_o + \alpha_1 L_o)) + \psi(B + wL_o - X^* r) \tag{5a}
\]

where \( u \) and \( \psi \) are the shadow prices of time and credit, respectively. Monotonicity of the utility function implies that the working capital constraint will be binding.

The First Order Conditions (FOC) are:

\[
\frac{\partial V}{\partial L_o} = w - U_l \alpha_1 + \psi w \leq 0 \tag{6a}
\]

\[
\frac{\partial V}{\partial L_f} = -U_l + pf_{L_f} \leq 0 \tag{6b}
\]

\[
\frac{\partial V}{\partial X} = pf_x - r - \psi r \leq 0 \tag{6c}
\]

\[
\frac{\partial V}{\partial \psi} = B + wL_o - rX \leq 0 \tag{6d}
\]

where \( U_l \) denotes marginal utility of leisure and \( f_x \) and \( f_{L_f} \) denote the marginal productivity of input \( X \) and on-farm labor respectively.

2.2 IMPLICATIONS FOR LABOR ALLOCATION

Under standard competitive capital and labor markets, the gender-specific effective off-farm work time would not affect the labor allocation decisions of the household, and thus the model would be separable. Therefore, the household’s production decisions would be independent from its consumption decisions and its composition; household characteristics
would play no role in the labor supply of the household. Thus, the household would allocate labor according to (from 6a and 6b):

$$w = pf_{L_f}$$  \hspace{1cm} (7)$$

We now assume that the wage paid for off-farm activities is different for men and women, thus $w^f = w(1 - \phi)$, where $0 \leq \phi \leq 1$ represents a gender discrimination factor in the labor market. The higher is $\phi$ the greater is the degree of discrimination for women.

Additionally, the shadow price of the working capital constraint $\psi(Z_{hh})$ is a function of the household characteristics; hence it will be different for male and female headed households. The household will allocate labor between on-farm and off-farm work as follows (from 6a and 6b):

Male headed households:

$$\frac{w(1 + \psi(Z_{hh}^M))}{\alpha_1^M} = pf_{L_f}^M$$  \hspace{1cm} (8)$$

Female headed households:

$$\frac{w(1 - \phi)(1 + \psi(Z_{hh}^F))}{\alpha_1^F} = pf_{L_f}^F$$  \hspace{1cm} (9)$$

Consequently, the household will allocate labor towards off-farm work until the net benefit from off-farm work is equal to the net benefit from on-farm work. Allocation of labor towards off-farm work will alleviate the working capital constraint, which can be shown through the effect of $(1 + \psi)$ on equation (8) or (9). The value of every hour worked in off-farm activities not only provides income but also relieves the working capital constraint. The marginal
productivity of labor is an endogenous function of household characteristics, preferences, assets and labor market discrimination. The model is thus non-separable.

Equations (8) and (9) show us that female-headed households tend to allocate more labor to on-farm work than male-headed households despite lower farm labor productivity of women compared to men.

The level of input $X$ used also depends on the degree of the household’s credit constraint. From (6c) we know that the more constrained the households are the less they will invest in $X$:

$$r(1 + \psi) = pf_x$$

The net effect of $\alpha_i$ on on-farm labor productivity is negative. A higher $\alpha_i$ causes an increase of on-farm labor as off-farm work is lower. Also, the use of non-labor inputs decreases with $\alpha_i$ because less off-farm income is received, thus exacerbating the working capital constraint. To sum up, the key mechanism is that fewer off-farm opportunities for women due to gender-specific effective off-farm work time as well as labor market discrimination leads to lower off-farm income, essentially exacerbating any pre-existing liquidity constraints faced by female-headed households. The end result is female-managed plots are quite likely to have fewer non-labor inputs and more on-farm work, ultimately reducing their on-farm labor productivity.

2.3 THE GENDER GAP
Dividing (8) over (9) we obtain a measure of differences in agricultural labor productivity or gender gap:

\[
\frac{f_{L_i}^M}{f_{L_i}^F} = \frac{\alpha_i^F}{\alpha_i^M (1 - \phi)} \frac{(1 + \psi(Z_{hh}^M))}{(1 + \psi(Z_{hh}^F))}
\]

The total labor market effect (gender discrimination and the gender specific effective off-farm work time) has a positive effect on the gender gap in agricultural labor marginal productivity. The first term in equation (11) is unambiguously greater than 1 given that \(\alpha_i^F > \alpha_i^M\), and \(0 \leq \phi \leq 1\) which implies that male agricultural labor productivity is higher than that of female-headed households. Female-headed households will still remain in farm activities even when the return is lower than the return in alternative off-farm activities. This is because they have to account for specific household labor activities they are required to fulfill, and for the differential treatment with regards to off-farm labor opportunities.

The second term in equation (11) reflects the effect of liquidity constraints on the gender gap in agricultural labor productivity, which is ambiguous. The difference in response to liquidity constraints may be different for male- and female-headed households, but the direction of the total effect remains an empirical question. If women-headed households are more discriminated in the credit market than male-headed households the resulting tighter liquidity constraint may reduce the on-farm inputs and increase the agricultural labor productivity gap. However, the tighter credit constraint may cause female-headed households to value more working in off-farm activities than male-headed households, thus encouraging a greater allocation of their labor to off-farm work. This indirect effect may reduce the gender productivity
gap. However, it is reasonable to assume that this second order effect will be dominated by the direct effect of credit constraint in which case male-headed households will have higher productivity than female-headed households and the gender gap will be enhanced.

2.4 PROPOSITIONS

The following proposition emerges from the theoretical model:

Higher GSEOW and labor market discrimination in capital-constrained households cause a greater expansion on farm work and lower off-farm labor and off-farm income in female-headed than in male-headed households. The greater reduction of off-farm labor income in female-headed households causes a greater reduction of the liquidity required to buy non-labor inputs than in male-headed households. This reduced liquidity effect may cause greater need for credit which in the context of credit market imperfections imply a larger negative impact of such imperfections for female than male households. Thus, labor market restrictions and credit market imperfections contribute to increase the farm productivity gender gap.

3. EMPIRICAL ANALYSIS

The econometric approach we use has been utilized in labor economics as part of the analyses of the gender wage gap, union wage gap, and growing wage inequality (O’Neill & O’Neill, 2006, Fortin, 2006). We use the mean decomposition methodology to look at the differences in agricultural productivity for male- and female-headed households. Regression-
based decomposition methods have been widely utilized in labor economics (Oaxaca 1973; Blinder 1973). This method however does require strong assumptions (Fortin et. al., 2011, Kilic et al., 2013).

Decomposition methods follow a partial equilibrium approach, where observed outcomes for one group can be used to construct various counterfactual scenarios for the other group (Fortin et. al., 2011). Decomposition methods are based on correlations, and hence cannot be interpreted as estimates of underlying causal parameters. However, decomposition methods do document the relative quantitative importance of factors in explaining an observed gap, thus suggesting priorities for further analysis and, ultimately, policy interventions (Fortin et. al., 2011, Kilic et al., 2013).

