Farm Size and Productivity -The Role of Family Labor

Muhammad Ayaz, Mazhar Mughal

Farm Size and Productivity - The Role of Family Labor

Muhammad Ayaz* and Mazhar Mughal†

May 16, 2022

Abstract

In this study, we draw a theoretical model to demonstrate that small farms achieve lower total factor productivity (TFP) compared to large farms, even though their yield may be higher. We argue that taking into account family labor modifies the farm size-productivity relationship. We test our hypotheses on geocoded data from 5,645 agriculture farms in Pakistan using Pakistan household integrated economic survey 2018-19 and labor force survey 2018 combined with remote sensing data to account for farm-specific topographic features. We base our analysis on OLS and stochastic frontier analysis. We find that family labor is the key to understanding the nature and strength of the farm size – productivity relationship. Farm size's association, both with yield and TFP, turns positive when we measure family labor in terms of market wage rate rather than marginal product. Farm yield decreases by -0.07% with a one percent increase in farm size but gets insignificant or increases by 0.034% when family labor cost is measured at market wages rather than the marginal product. We find that higher family labor intensity, labor market distortion due to the notion of family dishonor, and suboptimal crop selection by small farms play a crucial role in this context.

Keywords: Farm size, productivity, family labor, TFP, family dishonor, technical efficiency, Pakistan.

JEL codes: Q12, Q13, J3,J4,D13

*UMR TREE, E2S UPPA, University of Pau, France, and University of Balochistan, Quetta, Pakistan. Email: m.ayaz@univ-pau.fr

†ESC Pau Business School, Pau, France, and UMR TREE, E2S UPPA, University of Pau, France. Email: mazhar.mughal@esc-pau.fr

1 Introduction

Agricultural productivity plays a crucial role in maintaining food price stability, reducing rural poverty and providing food security in the developing world. A major challenge to enhancing agricultural productivity in the developing countries relates to farm size. Around 60 - 70% of farmland in low- and lower-middle income countries belong to smallholders (Lowder et al., 2016). In South Asia and Sub-Saharan Africa, where 80.6% of the world's poor live [1], 80% of the farms are small, i.e. less than two hectares (Lowder et al., 2016). A large body of literature has examined the role of farm size in determining productivity, starting with the study of Chayanov (1926) who observed an inverse farm size - productivity relationship in Russia. Studies from developing countries, such as Sen (1962). Bardhan (1973), Berry & Cline (1980), Bhalla (1988b), Assunção & Braido (2007), Michael Lipton (2009),Larson et al. (2014), Ali & Deininger (2015), Kagin et al. (2015), and Gautam & Ahmed (2018) found that small farms are more productive and efficient than large farms. In the developing countries, where land and capital are scarce while labor is surplus, small farms inevitably have an advantage over larger ones (Lipton, 2005). For example, Julien et al. (2019) analyze survey data from three east African countries and find that the total factor productivity (TFP) of small farms is higher than large farms.

However, there is no consensus on whether the inverse farm size - farm productivity relationship (hence forth referred to as IR) is real, a statistical artifact arising from data misreporting, or a result of factor market imperfections. The direction of the farm size – productivity relationship has important policy implications regarding land distribution and agriculture factor market reforms. An inverse farm size and agricultural productivity relationship would suggest the need for land to be parceled into smaller farms in order to make land distribution more efficient (Lipton, 2005), whereas an IR owing to factor market imperfections would underscore the need for input market reforms.

The literature discusses three main explanations for the IR: (1) omitted variable bias due to factors such as land quality; (2) measurement errors in the measures of farm-size, output or farm productivity, and (3) factor market imperfections, specifically, labor market distortions.

First, the average land quality of small farms is better than large farms (Bhalla, 1988a; Bhalla & Roy, 1988; Sen, 1975). Small holders tend to sellout their poor-quality land when they are in need of cash. As the farm size decreases due to selling out of poor-quality land, the small farmland's relative quality improves (Bhagwati & Chakravarty, 1969; Bhalla, 1988b). Controlling for land quality can therefore eliminate the IR (Lamb, 2003; Benjamin, 1995; Bhalla, 1988b). Conversely, recent studies find that soil quality does not explain the IR (Bevis & Barrett, 2020; Gourlay et al., 2019).

The second explanation for the IR relates to errors in measuring output, land size, and farm

[1] The statistics can be retrieve from world bank site using the following address: https://blogs.worldbank.org/opendata/updated-estimates-impact-covid-19-global-poverty-looking-back-2020-and-outlook-2021

productivity. Using data from Ethiopia, Abay et al. (2019) and Desiere & Jolliffe (2018) find that the IR disappears if the farms' output is measured by the quantity of the crops cut rather than the farmers' self-reported production. Similar results are found by Bevis & Barrett (2020) in case of Uganda. The choice of the measure of farm productivity also plays a significant role in explaining the IR. The negative relationship between farm size and productivity turns positive if agricultural productivity is measured in terms of TFP instead of crop yield(Helfand & Taylor, 2020; Sanchez et al., 2020). Nevertheless, merely correcting for production bias (using the crops cut instead of farmer self-reported yield) does not explain the IR, unless land size (measured using objective estimates of farm size instead of farmer self-reporting) is also taken into account (Abay et al., 2019).

The third explanation for the inverse relationship between farm size and productivity pertains to labor market imperfections. Small farms may obtain higher yields thanks to intensive use of inexpensive family labor, which creates agricultural labor market dualism (Bhalla, 1979, 1988b). Agriculture in the developing countries is characterized by surplus farm labor and relatively-expensive capital. Small farms optimize their yield by using surplus family labor with low opportunity cost. In contrast, large farms are more capital-intensive and employ hired labor with greater explicit costs. This provides a cost advantage to small farms over large farms (Sen, 1962; Bhalla, 1979). Sen (1962) argue that prima facie, agriculture in developing countries operates in two distinct wage regimes: the non-wage-based family labor and the wage-based hired labor. Small farms generally operate on a non-wage family labor basis, while large farms mainly depend on wage-based hired labor (Feder, 1985). Small farms employ higher labor per acre in production (Bhalla, 1988b). This is because the opportunity cost of surplus family labor is less than that of hired labor in the absence of sufficient work opportunities. In a surplus-labor economy, the opportunity cost of on-farm family labor is not equal to the market wage but rather to the market wage multiplied by the probability of getting another job in the market, which is invariably less than one (Mazumdar, 1965). Hence, the per acre labor intensity and the corresponding per acre output of small farms depending on non-wage family labor is higher than that of large farms (Bhalla, 1988b). Accounting for this labor market imperfection in agriculture can eliminate the inverse farm size - productivity relationship (Foster & Rosenzweig, 2021; Lamb, 2003). Here, it must be mentioned that the comparative advantage of small farms due to the availability of surplus family labor does not guarantee that they would be more efficient than large farms. An expanding industrial sector absorbs more and more of surplus labor looking for non-farm activities, with the result that the relative production efficiency advantage enjoyed by the small farms over large farms wanes over time (Deininger et al., 2018; Gautam & Ahmed, 2018) and eventually reverses (Otsuka et al., 2013, 2016; Rada & Fuglie, 2019). The productivity of family labor relative to hired labor also decreases over time, eliminating small farms' comparative advantage. In India, the allocative and technical efficiency of large farms is found to be higher than that of small farms (Jha et al., 2000). Small farms employ higher inputs per unit of land, and therefore produce less output at given inputs (Carter, 1984)(p.141). Similarly, farm mechanization and better protection from adverse income shocks make large farms in India more profitable and efficient than small farms (Foster & Rosenzweig, 2021).