We regress $Y$, the log of value of output per hectare (land productivity) for male- (M) and female- (F) headed household plots, on its determinants as expressed by the following equations:

$$Y^M = \beta_0^M + L^M \beta_L^M + \sum_k X_k^M \beta_k^M + \sum_j Z_j^M \delta_j^M + \varepsilon^M$$

$$Y^F = \beta_0^F + L^F \beta_L^F + \sum_k X_k^F \beta_k^F + \sum_j Z_j^F \delta_j^F + \varepsilon^F$$

where $L$ is the number of hours of managerial labor per hectare; $X$ is a vector of $k$ purchased inputs (pesticides, organic and inorganic fertilizer, hired labor, agricultural implements, improved seeds); $Z$ is a vector of characteristics of the household that includes human and physical capital (wealth, land assets, household composition, location of the household and location of the plot, access to off-farm income and transfers); $\beta_0, \beta_L, \beta_k, \delta_j$ are the associated vector of intercept and slope coefficients for male and female headed households; and $\varepsilon$ is the
error term under the assumption that $E(\varepsilon^M) = E(\varepsilon^F) = 0$. The decomposition of the gender gap in agricultural land productivity is presented on the online appendix A.

### 3.1 MEAN DECOMPOSITION OF THE GENDER PRODUCTIVITY GAP

We use the resulting vector of coefficients from the land productivity regressions indicated in equations (12a) and (12b) to create a measure of labor productivity (value of output per hour of managerial labor). We create measures of labor productivity in logarithm form by subtracting labor from land productivity $\ln(Y) = \ln(Y/ha) - \ln(L/ha)$.

We use equations (12a) and (12b) and subtract $L$ from both sides of the equation.

$$E(Y^M) - E(L^M) = \beta_0^M + E(L^M) \beta_L^M + \sum_k E(X_k^M) \beta_k^M + \sum_j E(Z_j^M) \delta_j^M - E(L^M)$$

$$E(Y^F) - E(L^F) = \beta_0^F + E(L^F) \beta_L^F + \sum_k E(X_k^F) \beta_k^F + \sum_j E(Z_j^F) \delta_j^F - E(L^F)$$

The gender gap in labor productivity “$D_L$” is expressed as the mean outcome difference:

$$D_L = [E(Y^M) - E(L^M)] - [E(Y^F) - E(L^F)]$$

Substituting (13a) and (13b) into (14) and adding and subtracting $E(L^M) \beta_L^F$, $\sum_k E(X_k^M) \beta_k^F$ and $\sum_j E(Z_j^M) \delta_j^F$, we decompose the gender gap in labor productivity into the following four components:
$$D = [E(L^M) - E(L^F)](\beta_L^F - 1) + \sum_k [E(X_k^M) - E(X_k^F)](\beta_k^F) + \sum_j [E(Z_j^M) - E(Z_j^F)]\delta_j^F +$$

$$E(L^M)[\beta_L^M - (\beta_L^F)] + \sum_k E(X_k^M)[\beta_k^M - \beta_k^F] + (\beta_0^M - \beta_0^F) + \sum_j E(Z_j^M)[\delta_j^M - \delta_j^F]$$

(15)

In practice, we estimate equation (12a) and (12b) using the value of output per hectare as the outcome variable. We use the resulting vector of coefficients, in combination with the mean values for each covariate of the male and female samples to compute the components of equation (15).

The difference between the decomposition of the gender gap in terms of land productivity and labor productivity is that the disparity in hours worked by men and women is weighted by \((\beta_L^F - 1)\), which is the measure of the elasticity of labor in female plots. Hence, in this case, the fact that women work more than men exacerbates the average labor productivity gap between men and women. In contrast, the increase in on-farm work by women relative to men actually increases land productivity albeit in an inefficient way.

The first component of equation (15) is the labor market effect, i.e. the portion of the gender gap driven by differences in quantities of labor allocated to on-farm work by the head of household. The second component is the purchased inputs effect, the portion of the gender gap that is explained by differences in levels of use of inputs that have to be bought such as fertilizer, pesticides, seeds, agricultural implements, and/or hired labor. The third component, the household endowment effect is comprised by differences in levels of observable characteristics of
the household, including human and physical capital. The fourth component is the *pure marginal productivity effect* and corresponds to the portion of the gender gap explained by differences in the coefficients of each observable covariate included in $L$ and in the $X$ and $Z$ vectors; as well as differences in the constant between male- and female-headed households.

We provide an explanation of the gender gap in land productivity in the online appendix A and a graphical representation of the gender gap in land and labor productivity and their respective components in Figure B1 in the online appendix B.

3.2 EMPIRICAL ISSUES

The decomposition methods described above are valid only under assumptions of (i) overlapping support and (ii) ignorability. The overlapping support assumption rules out cases where observable and unobservable covariates may be different across the two groups. Hence “overlap” refers to the similarity of the covariate distributions of both subpopulations. It implies that no single value of the covariates ($X, Z, L$) attain specific values ($X = x, Z = z, L = l$) or $\epsilon = e$ exists to identify female plot management.

Ignorability refers to the random assignment of female plot management conditional on observable attributes (Kilic et al., 2013). Specifically we worry that our male and female managed plots may not be randomly assigned. Ignorability allows us to assume that we have enough controls and thus, conditional on these controls, our assignment of female plot management is essentially randomized. It rules out what we typically call “self-selection” based
on unobservables. The additional essential assumptions required by detailed decomposition of
the individual contribution of each covariate include additive linearity and zero conditional
mean. The latter implies that \( \varepsilon \) is independent of the explanatory variables. In other words, we
assume that there is no unobservable heterogeneity that jointly determines the outcome and
observable attributes. The former assumes a linear functional form.

In exploring the existence and extent of the gender gap in a multivariate framework, the
validity of findings largely depends on the plausibility of the ignorability and zero conditional
mean assumptions, i.e. the extent to which the estimation strategy addresses possible
unobservable household-/plot-level heterogeneity that jointly determines plot agricultural
productivity and observable covariates, including whether a plot belongs to a household headed
by a female. We attempt to lend as much support to the assumptions of overlapping support,
ignorability, and zero conditional mean as possible by applying the added control approach and
checking whether the estimations are robust to a range of sample alteration in order to see if the
coefficients of interest change due to omitted variable bias. These sensitivity analyses are
presented in the online appendix D.

Additionally, we consider the possibility of reverse causality. This is less of a concern for
inputs as it is widely accepted in the literature that agricultural inputs may be regarded as
predetermined vis-à-vis the level of output (Griliches, 1963; Dinar et al., 2007). This is due to the
fact that agricultural production takes time to be completed and inputs are applied at the
beginning of the season while the corresponding output is harvested at the end. It seems
reasonable to assume that there is no correlation between the stochastic error and the predetermined inputs.

4. DATA

This study uses data from the Third Integrated Household Survey (IHS3), collected from March 2010 to March 2011 by the Malawi National Statistical Office, with support from the World Bank Living Standards Measurement Study - Integrated Surveys on Agriculture (LSMS-ISA) project. The IHS3 data were collected within a two-stage cluster sampling design, and are representative at the national, urban/rural, regional, and district levels, covering 12,271 households in 768 enumeration areas (EAs). The IHS3 instruments included Household, Agriculture, Fishery, and Community Questionnaires.