Recent studies show that IR persists even after controlling for household fixed effects (Gourlay et al., 2019), suggesting that the relationship exists at the plot rather than the farm level (Bevis & Barrett, 2020). However, studies such as Gourlay et al. (2019) and Larson et al. (2014) found that the relationship exist at both the plot and the farm level.

In this study, we propose a theoretical model and provide empirical evidence supporting the importance of family labor valuation and imputation. We examine how family-labor valuation at the marginal product of family labor and market wage rate can be crucial in determining the strength and direction of the association between farm size and productivity. We show how various farm productivity measures, such as yield or TFP, as well as overall or technical efficiency, can affect the IR with respect to family labor cost evaluated at marginal product of labor (MPL) and market wage rate (MWR). We find that the total factor productivity of small farms is lower than that of large farms, even though the yield of the small farms is higher if family labor is accounted for in terms of marginal product. Besides, the IR disappears if family labor is measured in terms of market wage instead of marginal product of family labor.

We propose the cultural norms related to family honor and status as a possible explanation for the observed farm size - productivity relationship. Self-employed family farm workers are often unwilling to offer their services as paid labor to the away-farms [2] because of the notion of honor and social status (Bardhan, 1973)(p.1380), and prefer to work on their family-owned farm even they may obtain higher wages on away farms or from off-farm activities. This notion of honor prevails in tribal or caste-based cultures with values of loyalty to the family, caste, and tribe (see e.g., Lyon (2012, 2005). In Pakistan, the ethnographies of two provinces, Khyber-Pakhtunkhwa (KPK) and Balochistan, are characterized by tribal affiliations and practices and discourses that overtly assert idealized notions of honor. The culture of other two provinces, Punjab and Sindh, is based on caste ('zaat') (Lyon, 2012). The proverbs such as "Give your life; take honor in return" and "Honor before bread" highlight the importance and endurance of honor in these cultural contexts. The 'Izzat' (honor) of one family member affects the family's honor in common (Lyon, 2005). The participation of men and women in agriculture other activities is a matter of honor (Izzat) and not only that of material benefits (Drucza & Peveri, 2018). Working for someone else implies subordinating and compromising one's family honor (Bardhan, 1973)(p.1380). In such a case, a farm worker does not compare his/her earnings from the family farm with the gross wage earned at the away-farm but with the potentially much lower net income that he/she brings home (Morten, 2019). As a result, even if family labor gets a significantly lower remuneration from the self-owned farm, the family laborers are hesitant to work on away-farms. Consequently, small farms become more dependent on family labor than large farms. Since self-employed labor costs are not properly

[2] Any agricultural farm/facility not owned or operated by the family.

accounted for by farmers, the IR could be attributed to small farms' intensive use of undervalued family labor.

We also examine various measurement issues related to the valuation and aggregation of farm inputs and outputs that have not simultaneously been addressed in the literature. Crop differentials, self-reported output price bias, and soil quality measures across farms are among the issues addressed.

We analyze representative geocoded data from 5,645 agriculture farms in Pakistan. For this purpose, we use Pakistan household integrated economic survey 2018-19 and labor force survey 2018 combined with the remote sensing data of SRTM version-1 at 3-arc-second, Aster GDEM, GLDAS-2.1, and HWSD to account for farm-specific factors including land quality, elevation, roughness and soil moisture. We employ OLS and Stochastic Frontier Analysis (SFA) techniques to analyze farms' overall and technical efficiency with respect to their size.

2 Theoretical Model

2.1 Economy

Consider an economy based on a single sector, namely agriculture. The economy is characterized by heterogeneous farm-households i. Each household is endowed with labor $T_i^* > 0$ and land ${N_i}^* \geq 0$. Labor endowment is similar across households. In other words, family size is identical across farms of different sizes. Land endowment ${N_i}^*$, on the other hand, varies across households. Each household maximizes its utility which depends positively on the output it produces and negatively on the labor it supplies to owned- ${Lf_i}^*$ and away-farms ${La_i}^*$.

2.2 Technology

Farm productivity varies according to the farmer's entrepreneurial skills and abilities $s_i \in (1, ..., S)$. Apart from the farmer's entrepreneurial skills, farms' productivity $\Gamma_i \in (1, ..., G)$ may vary due to differences in environment, land topography, climate, Soil Moisture, soil quality, and other conditions. Farm households use land ${N_i}^*$ and part of or total labor time ${L_i}^*$ to produce output ${y_i}^*$ which is given by:

$$ {y_i}^* = A\Gamma_i s_i \left({L_i}^{*\beta} {N_i}^{*1-\beta} \right)^{\sigma} \tag{1} $$

Where ${L_i}^*$ is household labor endowment and A is a common productivity parameter. Production exhibits constant elasticity of substitution CES technique in agriculture. The parameter $\sigma \in (0, 1)$ indicates return to scale. It is assumed that farms production exhibits constant return to scale. We begin by assuming that there are no agricultural labor and land markets. This implies that the distribution of agricultural land does not change over time, and that sale or leasing of additional land is restricted. Later on, we relax these assumptions. Each farm produces output ${y_i}^*$ exploiting their fixed land ${N_i}^*$ and family-labor ${Lf_i}^*$. The initial land-labor ratio ${Q_i}^*$ in production varies according to farm size and can be described as follows:

$\left(\frac{{L_i}^*}{{N_i}^*}\right) = {Q_i}^*$

$$\left(\frac{{L_i}^*}{{N_i}^*}\right) = \begin{cases} \left(\frac{{L_i}^*}{{N_i}^*}\right)^s \; if \; {N_i}^* = N^{*s} \\ \\ \left(\frac{{L_i}^*}{{N_i}^*}\right)^l \; if \; {N_i}^* = N^{*l} \end{cases}$$

where $\mathrm{N}^{*s} < N^{*l} \wedge \left({L_i}^*/{N_i}^*\right)^s > \left({L_i}^*/{N_i}^*\right)^l, s \hat{=} N^{*s} \wedge l \hat{=} N^{*l}$

Here s represents small farms and l represents large farms. Variables with $*$ denote initial values in the absence of land and labor markets. In contrast, variables without $*$ represent the respective final values in the presence of land and labor markets as introduced later.

Proposition-1: *In the absence of land and labor markets, small farms are more labor-intensive than large farms. Therefore, the marginal product of labor and wages will be lower in small farms than in large farms.*

In the presence of excess labor inputs in small farms and the absence of labor market, small farms are relatively more labor intensive, and as a result, the marginal product of labor (MPL) and the wages of the small farms is lower than that of large ones. This can be given as follows:

$$\left(\frac{{L_i}^*}{{N_i}^*}\right)^s > \left(\frac{{L_i}^*}{{N_i}^*}\right)^l \wedge \frac{{L_i}^*}{{N_i}^*} = {Q_i}^* \tag{2}$$

$$\left(MPL^*\right)^s = W^{*s} < \left(MPL^*\right)^l = W^{*l} \tag{3}$$

Subsequently, we relax the assumption of the absence of labor market. To achieve an efficient input-mix, small farm-households can now supply their excess labor hours La_i to the away-farms at a wage W_a, whereas large farms can hire-in additional labor hours Lh_i at a wage W_h. Similarly, small farm-households can hire-in additional land N_i, whereas large farm-households can supply their surplus land to the market at a land rental rate of r. As a result, small and large farms' demand for labor and land can be given as follows:

$$Small farms \begin{cases} L_i = {Lf_i}^* - La_i \\ N_i = {N_i}^* + N_r \end{cases} \tag{4}$$

$$Large farms \begin{cases} L_i = {Lf_i}^* + La_i \\ N_i = {N_i}^* - N_r \end{cases} \tag{5}$$

Proposition-2: *In the presence of incomplete land and labor markets, the disparities between small and large farms relative to labor intensity, marginal product of labor, and wages wane but do not entirely disappear.*

In the presence of labor market, the substitution between family and hired labor takes place as follows:

$$L_i = aLf_i + bLh_i \tag{6}$$

Normalising equation (6) with respect to family labor Lf_i we get:

$$L_i = Lf_i + \theta Lh_i \tag{7}$$

where

$$\theta = \frac{b}{a} \tag{8}$$

θ is the productivity of family labor relative to hired labor. Equation (7) shows that small farms, which have an abundance of self-employed labor, rely solely on family labor, whereas large farms use a combination of hired and family labor in their production techniques. The labor demand of the farms of different sizes is given as follows:

$$\text{Lh}_i \begin{cases} \geq 0 \text{ , if } Ni = N_l \\ \\ = 0 \text{ , if } Ni = N_s \end{cases}$$

The production technology of small farms y_{si} (based on self-employed family labor) and large farms y_{li} (based on both family and hired labor) can be given as follows:

$$y_{li} = A\Gamma s_i \left[\left(Lf_i^{\,*} + \theta Lh_i \right)^{\beta} N_i^{1-\beta} \right]^{\sigma} \tag{9}$$

$$y_{si} = A\Gamma s_i \left(Lf_i^{\beta} N_i^{1-\beta} \right)^{\sigma} \tag{10}$$

2.3 Household preferences

We assume that household welfare depends positively on consumption C_i and negatively on labor supply to the family farm $Lf_i^{\,*}$ and the away-farm La_i. The labor supply elasticity of the family- and away-farms depends on the discomfort of working on others' fields relative to working on the family farm. In traditional agrarian societies, people live together in tribes and clans as friends and rivals. Working on others' farms is considered below one's family honor or status (Bardhan, 1973)(p.1380). Moreover, off-farm work requires commuting which may involve travel cost, accommodation charges and staying away from family and friends for an extended period of time. A farm worker does not compare his/her earnings from the family farm with the gross wage earned at the away-farm but with the potentially much lower net income that he/she brings home (Morten, 2019). Individuals who live in rural areas in their ethnic communities, clans or social groups also benefit from a variety of social insurances. As a result, individuals choose to work on their family farms even if they can receive higher wages in farms elsewhere (Gollin et al., 2014). Individuals prefer to work on their own farm as long as the loss in net wages is lower than the disutility of not working on the away-farm. Consequently, much of the surplus labor belonging to mostly small farm households does not enter the market for paid labor. Evidence from the Indian National Sample Survey Organization (1972-73) data endorses this argument. According to the survey, no individual (male or female) in the Indian states of Himachal Pradesh and Delhi and the territory of Pondicherry was found to be willing to work away from family farms, even if he/she could get higher wages in away farms. In Pakistan, small farmers, despite having a surplus of family labor, account for only 9.6% of hired labor working in agriculture, whilst landless workers account for all the remaining hired labor (PSLM/HIES-2018-19). This creates a distortion in the agriculture labor market. The labor supply to away-farms by households of various farm sizes is given by:

$$\mathrm{La}_i \begin{cases} \geq 0 \text{ , if } N_i = N_s \text{ or } N_i = 0 \\ \\ = 0 \text{ , if } N_i = N_l \end{cases} \qquad \text{and } Lf_i > 0 \;\forall\; \mathrm{N}_i > 0$$

Small farmers seek to sell their surplus family labor to the market, whereas large farms hire them. The preferences of small, large and landless farm-households in terms of the labor supplied to family and away-farms, denoted by U_{si}, U_{li}, and U_{ni} respectively, are given as follows [3]:

$$U_{li}\left(C_i, Lf_i\right) = \left[\log C_i - \xi\left(\frac{Lf_i^{*1+\frac{1}{\gamma}}}{1+\frac{1}{\gamma}}\right)\right] \tag{11}$$

$$U_{si}\left(C_i, Lf_i, La_i\right) = \left[\log C_i - \xi\left(\frac{Lf_i^{1+\frac{1}{\gamma}}}{1+\frac{1}{\gamma}}\right) - \psi\left(\frac{La_i^{1+\frac{1}{\alpha}}}{1+\frac{1}{\alpha}}\right)\right] \tag{12}$$

$$U_{ni}\left(C_i, Ln_i\right) = \left[\log C_i - \psi\left(\frac{Ln_i^{1+\frac{1}{\alpha}}}{1+\frac{1}{\alpha}}\right)\right] \tag{13}$$

Where, $\gamma > 0$ and $\alpha > 0$ are the household labor supply elasticities for the family and away-farms with the condition $\gamma \geq \alpha$. $\xi > 0$ and $\psi > 0$ are scale dis-utility parameters associated with family- and away-farm labor, respectively. From the above discussion, we can deduce that the agricultural households labor supply to the family-owned and away-farms does not exclusively depend on economic (production) factors but is also influenced by specific cultural and social norms (household preference) prevalent in the society.