All sample households were administered the multi-topic Household Questionnaire. The sample households that were involved in agricultural activities (through ownership and/or cultivation of land, and/or ownership of livestock) were administered the Agriculture Questionnaire. The Agriculture Questionnaire solicited information on land areas, physical characteristics, labor and non-labor input use, and crop cultivation and production at the plot level, separately for the reference rainy and dry seasons. Table C1 in the electronic Appendix C presents the definitions of the variables used in the analysis.

5. RESULTS
5.1 DESCRIPTION OF GENDER DIFFERENCES

The descriptive statistics and the results from the tests of mean differences sorted by the head of household’s gender are presented in Table 1. The full sample consists of 14,044 plots managed by the head of household, 23 percent of them are female headed households. The sample has been restricted to include only plots that are managed by the head of household and in which the manager works at least one hour per day on the plot. In Malawi where households cultivate an average of 1.8 plots, most of the plots (94%) are managed by the head of household. Only 6.8% of households in the sample have plots that are managed by an individual other than the head of household. There are only 1.2% of households in which different plots are managed by more than one individual – some managed by the head and others by another household member.

It is important to note that female headed households include mostly divorced, separated, single, or never married women, which together account for 87 percent of all female headed households. The remaining 13 percent of female headed households are married. Thus our examination of gender differences mostly pertains to female headed households in which a male spouse is absent.

Table 1 provides evidence of the gender gap: the average of the log gross value of output per hour of managerial labor is 44 percent lower for female plots, while the average of the log of gross value of output per hectare is 24 percent lower for female plots. This result provides some support for our model which predicts a gender gap in labor productivity that is larger than the gender gap in land productivity. The reasons posited for this difference in the theoretical model
include labor market imperfection spillovers that lead to female-headed households allocating more labor to the farm thus decreasing labor productivity but increasing land productivity. The gender differences in agricultural productivity are also evident in the comparison of the Kernel density estimates of the log of land productivity for male and female plots, as displayed in Figure 1. The Kernel density estimates for the log of labor productivity for male and female plots are displayed in Figure 2.

The overwhelming majority of the differences in the average values of the observable covariates across male- vs. female plots in Table 1 are statistically significant at the 1 percent level. Female plots are, on average, overseen by individuals that are 8 years older and have 2.5 less years of schooling with respect to their male counterpart. A significantly higher percentage of female plots exhibit manager-owner correspondence than male plots (78 vs. 58 percent).ii

The average GPS-based plot area for female plots is 9 percent smaller than male plots. The use of inorganic fertilizer per hectare is in average 18 percent lower on female plots than on male plots.

In terms of household labor use, the dynamics are also highly different on female plots vis-à-vis their male comparators, as can be seen in Table 1. The average amount of hours of managerial labor per hectare is 22 percent higher on female plots than on male plots, while female plots have 34 percent less family labor than male plots. Female plots have higher levels of exchange labor use, while hired labor is not significantly different between male and female plots. Female plots are also 5 percentage points less likely to be associated with households that
receive agricultural extension service on topics that relate to crop production and marketing. Female heads of household are 10 percent less likely to participate in off-farm work than male-headed households. Lastly, male-headed households are, on average, more likely to be associated with higher levels of wealth and access to agricultural implements.

5.2 BASE REGRESSION RESULTS

Table 2 presents plot-level land productivity regression results for the male and female plot samples in columns 1 and 2 respectively. We include the explanatory variables that have been widely used in the literature (López 1984, López 1986, Peterman et al. 2011) including plot area, labor and non-labor inputs, plot characteristics, and farm attributes including type of crop and the presence of inter-cropping which are expected to have a direct effect on productivity. We also include covariates capturing manager and household characteristics which may affect agricultural productivity if the consumption and production decisions of the household are non-separable such as whether the head of household works off-farm and whether the household has an adult male household member, different from the manager (in the case of male-headed households). Additionally, we include district level effects to account for time-invariant omitted variables at the district level.

The land productivity regressions’ results are as expected; labor and non-labor inputs contribute positively to agricultural land productivity; while the area of the plot has a negative sign that reflects decreasing returns to scale of the production function. Additionally, household and manager characteristics are significant, possibly reflecting the non-separability nature of the
optimization problem caused by the liquidity constraints and the labor market imperfections faced by households, as explained in the theoretical model in section 2.

We now turn to the effect of different covariates on the production of male and female plots. Table 2 shows that only six coefficients are significantly different between the male and female regressions, at the 10 percent level of significance. The coefficients that are significantly different are inorganic fertilizer, area of the plot and area squared, exchange labor, child dependency ratio and extension.

Plot area has a negative coefficient that is statistically significant at the 1 percent level in the male and female samples; however the coefficient in the female regression is more than twice as high as in the male regression. We calculate the marginal productivity of land, which is shown in Table 3 and find that it is 27 percent lower in female managed plots compared to male-managed plots.

The log of inorganic fertilizer use per hectare is positively associated with the log of gross value of output per hectare, irrespective of the plot sample. However, the return to inorganic fertilizer use (i.e. the coefficient of inorganic fertilizer) is higher within the male plot sample than within the female.

The log of managerial labor has a positive coefficient that is statistically significant at the 1 percent level across the male and female samples. As shown in Table 3, the marginal productivity of female managers is 37 percent lower than the labor productivity of male
managers. The coefficient of household labor is positive and significant in the female and male plots, but larger on male plots, while the coefficient of exchange labor is only significant on the male plot sample.

The child dependency ratio, which is defined as the number of household members below the age of 10 divided by the number of household members aged 10 years and above, has a substantial negative coefficient that is statistically significant at the 1 percent level only within the female-managed plot sample. The comparable statistics for the male-managed plot sample is negative and statistically insignificant.

With respect to the household characteristics, household size has a positive coefficient that is statistically significant irrespective of the plot sample; the magnitude of the coefficient within the female plot sample is larger than within the male-managed plot sample. The gender differences in returns to child dependency ratio after controlling for household size imply that the burden of childcare is more likely to reduce female agricultural productivity than male agricultural productivity. The distance of the plot to the household compound is negative and statistically insignificant for the male and female samples, while the distance of the household to the nearest road is negative and not statistically significant for the female sample alone.

5.4 MEAN DECOMPOSITION

In the decomposition methodology, we uncovered four components of gender differences in agricultural productivity – labor market, purchased inputs, household endowment and pure
marginal productivity effects outlined in section 3. We decompose the mean gender gap in agricultural land productivity and the agricultural labor productivity in order to rank the importance of each of the four components as indicated in (15) and (A.4) respectively. The decomposition uses the base regressions (section 5.2) which correspond to equations (12a) and (12b). We find that the decomposition results are consistent with the theoretical model. Gender gaps exist in favour of plots managed by male-headed households for both labor and land productivity, however the gender differences are far greater for labor productivity.

5.4. LAND PRODUCTIVITY DECOMPOSITION

As mentioned in the theoretical model, we predict that the agricultural productivity of female headed households will be lower when compared to male managed plots, the theoretical model also states that the gender gap in agricultural productivity is explained by how differently households are affected by liquidity and labor market constraints which are influenced by the composition of the household and its preferences. From the theoretical model we expect the gender gap to be decomposed into differences in levels of use of purchased inputs (male managed plots will have higher use), managerial labour (higher levels for female-managed plots) and differences in levels of the assets and characteristics of the household.

Table 4 shows the decomposition of the mean gender differential in agricultural land productivity into the four effects specified in equation (15). The gender gap in land productivity is estimated at 24 percent. The four effects are all significant at the 1 percent level.
The decomposition indicates that the purchased inputs and household endowment effects account for 52 and 36 percent of the total land gender gap respectively. According to our model, these two effects are a manifestation of the impact that liquidity constraints have on households as well as labor market imperfections, which are different for male and female headed households. Differences in purchased inputs reflect the direct effect of liquidity constraints and the indirect effect of labor market imperfections, while differences in the endowment of the household may be a result of the long term impact of such constraints.

The labor market effect is estimated to be -0.5 percentage points, which represents 19 percent of the gender gap. The latter means that if women worked the same number of hours as men, their land productivity would be lower and the gender gap in land productivity would rise to 30 percent. The fact that female managers work more on farm than male managers, implies that the land productivity will tend to reduce the male-female agricultural land productivity gap. This is so whenever labor and land are gross complements, which is apparently the case in the data sample.

The labor, endowment and purchased inputs effects account for 69 percent of the gender gap in land productivity. The remaining 31 percent of the gender gap in agricultural land productivity is explained by differences in coefficients (the pure marginal productivity effect), which includes the differential effect of the child dependency ratio, distance of the plot to the household and distance to the closest market, all of which are proxies to the gender specific effective off-farm work time.
5.4.B LABOR PRODUCTIVITY DECOMPOSITION

The theoretical model predicts specifically that labor productivity will be lower in female-headed households due to the different impact that the liquidity constraints, labor market discrimination and gender specific effective off-farm work time have over them compared to male-headed households (Equation (11)). As can be seen in Table 5, these predictions are consistent with the empirical results.

The gender gap in agricultural labor productivity is estimated at 44 percent. The agricultural labor productivity gender gap is larger than the agricultural land productivity gender gap, due to the labor market effect which accounts for 34 percent of the total gap. In this case the fact that female managers work more hours in the plot than male managers exacerbates the agricultural labor productivity gender gap, unlike its effect on land productivity, where it reduces the gap. The labor market effect may be attributed in part to the direct effect of gender differential treatment in the labor market and should be considered as an upper bound.

The purchased inputs and the household endowment effects are 29 and 20 percent of the gender gap respectively, which together with the labor effect explain 83 percent of the gender gap in agricultural labor productivity; the remaining 17 percent is explained by the pure marginal productivity effect.

So far we have performed the decomposition at the mean. However going beyond the “average” farmer and understanding the heterogeneity in constraints faced by farmers with
different gender and productivity profiles is crucial for the design and implementation of better targeted interventions aimed at bridging the gender gap. An important question is whether our findings, which are based on the sample means, are robust to the decomposition of alternative distributional statistics beyond the mean.

A method that is similar in spirit to the mean decomposition uses the recentered influence function (RIF) regressions proposed by Firpo et al. (2009) and provides a straightforward framework within which across-group differences in any distributional statistic could be decomposed. We rely on the RIF decomposition to provide estimates of the decomposition of the gender gap at different deciles of the agricultural productivity distribution. We find that when analyzing the gender gap at different deciles of agricultural labor productivity, the gender gap increases across the distribution, reaching a maximum value of 54 percent at the 70th percentile of the labor productivity distribution. Results are available on request from the authors.

We also conduct several sensitivity analyses to check the validity of the assumptions of overlapping support and ignorability. These include using the scale-free normalized difference of the regression covariates, an investigation of the robustness of the results to additional controls, and sample composition alterations. Due to space restrictions, we present these results in the online appendix and the tables are available on request.

7. CONCLUSION

This study presents a theoretical and empirical analysis on the mechanisms underlying gender differences in agricultural productivity. It focuses on the interactive effects of labor and
credit market imperfections. The empirical approach provides evidence of the relative quantitative importance of factors that lie behind the gender gap in agricultural labor productivity, providing evidence consistent with the predictions of the theoretical model.

This study theoretically and empirically uncovers the importance of market imperfections behind the gender gap in agricultural labor productivity. We have shown that liquidity constraints, labor market discrimination and effective off-farm work time, which differ greatly between men and women, result in lower agricultural labor productivity in plots managed by female- in comparison to male-headed households.

In terms of policies, this paper shows that correcting credit and labor market failures is likely to have a bigger impact on female-headed households than on male-headed households. That is, apart from the well-recognized overall efficiency improvement of these policies, one obtains additional gains by reducing gender differentials in agricultural productivity.

---

Footnotes:

1 Female plots refer to “plots managed by female heads of household”
2 The majority of the owned plots (81 percent) are acquired through inheritance. Another 12 percent is reported to have been granted by local leaders. The remaining are acquired as bride price (2 percent), purchased with title (1 percent) and purchased without title (2 percent).
REFERENCES


Table 1: Descriptive Statistics & Results from Tests & Mean Differences by Gender of the Plot Manager

<table>
<thead>
<tr>
<th>Sample size</th>
<th>Pooled Sample</th>
<th>Male Sample</th>
<th>Female Sample</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>14,044</td>
<td>10,822</td>
<td>3,222</td>
<td></td>
</tr>
</tbody>
</table>

**Outcome Variable**

- **Ln Value Output per hectare**: 10.42, 10.47, 10.23, 0.2 *****
- **Ln Value Output per hour of managerial labor**: 4.56, 4.66, 4.22, 0.4 *****

**Plot Managerial Labor Input Use**

- **Ln Managerial Labor (hours/ha)**: 5.9, 5.8, 6.0, -0.2 *****

**Purchased Inputs**

- **Pesticide/herbicide use yes/no**: 0.019, 0.021, 0.012, 0.009 *****
- **Organic Fertilizer use yes/no**: 0.112, 0.114, 0.106, 0.008
- **Ln Inorganic Fertilizer (kg/ha)**: 3.274, 3.317, 3.130, 0.2 *****
- **Ln Hired labor (days/ha)**: 0.545, 0.552, 0.521, 0.0
- **Agricultural implements Asset Index**: 0.682, 0.838, 0.158, 0.7 *****
- **Proportion of area of the plot under improved varieties**: 0.372, 0.384, 0.333, 0.1 *****
- **Proportion of area of the plot under export crops**: 0.078, 0.093, 0.028, 0.1 *****