Proposition-3: *If family labor engagement depends on farm household preferences, the production and utility functions become non-separable.*

Following the studies from countries such as Canada (Lopez, 1984), Peru (Vakis et al., 2004;

[3] The function is similar to the one stated (Long et al., 1983)

Jacoby, 1993), India (Skoufias, 1994) and Côte d'Ivoire (Grimard, 2000), we assume that the production and consumption functions are non-separable, i.e., households' own- and away-farm labor supply depends not only on production factors but also on the households' consumer preferences. Here, the resource constraints of small, medium, and landless farmers can be given by the equation:

$$C_{li} \leq A\Gamma_i s_i \left[\left(L{f_i}^* + \theta Lh_i \right)^{\beta} N_i^{1-\beta} \right]^{\sigma} - W_h L_{hi} + rN_{ri} \tag{14}$$

$$C_{si} \leq A\Gamma s_i \left(Lf_i^{\beta} N_i^{1-\beta} \right)^{\sigma} + W_a La_i - rN_{ri} \tag{15}$$

$$C_{ni} \leq W_a Ln_i \tag{16}$$

The household can supply labor to the own-farm Lf_i, and excess labor to the away-farm La_i, such that:

$$L_i = La_i + Lf_i \tag{17}$$

Similarly, households can substitute total time allocation between work L_i and leisure ℓ_i:

$$T_i = L_i + \ell_i \tag{18}$$

2.4 Optimum Choices

The representative household chooses the sequence $\{Ci, Lf_i, L_{ai}, N_i\}$ that maximises its utility function given by equations (11), (12),and (13) subject to the resource constraints given by equations (14), (15), and (16). The corresponding optimization program can be written as:

$$\underset{\{Ci, Lf_i, L_{ai}, N_i\}}{Max} \quad \begin{cases} U_{li}\left(C_i, Lf_i, N_i\right) \\ U_{si}\left(C_i, Lf_i, La_i, N_i\right) \\ U_{ni}\left(C_i, Ln_i\right) \end{cases}$$

$$\text{s.to} \begin{cases} C_{li} = A\Gamma s_i \left[\left(L{f_i}^* + \theta Lh_i \right)^{\beta} N_i^{1-\beta} \right]^{\sigma} - W_h L_{hi} + rN_{ri} \\ C_{si} = A\Gamma s_i \left(Lf_i^{\beta} N_i^{1-\beta} \right)^{\sigma} + W_a L_{ai} - rN_{ri} \\ C_{ni} = W_a Ln_i \\ L_i = La_i + Lf_i \\ T_i = L_i + \ell_i \end{cases}$$

The Lagrangian functions for maximization are given by:

$$\int = \left[\log C_i - \xi \left(\frac{Lf_i^{1+\frac{1}{\gamma}}}{1+\frac{1}{\gamma}} \right) \right] + \lambda_l \left[A\Gamma s_i \left(\left(Lf_i^{\ *} + \theta Lh_i \right)^{\beta} N_i^{1-\beta} \right)^{\sigma} - W_h Lh_i + rN_{ri} - C_i \right]$$

$$\mathcal{L} = \left[\log C_i - \xi \left(\frac{Lf_i^{1+\frac{1}{\gamma}}}{1+\frac{1}{\gamma}} \right) - \psi \left(\frac{La_i^{1+\frac{1}{\alpha}}}{1+\frac{1}{\alpha}} \right) \right] + \lambda_s \left[A\Gamma s_i \left(Lf_i^{\beta} N_i^{1-\beta} \right)^{\sigma} + W_a La_i - rN_{ri} - C_i \right]$$

$$\mathcal{N} = \left[\log C_i - \psi \left(\frac{Ln_i^{1+\frac{1}{\alpha}}}{1+\frac{1}{\alpha}} \right) \right] + \lambda_n \left[W_a Ln_i - C_i \right]$$

Where λ_i are the Lagrange multipliers associated with household resource constraints. Here, the first order conditions are:

$$\frac{1}{C_i} = \lambda_l = \lambda_s = \lambda_n \tag{19}$$

$$\xi Lf_i^{\frac{1}{\gamma}} = \lambda_s \beta\sigma A\Gamma s_i \left(Lf_i^{\beta} N_i^{1-\beta} \right)^{\sigma} Lf_i^{-1} \qquad (focLf_s) \tag{20}$$

$$\xi Lf_i^{\frac{1}{\gamma}} = \lambda_l \beta\sigma A\Gamma s_i \left[\left(Lf_i^{\ *} + \theta Lh_i \right)^{\beta} N_i^{1-\beta} \right]^{\sigma} Lf_i^{-1} \qquad (focLf_l) \tag{21}$$

$$\psi La_i^{\frac{1}{\alpha}} = \lambda_s Wa \tag{22}$$

$$\psi Ln_i^{\frac{1}{\alpha}} = \lambda_n Wa \tag{23}$$

$$Wh = \beta\sigma A\Gamma s_i \left[\left(Lf_i^{\ *} + \theta Lh_i \right)^{\beta} N_i^{1-\beta} \right]^{\sigma} \theta Lh_i^{-1} \tag{24}$$

$$r = (1-\beta)\sigma A\Gamma s_i \left[Lf_i^{\beta} N_i^{1-\beta} \right]^{\sigma} N_i^{-1} \tag{25}$$

By definition, the wage paid to the labor hired by large farms households is identical to the wage received by the away farm labor:

$$Wa = Wh \tag{26}$$

Combining (20) and (22) gives:

$$Wa = MPL_f . \left(\frac{\psi La_i^{\frac{1}{\alpha}}}{\xi Lf_i^{\frac{1}{\gamma}}} \right) \tag{27}$$

Similarly, combining equation (27) with (24) and (26), we get:

$$\frac{MPL_h}{MPL_f} = \frac{\psi La_i^{\frac{1}{\alpha}}}{\xi Lf_i^{\frac{1}{\gamma}}} \tag{28}$$

Dividing equation (24) by (25) we get:

$$\frac{MPL_h}{MPN} = \frac{W_h}{r} \tag{29}$$

At a given market wage rate (MWR) of the away-farm labor W_a,the allocation of labor between home production and away-farms is determined by the marginal product of self-employed labor, the relative discomfort of working at the away-farm, and the labor supply elasticity of the owned and away-farms. In other words, the labor market is in equilibrium if the marginal product of family-labor is equal to the given market wage W_a. In such case, the ratio of the marginal dis-utility of the family-farm labor to the away-farm labor is equal to $\left(\frac{\psi La_i^{\frac{1}{\alpha}}}{\xi Lf_i^{\frac{1}{\gamma}}}\right) = 1$.

2.5 Market distortion due to the perception of family dishonor

The ratio $\left(\frac{\psi La_i^{\frac{1}{\alpha}}}{\xi Lf_i^{\frac{1}{\gamma}}}\right) \geq 1$ if the perception of family dishonor or hesitation for any other reason prevents individuals from working away from the family farm. This creates distortion in the labor market. As a result, the marginal product of self-employed labor decreases by the factor $\left(\frac{\xi Lf_i^{\frac{1}{\gamma}}}{\psi La_i^{\frac{1}{\alpha}}}\right)$ against the given agriculture MWR W_a. Deviation of the term $\left(1 - \frac{\xi Lf_i^{\frac{1}{\gamma}}}{\psi La_i^{\frac{1}{\alpha}}}\right)$ from the unitary value indicates the extent of distortion in the agriculture labor market.

From the above discussion, we note that

$$\frac{\psi La_i^{\frac{1}{\alpha}}}{\xi Lf_i^{\frac{1}{\gamma}}} \begin{cases} = 1, \text{ if } W_a = W_f = MPL_f = W^s = W^l (\text{No family dishonor}) \\ \\ > 1 \wedge W_a > MPL_f \vdash W^l > W^s, \text{otherwise,(family dishonor)} \end{cases} \tag{30}$$

Combining equations (27) and (29), we can see that the family and hired labor intensity of production depends on the notion of family dishonor while working away from the family-owned farm. This can be shown as follows:

$$\frac{Lf_i}{N_i} = \frac{\beta}{(1-\beta)} \frac{r}{W_a} . \left(\frac{\psi La_i^{\frac{1}{\alpha}}}{\xi Lf_i^{\frac{1}{\gamma}}}\right) \tag{31}$$

$$\frac{Lh_i}{N_i} = \frac{\theta\beta}{(1-\beta)}\frac{r}{W_f}.\left(\frac{\xi Lf_i^{\frac{1}{\gamma}}}{\psi La_i^{\frac{1}{\alpha}}}\right) \tag{32}$$

Equations (31) and (32) show that, , in addition to production factors, family and hired labor intensity of production are affected by relative family dishonor $\left(\frac{\psi La_i^{\frac{1}{\alpha}}}{\xi Lf_i^{\frac{1}{\gamma}}}\right)$ in working on the away-farm. Nevertheless, it depends exclusively on production factors in the absence of market distortion arising due to the perception of family dishonor. Solving equations (20) and (22) simultaneously gives household labor supply to the family- and away-farms:

$$Lf_i = \left(\frac{MPL_f.\psi La_i^{\frac{1}{\alpha}}}{\xi Wa}\right)^{\gamma} \tag{33}$$

$$La_i = \left(\frac{Wa.\xi Lf_i^{\frac{1}{\gamma}}}{\psi MPL_f}\right)^{\alpha} \tag{34}$$

Equation (34) shows that away-farm agricultural labor supply is associated directly with the agriculture MWR and the dis-utility derived from working on the family-farm. Supply, on the other hand, is inversely proportional to the marginal product of the family-farm labor. Similarly, equation (33) illustrates that family labor supply to the family-owned farm is directly proportional to the marginal product of the family labor and the family dishonor associated with working on the away-farm. However, the ratio is inversely related to the given market wage.

This shows that, in the presence of cultural values which impede small land-owners from working on the away-farms, small farms become more labor-intensive than large farms. Consequently, the MPL and wage rate of the family labor is less than that of the hired labor, as shown below:

$$\left(\frac{L_i}{N_i}\right)^l \begin{cases} = \left(\frac{L_i}{N_i}\right)^s \vdash w_a = W_f = MPL_f \text{ if } \left(\frac{\psi La_i^{\frac{1}{\alpha}}}{\xi Lf_i^{\frac{1}{\gamma}}}\right) = 1 \\ < \left(\frac{L_i}{N_i}\right)^s \vdash W_a > MPL_f \text{ , otherwise} \end{cases} \tag{35}$$

Combining equation (25) with (27) we associate land productivity (yield) with family dishonor:

$$\frac{y_i}{N_i} = \frac{MPL_f}{W_a}.\left(\frac{\psi La_i^{\frac{1}{\alpha}}}{\xi Lf_i^{\frac{1}{\gamma}}}\right).\frac{r}{(1-\beta)\sigma} \tag{36}$$

From equation (36) we can deduce that, apart from production factors, the higher the perception of family dishonor from working on the away farms, the higher the land productivity (yield).

Furthermore, large farms do not need to send their family labor to away farms, given their high land endowment. On the other hand, smallholders with limited land and surplus family labor are subject to family dishonor if they choose to work on away farms. As a result, only the small-farm families potentially face family dishonor, while the labor of large farm depends solely on production factors. Subsequently, the yield and labor intensity across farms of varying sizes in the context of family dishonor is described as follows:

$$\left(\frac{y_i}{N_i}\right)^l \begin{cases} \leq \left(\frac{y_i}{N_i}\right)^s \text{ if } \left(\frac{L_i}{N_i}\right)^l \leq \left(\frac{L_i}{N_i}\right)^s f_L > 0 \wedge f_{LL} < 0 \\ > \left(\frac{y_i}{N_i}\right)^s \text{ otherwise} f_{NL} > 0 \end{cases} \tag{37}$$

From the above equation, we can infer that keeping other factors constant, small farms become more labor-intensive due to the notion of family dishonor, therefore, their yield per acre will be higher than that of large farms. This may however diminish the productivity of other inputs that are directly associated with farm size, such as labor and machinery. The overall productivity of the small farms may therefore be lower, even though their yield per acre may be higher. Hence, the TFP depends on the second-order own and cross partial derivatives of all inputs with respect to land and labor. Consequently, the relationship between TFP and farms size may be positive or negative depending upon the combined efficiency and productivity of all farm inputs. This can be given as follows:

$$\left(\frac{y_i}{N_i}\right)^l < \begin{cases} \left(\frac{y_i}{N_i}\right)^s \left(\frac{L_i}{N_i}\right)^l < \left(\frac{L_i}{N_i}\right)^s \dashv \left(\frac{y_i}{I_i}\right)^l < \left(\frac{y_i}{I_i}\right)^s \\ \quad if \ (f_{NL} + f_{ZL} + f_{LL}) < (f_{LN} + f_{ZN} + f_{NN}) \\ \\ \left(\frac{y_i}{N_i}\right)^s \vdash \left(\frac{y_i}{I_i}\right)^l < \left(\frac{y_i}{I_i}\right)^s, otherwise \end{cases} \tag{38}$$

Here,

f$_{NL}, f_{ZL}, f_{LN}, f_{ZN} > 0, \quad and \ f_{LL}, f_{NN} < 0$

I_i is a vector of total input mix and $f_{NL}, f_{ZL}, f_{LN}, f_{ZN}$ are cross partial derivatives. Correspondingly, f$_{LL}$ *and* f$_{NN}$ are second-order partial derivatives with respect to labor and land, respectively.