**Endowment of the Household**

- **Ln Area of the plot (ha)**: -1.223, -1.201, -1.295, 0.094 *****
- **Ln Area of the plot (ha) Squared**: 1.975, 1.921, 2.156, -0.235 *****
- **elevation (m)**: 892.5, 907.2, 843.0, 64.2 *****
- **plot distance to hh**: 1.973, 2.063, 1.670, 0.393 **
- **Inter-croppped**: 0.353, 0.325, 0.448, -0.123 *****
- **Manager is equal to one of the owners**: 0.625, 0.578, 0.783, -0.205 *****
- **Age of the manager**: 42.961, 41.145, 49.063, -7.918 *****
- **Years of Schooling of the manager**: 5.203, 5.796, 3.212, 2.583 *****
- **Non-Managerial Household Labor (hours/ha)**: 530.6, 576.3, 377.2, 199.1 *****
- **Ln Non-Managerial Household Labor (hours/ha)**: 5.313, 5.766, 3.789, 1.977 *****
- **Ln Exchange labor (days/ha)**: 0.211, 0.180, 0.316, -0.136 *****
- **Household Size**: 4.880, 5.148, 3.982, 1.166 *****
- **Dependency Ratio**: 0.704, 0.701, 0.713, -0.012
- **Ag extension services receipt**: 0.311, 0.323, 0.271, 0.052 *****
- **Head of household works off-farm**: 0.378, 0.402, 0.296, 0.106 *****
- **HH has male adult that works off-farm (no manager)**: 0.046, 0.031, 0.094, -0.063 *****
- **HH receives other transfers/safety net help**: 0.215, 0.211, 0.227, -0.016 *
- **Wealth Index**: -0.707, -0.612, -1.025, 0.413 *****
- **HH Distance (KMs) to Nearest ADMARC**: 8.203, 8.201, 8.210, -0.009

Note: ** *** p<0.01, ** p<0.05, * p<0.1
Table 2: Base OLS Regression Results Underlying the Mean Decomposition  
*Dependent Variable: Ln[Plot Gross Value of Output per hectare]*

<table>
<thead>
<tr>
<th></th>
<th>Male Managed Plot Sample</th>
<th>Female Managed Plot Sample</th>
<th>Difference in Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Labor</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ln Managerial Labor (hours/ha)</td>
<td>0.229*** (0.011)</td>
<td>0.236*** (0.018)</td>
<td></td>
</tr>
<tr>
<td><strong>Purchased Inputs</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pesticide/herbicide use yes/no</td>
<td>0.442*** (0.054)</td>
<td>0.586*** (0.123)</td>
<td></td>
</tr>
<tr>
<td>Organic fertilizer use yes/no</td>
<td>0.021 (0.024)</td>
<td>0.034 (0.044)</td>
<td></td>
</tr>
<tr>
<td>Ln Inorganic Fertilizer (kg/ha)</td>
<td>0.070*** (0.003)</td>
<td>0.058*** (0.006)</td>
<td>***</td>
</tr>
<tr>
<td>Ln Hired labor (days/ha)</td>
<td>0.102*** (0.007)</td>
<td>0.113*** (0.012)</td>
<td></td>
</tr>
<tr>
<td>Agricultural Implements Access Index</td>
<td>0.036*** (0.007)</td>
<td>0.048*** (0.012)</td>
<td></td>
</tr>
<tr>
<td>Proportion of Plot Area Under Improved Seeds</td>
<td>0.040** (0.019)</td>
<td>0.020 (0.034)</td>
<td></td>
</tr>
<tr>
<td>Proportion of Plot Area Under Export Crops</td>
<td>1.076*** (0.030)</td>
<td>1.131*** (0.086)</td>
<td></td>
</tr>
<tr>
<td><strong>Household Characteristics and Endowment</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ln GPS Total Area of the plot (ha)</td>
<td>-0.164*** (0.040)</td>
<td>-0.345*** (0.074)</td>
<td>**</td>
</tr>
<tr>
<td>Ln GPS Total Area of the plot (ha) Squared</td>
<td>0.052*** (0.014)</td>
<td>0.008 (0.025)</td>
<td>*</td>
</tr>
<tr>
<td>Elevation (m)</td>
<td>0.000** (0.000)</td>
<td>0.000** (0.000)</td>
<td></td>
</tr>
<tr>
<td>Plot distance to household</td>
<td>-0.000 (0.001)</td>
<td>-0.000 (0.002)</td>
<td></td>
</tr>
<tr>
<td>Intercropped</td>
<td>0.230*** (0.022)</td>
<td>0.269*** (0.037)</td>
<td></td>
</tr>
<tr>
<td>Manager is equal to one of the owners</td>
<td>-0.007 (0.016)</td>
<td>-0.003 (0.033)</td>
<td></td>
</tr>
<tr>
<td>Age of the manager</td>
<td>-0.001** (0.001)</td>
<td>-0.000 (0.001)</td>
<td></td>
</tr>
<tr>
<td>Years of Schooling of the manager</td>
<td>0.003 (0.002)</td>
<td>0.011** (0.005)</td>
<td></td>
</tr>
<tr>
<td>Ln Non-Managerial Household Labor (hours/ha)</td>
<td>0.015** (0.006)</td>
<td>0.013** (0.005)</td>
<td></td>
</tr>
<tr>
<td>Ln Exchange labor (days/ha)</td>
<td>0.041*** (0.012)</td>
<td>0.013 (0.015)</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>Male Managed Plot Sample</td>
<td>Female Managed Plot Sample</td>
<td>Difference in Coefficients</td>
</tr>
<tr>
<td>--------------------------</td>
<td>--------------------------</td>
<td>----------------------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>Household Size</td>
<td>0.011***</td>
<td>0.021***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.008)</td>
<td></td>
</tr>
<tr>
<td>Dependency Ratio</td>
<td>-0.005</td>
<td>-0.066***</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>(0.016)</td>
<td>(0.018)</td>
<td></td>
</tr>
<tr>
<td>Agricultural Extension Receipt</td>
<td>0.031*</td>
<td>0.134***</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>(0.017)</td>
<td>(0.031)</td>
<td></td>
</tr>
<tr>
<td>Head of HH works off-farm</td>
<td>-0.058***</td>
<td>-0.017</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.016)</td>
<td>(0.031)</td>
<td></td>
</tr>
<tr>
<td>HH has male adult that works off-farm (different from manager)</td>
<td>-0.037</td>
<td>0.022</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.044)</td>
<td>(0.047)</td>
<td></td>
</tr>
<tr>
<td>HH receives other transfers/safety net help</td>
<td>0.009</td>
<td>-0.032</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.020)</td>
<td>(0.034)</td>
<td></td>
</tr>
<tr>
<td>Wealth Index</td>
<td>0.064***</td>
<td>0.059***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.009)</td>
<td></td>
</tr>
<tr>
<td>Distance to Nearest ADMARC (KM)</td>
<td>0.004**</td>
<td>-0.000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.003)</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>7.946***</td>
<td>8.321***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.261)</td>
<td>(0.757)</td>
<td></td>
</tr>
<tr>
<td>Number of observations</td>
<td>10,822</td>
<td>3,222</td>
<td></td>
</tr>
<tr>
<td>R2</td>
<td>0.379</td>
<td>0.373</td>
<td></td>
</tr>
<tr>
<td>Adjusted R2</td>
<td>0.376</td>
<td>0.361</td>
<td></td>
</tr>
</tbody>
</table>

Robust standard errors in parenthesis.

note: *** p<0.01, ** p<0.05, * p<0.1
Table 3. Marginal Productivity of Land and Labor