2.6 Aggregates

In equilibrium, the total labor endowment (labor supply) of agricultural families, with and without land ownership, equals the sum of self-employed labor and hired labor (demand for labor):

$$\sum Lf_i^* + \sum Ln_i = \sum Lf_i + \sum Lh_i \tag{39}$$

Alternatively,

$$\sum Lf_i + \sum La_i + \sum Ln_i = \sum Lf_i + \sum Lh_i \tag{40}$$

where $\text{Lf}_i^* = Lf_i + La_i$

In other words, the demand for excess labor by the large farms is equal to small farms' surplus labor supplied to the away farms and the labor supply of the landless families.

$$\sum La_i + \sum Ln_i = \sum Lh_i \tag{41}$$

Correspondingly, the total consumption of small farms, large farms, and landless agriculture-related families is equal to the combined output produced by the small and the large farms.

$$\sum y_{li} + \sum y_{si} = \sum C_{ni} + \sum C_{li} \sum C_{si} \tag{42}$$

3 Data and methodology

3.1 Data

We primarily used Pakistan social and living standard measurement survey (PSLM/HIES 2018-19) for our empirical analysis. PSLM 2018-19 is an extensive survey of 25,676 households, out of which 5,645 (22.73%) are farm households that are involved in agriculture or livestock farming [4] [5]. The PSLM/HIES gathers extensive information on agriculture and livestock including the types of input and their cost, crops and livestock produced, consumed and sold, and the value of farm production. We obtain information about farmland quality by linking the survey data with remote sensing satellite imagery and geo-coded data including Harmonized World Soil Database (HWSD), USGS/NASA Shuttle Radar Topography Mission "SRTM" Version-1 at 3-arc-second, or 90m resolution, ASTER GDEM (Global Digital Elevation Model), and NASA's Earth Data/ NASA Global Land Data Assimilation System Version 2.1 (GLDAS-2.1).

3.2 Merging satellite based remote sensing and geo coded data with PLSM-2018-19

We construct a number of farm productivity control measures related to land suitability for agriculture by using remote sensing data. For this purpose, we link satellite imagery with PSLM/HIES data. Given that the HIES data is not geocoded, we first linked the PSLM-HIES survey with Pakistan's population census at the enumeration block-level to obtain each household's locality

[4] We exclude 83 observations showing negative livestock income even though the households did not report any loss or theft of livestock.

[5] These households own agricultural land and operate the farm either as owner or tenant. Households with hired labor that do not own land, and those engaged solely in livestock without farming are not included.

address. Next, we locate the geo-coordinates of all enumeration blocks manually from the Google Earth software. We identify the land suitability of all farms by using geo information-based soil quality data from the Harmonized World Soil Database (HWSD), constructed jointly by the International Institute for Applied System Analysis (IIASA) and Food and Agriculture Organization (FAO) GIS information. We construct other measures of land suitability for agriculture by using remote sensing data of the United States Geological Survey USGS/NASA Shuttle Radar Topography Mission "SRTM" Version-1 at 3-arc-second, or about 90 meters (295 feet) resolution. The measures include farm elevation (in meters), terrain ruggedness, and distance (in kilometers) to the nearest water body. Most voids in SRTM Version-1 are filled with elevation data from the ASTER GDEM (Global Digital Elevation Model). ASTER is a sensor on NASA's Terra satellite that uses stereoscopic imaging via optical parallax to measure elevations from where clouds do not mask it. Other small void areas are filled using interpolation methods described by Reuter et al. (2007). Finally, we construct the indicator for underground soil moisture from the data taken from NASA's Earth Data/ NASA Global Land Data Assimilation System Version 2.1 (GLDAS-2.1). The construction and description of all the above variables are discussed as follows:

3.3 Variable construction and measurement bias

The relationship between farm size and farm productivity may vary depending on soil quality, topography, soil moisture, type of inputs available, and source of irrigation (Vollrath, 2007). The higher yield of small farms may also be attributed to the asymmetric distribution of soil quality across farms of different sizes, as small and medium farms usually have better quality soils.

3.3.1 Soil quality

We base the farms' soil quality indicator on the Harmonized World Soil Database (HWSD) constructed jointly by the International Institute for Applied System Analysis (IIASA) and Food and Agriculture Organization (FAO) GIS information. The algorithm used to measure Soil Quality is based on the nutrients available in the soil. The model integrates location-specific soil traits like soil texture/mineralogy, soil pH, organic carbon, and total exchangeable bases (TEB) to compute the nutrients availability index for 0-30cm topsoil. However, for the 30-100cm subsoil, the index is measured using the same nutrients apart from organic carbon. The index ranges from 1 (no or slight productivity constraint) to 7 (severe productivity constraint). The data is available with geographical information in spatial Grid cells of 5 minutes or 30 arc seconds [6]. Figure 1 presents farm localization by corresponding soil quality. Most small farms are of good soil quality. The proportion of large farms with topsoil (60%) is much lower than that of small farms (80%). On the other hand, 3% and 8% of small and large farms have soils that lie in the bottom soil category, respectively. Small farms have an average soil quality of 1.27 (7 being the poorest-quality soil)

[6] For details, see (Fischer et al., 2008)

compared to 1.57 for medium farms and 1.62 for large farms (Figure 2). This is in part because farmers generally sell off less productive or poor-quality land in times of need which progressively leads to higher average soil quality among smaller farms and lower average quality among larger farms.

3.3.2 Root soil moisture

The indicator for Volumetric Soil Moisture (VSM) is taken from NASA's Earth Data/ NASA Global Land Data Assimilation System Version 2.1 (GLDAS-2.1). It corresponds to the volume of water per unit volume of soil (g/cm^3) found 0-10cm underground. It is measured as the average moisture of the soil within a 2km radius of the farm locality. figure 3 shows the farms' geographical localization by the availability of 0-10cm underground soil moisture. This indicates that the northern parts and upper Punjab have considerably higher levels of underground soil moisture.

3.3.3 Land elevation and ruggedness

Variation in land ruggedness plays an essential role in determining the farm productivity Nunn & Puga (2012). Land ruggedness measures the mean absolute variation of the central pixel relative to eight surrounding cells in a given DEM cell. The index value is categorized as level: 0-80m, nearly level: 81-116m, slightly rugged: 117-161m, intermediately rugged: 162-239m, moderately rugged: 240-497m, highly rugged: 498-958m, and extremely rugged: 959-4367m . The elevation and land ruggedness of individual farms is calculated as the mean value for the area within a 2km radius around the farms' locality based on data from SRTM Version-1 at 3-arc-second. Given some vides in the SRTM Digital Elevation Model and the unavailability of exact farm geolocation, we create a buffer of 2km radius around the farm locality to measure the average elevation and ruggedness of the farmland within the 2km radius. Elevation is measured in meters Above Mean Sea Level (AMSL). Figures 4 and 5 respectively show geo-specific land elevation and ruggedness patterns. The ruggedness of farm terrain is mainly mild to medium, implying that most farms are located in or near the river plains (Figure 6). Farms are on average 46 km from the nearby water channel. Around 80% of the farms are irrigated by tube-wells, particularly those located at higher altitude, rough surfaces, or far from the main watercourse. Most farms in the KPK and Balochistan provinces are watered by tubewells, while the Indus and its tributaries irrigate most farms in the provinces of Punjab and Sindh.

3.3.4 Other controls

Other variables used as determinants of farm productivity include per acre input cost, family labor participation, irrigation status, land ownership, household head's education attainment, and distance from farm to market (measured as transportation cost as a proportion of total output).

Family labor participation is measured in three ways: 1) family workers per acre, 2) the ratio of family labor cost to total labor cost, and 3) the number of on-farm family workers.

3.3.5 Crop choice

Another source of bias in productivity measurement may result from the type of crops cultivated. For instance, fruit producers incur initial fixed input cost without producing any output in the early years. Later, they produce more output with little additional input cost compared to farms producing seasonal or annual crops such as cereals or vegetables. The TFP of fruit farms may therefore appear higher than that of farms producing vegetables and cereals. We introduce a control variable for the type of crops produced at the farm to account for any related bias. We divide crops into five categories: cereals, cash crops, fodder, vegetables and fruits. Farms are grouped into these categories on the basis of the share of principal crop they produce. The division of crops and the distribution of farms by the crops they produce is presented in Table 1. Most small and medium farms produce cereals, cash crops, and fodder. Large farms mainly produce cereals, cash crops, and vegetables and pulses. A minimal portion of small and medium farms; 4% each produce vegetables as their primary crop. Similarly, only 4% and 2% of small and medium farms principally produce fruits. This is in spite of the fact that vegetables and fruits are labor-intensive crops. In the result section, we discuss the optimal crop choice for the three farm categories.

3.3.6 Reported price bias

A typical problem in the aggregation of inputs and multiple outputs produced in farms of different sizes involves price differentials. Small farms often sell their products to retailers or end-use consumers at higher stated prices than large farms that sell their produce to wholesalers. We cross-check the reported prices of six major cultivated crops, namely wheat, rice, maize, fodder, sugarcane, and cotton. Except for sugarcane and cotton, prices of all major crops show substantial variability by farm size, ranging from 15% for rice to 270% for maize and wheat. Small and medium farms report higher crop sale prices compared to large farms (Figure 7). To check the robustness of the results to this issue, we harmonize output prices by re-estimating aggregate farm output using wholesale price indices (i.e. wholesale price per 40 kg) [7] . The revised results (shown in tables 9 and 10 in the robustness section) remain identical to the baseline estimates.

3.3.7 Imputation of land rent

A potential challenge to the accurate measurement of total factor productivity is the unavailability of the imputed rent of the family-owned farms. Out of the total 5,561 farms surveyed in PSLM-HIES 2018-19, only 748 are partially or entirely rented. All the other farms are owner-operated,

[7] except for fruits and vegetables, given their high price variability.

for which imputed rent figures are not available. The absence of rent cost from the calculation of TFP can affect the results, especially if the cost differs by farm size. To address this issue, we re-estimated the model with imputed land rent. We impute the missing rent cost from the available data for partially or wholly-rented farms. The calculations show that the per acre land rent of small farms is 16% higher, i.e., 35,072 rupees, than that of large farms, 30,276 rupees (Figure 8). This difference in rent owes, in part, to the difference in soil quality and farms locality around water bodies.

4 Empirical Model and Estimation Method

Agricultural productivity can be attributed to two components: production technology (i.e. productivity growth due to type and quality of inputs and the available technology used in production), and technical efficiency (i.e. how well are the inputs combined in production) (Mechri et al., 2017). For instance, the relatively higher overall productivity of large farms compared to small farms could be due to better technology available to them, even though they might be technically inefficient. Farm productivity is often taken as a synonym of technical efficiency as farms are considered technically efficient. However, not all farms are technically efficient (Grosskopf, 2003; Färe et al., 1989; Nishimizu & Page, 1982). Therefore, the overall productivity growth could be due to technological progress as well as technical efficiency. The variation in productivity across farm size could be mainly attributed to production technology, while the variation within the farms of similar size could mainly be attributed to technical efficiency.

4.1 Agriculture productivity

We use the Cobb-Douglas production function, described in equation (1), to measure agriculture productivity. The generalized form of agriculture production with farm controls is presented as follows:

$$y_i = A\Gamma_i s_i \left(L_i^{\beta} N_i^{1-\beta}\right)^{\sigma} \tag{43}$$

Where 'N_i ' corresponds to farmland, 'L' is labor, 'A' is a common technological productivity driver, s_i are farm-specific productivity factors that depend on the type of inputs and the farmer's ability, the way he or she combines the given inputs. Γ_i is a regional/environmental productivity parameter that captures the influence of factors such as soil quality, soil moisture, topography, elevation, farm specialization, and other factors.

To measure the impact of farm size on productivity using crop yield, we divide both sides by N_i and then take the natural log. This yields the following equation:

$$ln\frac{y_i}{N}_i = lnA + ln\Gamma_i + lns_i + \beta\sigma ln\frac{L_i}{N_i} + (1-\beta)\sigma lnN_i \tag{44}$$

The direct measure of farm productivity s_i is not available. It can be represented by a set of variables as follows:

$$lns_i = \theta ln\frac{Lf_i}{L_i} + \gamma lnZ_i + \epsilon_i \tag{45}$$

where $\frac{Lf_i}{L_i}$ represents the share of family labor in total labor, Z_i is a vector of farm specific characteristics such as farmer education, farm irrigation and raising livestock along with crops, and ϵ is the unexplained random productivity component.

Substituting the farm productivity measure s_i from equation (45) into (44), we get:

$$ln\frac{y_i}{N}_i = lnA + ln\Gamma_i + \theta ln\frac{Lf_i}{L_i} + \gamma lnZ_i + \beta\sigma ln\frac{L_i}{N_i} + (1-\beta)\sigma lnN_i + \epsilon_i \tag{46}$$

If the assumption of constant return to scale holds (= 1), then the above equation can be used to estimate the impact of farm size on productivity (crop yield). Otherwise, the farm-size parameter $i - \beta$ must be inflated or deflated by the term σ with decreasing or increasing returns to scale, respectively.

Given that each farm can produce multiple outputs in a given year, and the area allocated to individual crops is not classified, we aggregate the value of all crops to obtain the combined value of product per acre. Similarly, in equation (46), the term $\frac{L_i}{N_i}$ shows the number of workers per acre. However, the survey reports the cost of hired-labor but not the units of unpaid family labor. In order to harmonize the labor employed in farm production, we convert the units of family labor into monetary terms by multiplying it with market wage W_h. This can be given as follows:

$$ln\frac{Y_i}{N}_i = lnA + ln\Gamma_i + \theta ln\frac{Lf_i}{L_i} + \gamma lnZ_i + \beta\sigma ln\frac{W_h.L_i}{N_i} + (1-\beta)\sigma lnN_i + \epsilon_i \tag{47}$$

where

$$(Y_i = P_j.y_{ji})$$

P_j is the price of the product j^{th} produced by the i^{th} farm-household. If $\sigma \in 1$ and there is no labor market imperfection across farm-size, the term $(1-\beta)$ in equation (47) will represent the actual relationship between farm-size and the yield measure of productivity. According to the theoretical model, small farms that rely heavily on family workers earn lower wages than large farms that rely on hired labor. In this case, the wage rate in large farms is W_h. However, the given wage rate decreases by the factor $(1-\vartheta_i)$ as farm size decreases. The parameter $\vartheta \in (0,1)$ is inversely proportion to farm-size. Failing to account for size-dependent wage imperfection therefore leads to biased estimates of the size-productivity relationship. To account for this bias, we include the cost of self-employed (family) labor equivalent to the marginal product of labor (MPL) rather than the given market wage W_h. The subsequent equation is given as follows:

$$ln\frac{Y_i}{N_i} = lnA + ln\Gamma_i + \theta ln\frac{Lf_i}{L_i} + \gamma lnZ_i + \beta\sigma(1-\vartheta)ln\frac{W_h.L_i}{N_i} + (1-\beta)\sigma lnN_i + \epsilon_i \quad (48)$$

Any significant difference between the size-parameter of equations (47) and (48) can be used to test the impact of wage differences across farms of varying sizes on farm productivity.