<table>
<thead>
<tr>
<th>Marginal Productivity of Land in Agricultural Production</th>
<th>Male Headed Households</th>
<th>Female Headed Households</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>113,111***</td>
<td>82,454***</td>
<td>30,656***</td>
</tr>
<tr>
<td></td>
<td>(2,254)</td>
<td>(3,158)</td>
<td>(3,880)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Marginal Productivity of Labor in Agricultural Production</th>
<th>Male Headed Households</th>
<th>Female Headed Households</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>28.6***</td>
<td>17.9***</td>
<td>10.6***</td>
</tr>
<tr>
<td></td>
<td>(1.39)</td>
<td>(1.41)</td>
<td>(1.98)</td>
</tr>
</tbody>
</table>

Note: *** p<0.01, ** p<0.05, * p<0.1
Table 4. Mean Decomposition of the Gender Gap in Agricultural Land Productivity

|                          | A. Mean Gender Gap |                    | B. Decomposition of the Mean Gender Gap+
|--------------------------|--------------------|--------------------|---------------------------------------------
|                          | Male Plots         |                    | Labor Market Effect                       |
|                          | 10.474***          | (0.009)            | -0.046***                                  |
|                          | Female Plots       | 10.229***          | Purchased Inputs Effect                    |
|                          |                    | (0.016)            | 0.128***                                   |
|                          | Gender Gap         | 0.245***           | Household Endowment Effect                 |
|                          |                    | (0.019)            | 0.088***                                   |
|                          |                    |                    | Pure Marginal Productivity Effect          |
|                          |                    |                    | 0.075***                                   |
|                          |                    |                    |                                             |

Note: *** p<0.01, ** p<0.05, * p<0.1,
+ all the effects sum up to the Mean Gender Gap
Table 5. Mean Decomposition of the Gender Gap in Agricultural Labor Productivity

<table>
<thead>
<tr>
<th>A. Mean Gender Gap</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male Plots</td>
</tr>
<tr>
<td>4.662***</td>
</tr>
<tr>
<td>(0.011)</td>
</tr>
<tr>
<td>Female Plots</td>
</tr>
<tr>
<td>4.221***</td>
</tr>
<tr>
<td>(0.019)</td>
</tr>
<tr>
<td>Mean Gender Gap</td>
</tr>
<tr>
<td>0.441***</td>
</tr>
<tr>
<td>(0.022)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B. Decomposition of the Mean Gender Gap+</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labor Market Effect</td>
</tr>
<tr>
<td>0.150***</td>
</tr>
<tr>
<td>(0.014)</td>
</tr>
<tr>
<td>Purchased Inputs Effect</td>
</tr>
<tr>
<td>0.128***</td>
</tr>
<tr>
<td>(0.012)</td>
</tr>
<tr>
<td>Household Endowment Effect</td>
</tr>
<tr>
<td>0.088***</td>
</tr>
<tr>
<td>(0.019)</td>
</tr>
<tr>
<td>Pure Marginal Productivity Effect</td>
</tr>
<tr>
<td>0.075***</td>
</tr>
<tr>
<td>(0.024)</td>
</tr>
</tbody>
</table>

Note: *** p<0.01, ** p<0.05, * p<0.1,
+ all the effects sum up to the Mean Gender Gap
Figure 1. Kernel Density Estimates of the Log of Gross Value of Output per Hectare for Male- and Female-Managed Plot Samples
Figure 2. Kernel Density Estimates of the Log of Gross Value of Output per Managerial Labor for Male- and Female-Managed Plot Samples
ELECTRONIC APPENDIX

APPENDIX A
GENDER GAP IN LAND PRODUCTIVITY

The *gender gap* “$D$” is expressed as the mean outcome difference:

$$D = E(Y^M) - E(Y^F)$$

(A.1)

Equations (12a) and (12b) imply that:

$$E(Y^M) = \beta_0^M + E(L^M)\beta_L^M + \sum_k E(X_k^M)\beta_k^M + \sum_j E(Z_j^M)\delta_j^M$$
\hspace{5.3cm} (A.2a)

$$E(Y^F) = \beta_0^F + E(L^F)\beta_L^F + \sum_k E(X_k^F)\beta_k^F + \sum_j E(Z_j^F)\delta_j^F$$
\hspace{5.3cm} (A.2b)

We rewrite equation (13) using equations (12a) and (12b):

$$D = E(Y^M) - E(Y^F) = \beta_0^M + E(L^M)\beta_L^M + \sum_k E(X_k^M)\beta_k^M + \sum_j E(Z_j^M)\delta_j^M -$$

$$\beta_0^F - E(L^F)\beta_L^F - \sum_k E(X_k^F)\beta_k^F - \sum_j E(Z_j^F)\delta_j^F$$
\hspace{5.3cm} (A.3)

Rearranging Equation (15) by adding and subtracting $E(L^M)\beta_L^F$, $\sum_k E(X_k^M)\beta_k^F$, $\sum_j E(Z_j^M)\delta_j^F$, we decompose the gender gap into the following components:

$$D = \underbrace{[E(L^M) - E(L^F)](\beta_L^F)}_{\text{labor market effect}} + \underbrace{\sum_k [E(X_k^M) - E(X_k^F)](\beta_k^F)}_{\text{purchased inputs effect}} + \underbrace{\sum_j [E(Z_j^M) - E(Z_j^F)]\delta_j^F}_{\text{household endowment effect}} +$$

$$\underbrace{E(L^M)[\beta_L^M - (\beta_L^F)]}_{\text{pure marginal productivity effect}} + \sum_k E(X_k^M)[\beta_k^M - \beta_k^F] + (\beta_0^M - \beta_0^F) + \sum_j E(Z_j^M)[\delta_j^M - \delta_j^F]$$
\hspace{5.3cm} (A.4)
where $\beta_0, \beta_l, \beta_k, \delta_i$ are the estimated intercept and slope coefficients of each covariate included in the regressions for the male and female plot samples.

Equation (16) is the aggregate decomposition. The first component is the labor market effect, i.e. the portion of the gender gap driven by differences in quantities of labor allocated to farm activities by the head of household. The second component is the purchased inputs effect, the portion of the gender gap that is explained by differences in levels of use of inputs that have to be bought such as fertilizer, pesticides, seeds, agricultural implements, and/or hired labor. The third component, the household endowment effect is comprised by differences in levels of observable characteristics of the household, including human and physical capital. The fourth component is the pure marginal productivity effect and corresponds to the portion of the gender gap explained by differences in the coefficients of each observable covariate included in L and in the $X$ and $Z$ vectors; as well as differences in the constant between male and female headed households.
Figure B1. Land and Labor Productivity Gender Gap Decomposition into Endowment, Purchased Inputs, Labor Market and Pure Marginal Productivity Effects

The graphical representation of the gender gap in land and labor productivity and their respective components is presented in Figure B1 below. It presents the value of output per hectare \((Y=y/ha)\) at each level of managerial labor hours per hectare \((L=l/ha)\). Curve \(Y_M^{M} (\text{Male Endowment})\) and curve \(Y_F^{F} (\text{Female Endowment})\) are derived from the male and female
regressions respectively. Curve \( Y^F \) derived using the coefficients from the female regression and the male average value of the variables included in the purchased inputs vector; this curve represents the level of agricultural productivity that a female-headed household would have if it had the same level of purchased inputs as a male-headed household, but the female level of endowment. Similarly, curve \( Y^F \) derived using the coefficients of the female regressions and the male average level of the variables included in the endowment and purchased inputs vectors. \( L^M \) and \( L^F \) are the observed average levels of managerial labor for the male and female samples respectively.

Point A is the observed average male land productivity and point D is the observed average female land productivity. Point E is the productivity that women would get if they would work the same number of hours as men. Point C is the land productivity that women would get if they had only the same level of purchased inputs as men, but the female endowment and would work the same number of hours as men, while point B is the land productivity that women would attain if they had access to the same level of endowment and purchased inputs and worked the same number of hours as men.

The observed gender gap in land productivity is equal to the distance between \( Y^A \) and \( Y^D \) and can be decomposed into the four components presented in equation (A.4): the pure marginal productivity effect is the distance \( Y^A - Y^B \), the endowment effect \( Y^B - Y^C \), the purchased inputs effect \( Y^C - Y^E \) and the labor effect \( Y^E - Y^D \). In the context of land productivity, the fact that women work more hours in average than their male counterparts has the effect of reducing the gender gap, counteracting the purchased inputs, endowment and pure marginal productivity effects.
The average labor productivity is obtained by dividing $Y$ over $L \left( \frac{Y}{L} \right)$ which is equal to the slope of a straight line from the origin to any point on the land productivity curve.

The observed gender gap in labor productivity is equal to $D_L = \frac{Y^A}{L^M} - \frac{Y^D}{L^F}$ which is decomposed into the four components of equation (15): the pure marginal productivity effect is equal to $\frac{Y^A}{L^M} - \frac{Y^B}{L^M}$, the endowment effect is $\frac{Y^B}{L^M} - \frac{Y^C}{L^M}$, the purchased inputs effect $\frac{Y^C}{L^M} - \frac{Y^E}{L^M}$ and the labor effect is $\frac{Y^E}{L^M} - \frac{Y^D}{L^F}$. For comparison purposes, it is important to note that the average productivity in D is equal to the average productivity in D*. The labor effect is positive unlike the labor effect in the decomposition of the land productivity. In the case of labor productivity, the labor effect amplifies the gender gap.

This is consistent with the theoretical model which predicts that the gender gap in labor productivity will be larger than the gender gap in land productivity given that the labor market imperfections spillovers increase the allocation of labor to the farm. As discussed earlier, this increases production in an inefficient way. Figure B1 illustrates this prediction.

**APPENDIX C**

Table C1: Variable Definitions

5
<table>
<thead>
<tr>
<th>Variable</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Agricultural Land Productivity</strong></td>
<td>The agricultural productivity variable is proxied by the plot level gross value of output in Malawi Kwacha (MK) calculated by first multiplying the kilogram-equivalent quantity of production for each crop on a given plot by the median crop sales value per kilogram within the corresponding enumeration area (EA), and then aggregating across values of crop production. The median crop sales value per kilogram is computed within the corresponding EA only if at least 10 values are available from the survey data. Otherwise, the median crop sales value per kilogram is computed at a higher level, in the order of traditional authority, district, region, and country. Our outcome variable is computed by normalizing plot-level gross value of output with the area of the plot (land productivity).</td>
</tr>
<tr>
<td><strong>Agricultural Labor Productivity</strong></td>
<td>The measure of labor productivity is calculated by normalizing the plot-level gross value of output with the total number of hours of managerial labor on the plot.</td>
</tr>
<tr>
<td><strong>Plot</strong></td>
<td>A plot was defined as a continuous piece of land on which a unique crop or a mixture of crops is grown, under a uniform, consistent crop management system, not split by a path of more than one meter in width. Plot boundaries were defined in accordance with the crops grown and the operator.</td>
</tr>
<tr>
<td><strong>Plot manager</strong></td>
<td>For each plot, the following question was asked to identify the primary decision maker/manager: “Who in this household makes the decisions concerning crops to be planted, input use and the timing of cropping activities on this plot?” The questionnaire allowed for identification of one manager per plot, on whom individual-level information could be recovered from the Household Questionnaire.</td>
</tr>
<tr>
<td><strong>Plot Owner</strong></td>
<td>For each plot, the following question was asked to identify the plot owners: “Who owns this plot?” The question allowed up to 2 household members to be specified as owners</td>
</tr>
<tr>
<td><strong>Plot Labor Input</strong></td>
<td>The plot-level measures of household labor input are the summations of rainy season labor hours across household members reported to have worked on a given plot. Individual labor input is computed as the multiplication of the number of weeks a household member worked on a given plot during the reference rainy season, the typical number of days worked per week during the reported number of weeks, and the typical number of hours worked per day during the reported number of weeks. The plot-level measure of hired labor (exchange) input is the sum of aggregate men, women, and child hired (exchange) labor days.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Household Wealth Index</strong></td>
<td>The household wealth index is constructed using principal component analysis, and takes into account the number of rooms</td>
</tr>
</tbody>
</table>
in the dwelling, a set of dummy variables accounting for the ownership of (i) dwelling, (ii) mortar, (iii) bed, (iv) table, (v) chair, (vi) fan, (vii) radio, (viii) tape/CD player, (ix) TV/VCR, (x) sewing machine, (xi) paraffin/ kerosene/ electric/ gas stove, (xii) refrigerator, (xiii) bicycle, (xiv) car/motorcycle/minibus/lorry, (xv) beer brewing drum, (xvi) sofa, (xvii) coffee table, (xviii) cupboard, (xix) lantern, (xx) clock, (xxi) iron, (xxii) computer, (xxiii) fixed phone line, (xxiv) cell phone, (xxv) satellite dish, (xxvi) air-conditioner, (xxvii) washing machine, (xxviii) generator, (xxix) solar panel, (xx) desk, and a vector of dummy variables capturing access to improved (i) outer walls, (ii) roof, (iii) floor, (iv) toilet, and (v) water source. The household agricultural implement access index is also computed using principal components analysis, and covers a range of dummy variables on the ownership of (i) hand hoe, (ii) slasher, (iii) axe, (iv) sprayer, (v) panga knife, (vi) sickle, (vii) treadle pump, (viii) watering can, (ix) ox cart, (x) ox plough, (xi) tractor, (xii) tractor plough, (xiii) ridger, (xiv) cultivator, (xv) generator, (xvi) motorized pump, (xvii) grain mail, (xviii) chicken house, (xix) livestock kraal, (xx) poultry kraal, (xxi) storage house, (xxii) granary, (xxiii) barn, and (xxiv) pig sty.
APPENDIX D

SENSITIVITY ANALYSES

We are concerned about two main issues: (i) the validity of the decomposition methodology employed and (ii) whether the gender differences of interest – labor, purchased inputs, endowment including child dependency ratio, plot and household locations, - are robust to various specifications. As noted earlier, the crucial assumptions for the validity of the aggregate decomposition include overlapping support and ignorability. The key assumptions additionally required by the detailed decomposition are additive linearity and zero conditional mean. A methodology that is proposed by Imbens and Rubin (2009) to assess the feasibility of the overlapping support assumption is centered on the idea of calculating a scale-free normalized difference for each covariate. They assert that the overlapping support across the groups of interest, in our case female vs. male plots, is adequate if the scale-free normalized differences across most of the covariates are less than 0.25. Table D.1 in the Appendix presents the scale-free normalized difference of the variables used in the regressions. Only 5 out of 26 independent variables (and 30 district dummies) have a normalized difference greater than 0.25.

In trying to lend support to ignorability and zero conditional mean assumptions, we use all available data and econometric tools at our disposal, and first rely on an empirical approach that was pioneered by Altonji (1988), Murphy and Topel (1990), and Altonji et al. (2005), based on the idea that the amount of selection on observable variables provides a guide to the extent of selection on unobservable counterparts. We use an informal version of the methodology applied by Acemoglu et al. (2001), and incorporate into our base specification, in a phased-in fashion, thematically-grouped control variables such that each regression is estimated with a different set
of additional independent variables and that the results are compared to those from the base specification. Our purpose is to gauge the stability of the key regression coefficients that underlie our decomposition results. If the coefficients on the covariates included in the base specification are stable subsequent to the incorporation of additional covariates, they are less likely to change if we are able to take into account potentially missing omitted variables.

There are certain factors that may affect the gender gap in agricultural productivity that we may not have fully accounted for in our base estimations. Various village specific policies, inputs, and quality of infrastructure may influence agricultural productivity and thus the gender gap in agricultural productivity. The distribution of subsidies may be decided at the village level by the village head, and this distribution may have gender-specific productivity effects. Geospatial characteristics may play an important role as well, as male-managed plots may have different characteristics than female-managed plots. Finally, it is possible that the composition of the household that we proxy for is not detailed enough.

To account for the factors described above, we consider the following sets of variables: (i) enumeration area fixed effects, (ii) plot geospatial characteristics informed by GIS data, (iii) other plot characteristics solicited by the IHS3 (farmer’s perception of plot characteristics), (iv) additional household characteristics such as disaggregated household composition controls, and (v) additional community characteristics such as number of fertilizer sellers, hyrbird maize sellers, distance to microfinance institutions, credit cooperatives, and extension services. An overwhelming majority of the coefficients, with respect to the base specification, are stable across the specifications and the plot samples, and do not change sign or significance. This suggests that
the assumptions of ignorability and zero conditional mean might not be unfounded. These results are available on request from the authors.

In addition, we checked for extreme data points that may dominate the sign and significance of key estimates. We conducted three types of dominance tests. In order to account for extreme data points and different samples, we first re-estimated the model by excluding observations in the top and bottom 1 percent of the land productivity. The same procedure is followed by re-estimating the model without observations in the top and bottom 1 percent of managerial labor. In the last specification we reduce the sample to those plots in which the manager has contributed with at least 50% of the household labor. The parameters are robust to the sample changes. These results are available on request from the authors.
Table D.1 Scale-Free Normalized Difference

<table>
<thead>
<tr>
<th>Variable</th>
<th>Male Sample</th>
<th>Female Sample</th>
<th>Difference</th>
<th>Normalized Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ln Managerial Labor (hours/ha)</td>
<td>5.81</td>
<td>6.01</td>
<td>-0.20</td>
<td>-0.16</td>
</tr>
<tr>
<td>Pesticide/herbicide use yes/no</td>
<td>0.02</td>
<td>0.01</td>
<td>0.01</td>
<td>0.06</td>
</tr>
<tr>
<td>Organic fertilizer use yes/no</td>
<td>0.11</td>
<td>0.1</td>
<td>0.01</td>
<td>0.02</td>
</tr>
<tr>
<td>Ln Inorganic Fertilizer (kg/ha)</td>
<td>3.32</td>
<td>3.13</td>
<td>0.19</td>
<td>0.05</td>
</tr>
<tr>
<td>Ln Hired labor (days/ha)</td>
<td>0.55</td>
<td>0.52</td>
<td>0.03</td>
<td>0.02</td>
</tr>
<tr>
<td>Agricultural implements Asset Index</td>
<td>0.84</td>
<td>0.16</td>
<td>0.68</td>
<td>0.38</td>
</tr>
<tr>
<td>Proportion of area of the plot under improved varieties</td>
<td>0.38</td>
<td>0.33</td>
<td>0.05</td>
<td>0.08</td>
</tr>
<tr>
<td>Propportion of area of the plot under export crops</td>
<td>0.09</td>
<td>0.03</td>
<td>0.06</td>
<td>0.18</td>
</tr>
<tr>
<td>Ln GPS Total Area of the plot (ha)</td>
<td>-1.2</td>
<td>-1.29</td>
<td>0.09</td>
<td>0.09</td>
</tr>
<tr>
<td>Ln GPS Total Area of the plot (ha) Squared</td>
<td>1.93</td>
<td>2.15</td>
<td>-0.22</td>
<td>-0.08</td>
</tr>
<tr>
<td>Elevation (m)</td>
<td>909</td>
<td>843</td>
<td>65.32</td>
<td>0.14</td>
</tr>
<tr>
<td>Plot distance to hh</td>
<td>2.06</td>
<td>1.68</td>
<td>0.38</td>
<td>0.03</td>
</tr>
<tr>
<td>Inter-cropped</td>
<td>0.32</td>
<td>0.45</td>
<td>-0.13</td>
<td>-0.19</td>
</tr>
<tr>
<td>Manager is equal to one of the owners</td>
<td>0.58</td>
<td>0.78</td>
<td>-0.20</td>
<td>-0.31</td>
</tr>
<tr>
<td>Age of the manager</td>
<td>41.18</td>
<td>49.07</td>
<td>-7.89</td>
<td>-0.34</td>
</tr>
<tr>
<td>Years of Schooling of the manager</td>
<td>5.8</td>
<td>3.21</td>
<td>2.59</td>
<td>0.50</td>
</tr>
<tr>
<td>Ln Non-Managerial Household Labor (hours/ha)</td>
<td>5.77</td>
<td>3.8</td>
<td>1.97</td>
<td>0.60</td>
</tr>
<tr>
<td>Ln Exchange labor (days/ha)</td>
<td>0.18</td>
<td>0.32</td>
<td>-0.14</td>
<td>-0.13</td>
</tr>
<tr>
<td>Household Size</td>
<td>5.16</td>
<td>4</td>
<td>1.16</td>
<td>0.38</td>
</tr>
<tr>
<td>Dependency Ratio</td>
<td>0.7</td>
<td>0.71</td>
<td>-0.01</td>
<td>-0.01</td>
</tr>
<tr>
<td>Ag extension services receipt</td>
<td>0.32</td>
<td>0.27</td>
<td>0.05</td>
<td>0.08</td>
</tr>
<tr>
<td>HH has any off-farm income</td>
<td>0.44</td>
<td>0.35</td>
<td>0.09</td>
<td>0.13</td>
</tr>
<tr>
<td>HH receives other transfers/safety net help</td>
<td>0.21</td>
<td>0.23</td>
<td>-0.02</td>
<td>-0.03</td>
</tr>
<tr>
<td>Wealth Index</td>
<td>-0.6</td>
<td>-1.03</td>
<td>0.43</td>
<td>0.16</td>
</tr>
<tr>
<td>HH Distance (KMs) to Nearest ADMARC</td>
<td>8.19</td>
<td>8.2</td>
<td>-0.01</td>
<td>0.00</td>
</tr>
</tbody>
</table>