Next, we use the overall farm productivity measure called Total or Multiple Factor Productivity (TFP / MFP) introduced by O'Donnell (2012); Zhao et al. (2012). The following equations are used to estimate the size and TFP relationship:

$$lnT_i = lnA + ln\Gamma_i + \theta ln\frac{Lf_i}{L_i} + \gamma lnZ_i + (1-\beta)\sigma lnN_i + \epsilon_i \quad (49)$$

$$ln\hat{T}_i = ln\hat{A} + ln\Gamma_i + \hat{\theta} ln\frac{Lf_i}{L_i} + \hat{\gamma} lnZ_i + (1-\hat{\beta})\hat{\sigma} lnN_i + \hat{\epsilon}_i \quad (50)$$

Ti is a measure of total factor productivity that assumes similar wages across family and hired labor. However, $\hat{T}i$ accounts for wage discrepancy between family and hired labor.

4.2 Technical efficiency

We explore farms' technical efficiency, i.e., how well given inputs are used across farms of various sizes to produce a given level of output using the technology available to each farm. We analyze the efficiency of family labor when they are valued at MWR instead of MPL. Various techniques are used in the literature to estimate agricultural technical efficiency. These include Average Production Function (APF) introduced by (Yotopoulos & Lau, 1973), the Engineering approach by (Herdt & Mandac, 1981), and Data Envelopment Analysis (DEA) by (Charnes et al., 1978). In this study, we use the Stochastic Frontier Analysis (SFA) model simultaneously introduced by Aigner et al. (1976) and Meeusen & van Den Broeck (1977) to analyze farm technical efficiency relative to farm size using the two productivity measures (i.e. yield and TFP). Using SFA, farms' technical efficiency in relation to their size (N_k) can be estimated through the following regression model:

$$ln\Im_i = \alpha + \varrho_k N_k + \pi_j lnX_{ji} + \vartheta_i - \upsilon_i \quad (51)$$

$\Im_i$ is a productivity measure, either yield $\frac{Y_i}{N_i}$ or total factor productivity;, T_i, and $\hat{T}_i$, where $\vartheta_i \sim iidN\,(0,\ \sigma^2_\vartheta\ )$ is a two-sided error component while $\upsilon_i \sim iid\ N^+\,(0,\ \ \sigma^2_\upsilon)$ is a non-negative half-normal technical inefficiency component. ϑ_i *and* υ_i are distributed independently of the regressors and of each other. Farm size is defined as a categorical variable that takes two categories: small (less than 2.5 acres) and large (more than 2.5 acres). With the help of estimates obtained from equation (51)), we obtain inefficiency (including and without family labor) for the two farm categories.

This estimate also helps compare the efficiency of family-farm labor with that of hired labor. The slope parameters of the above model are consistent but the intercept parameter is biased when estimated using OLS (Aigner et al., 1976; Schmidt, 1976; Subal C. Kumbhakar and C. A. Knox Lovell, 2012). Consequently, we use the Maximum Likelihood Estimator which generates the most consistent estimates of all parameters (Subal C. Kumbhakar and C. A. Knox Lovell, 2012; Aigner et al., 1976)

4.3 Testing wage differential across farm sizes

We test the hypothesis of wage imperfection across farms of varying sizes. For this, we compare the statistical difference between the MPL of the three farm categories, as described in equation (30) of the theoretical model. The following regression is used to measure the wage differences between family and hired labor involved in agriculture across farms of different sizes:

$$lnY_i = \nu + \omega_k N_k * lnL_i + \theta ln\frac{Lf_i}{L_i} + \zeta_{ji} lnZ_{ji} + \xi_i \tag{52}$$

Where Y_i is the farm output, 'ω_k' is the output elasticity of labor for the three farm categories. 'θ_i' shows the productivity of family labor relative to that of hired labor, while k k represents the three farm sizes. The sign and statistical significance of ω_k *and* θ_i help gauge the extent of labor productivity (wage) differential across farms of different sizes and the type of labor used in agriculture, respectively. OLS produces consistent estimates of equation(52), which can be given as follows:

$$\omega_k = \begin{cases} \omega_S = \omega_M = \omega_L = \text{ if } W^s = W^l \vdash \text{No wage differential} \\ \\ otherwise, W^s > W^l \vdash \text{ wage imperfection due to other factors} \\ \\ or, W^s < W^l \vdash \text{ wage imperfection due to unwillingness of family labor} \end{cases} \tag{53}$$

Where $\omega_S, \omega_M, and\ \omega_L$ represent the relative labor productivity of small, medium, and large farms respectively. Alternatively, we look for labor market distortion using the productivity of family-farm labor relative to hired labor θ, which is interpreted as follows:

$$\theta = \begin{cases} 0 \text{ if } MPL_f = MPL_h \text{ No reluctance for family labor to work in away farms} \\ \\ \text{otherwise, } R_+^* \vdash MPL_f > MPL_h \text{ No need to work for others} \\ \\ or, R_- \vdash MPL_f < MPL_h \text{ family labor are unwilling to work for others} \end{cases} \tag{54}$$

4.3.1 Potential endogeneity

The association of unobserved potential factors with farm size may result in endogeneity bias in equations (47) to (50). Given that both the potentially-endogenous covariate and the outcome variable pertain to the farm, finding a pertinent exogenous instrument is challenging. Some studies used renting-in land and double cropping as instruments to estimate the farm size and productivity relationship (Lamb, 2003). However, several studies do not use an instrumental strategy despite acknowledging the presence of endogeneity (see for example, (Omotilewa et al., 2021; Garzón Delvaux et al., 2020). We too rely on the OLS results due to lack of a relevant exogenous instrument. As a result, despite accounting for all omitted variables biases identified in the literature, we could not claim a causal relationship in the present scenario.

5 Variable Description

5.1 Dependent Variables

Crop yield, i.e. crop output per unit of land, is widely used in the literature to measure farm productivity (Sheng & Chancellor, 2019; Michael Lipton, 2009; Akram-Lodhi, 2001; Rosset, 2000; Bardhan, 1973; Berry & Cline, 1980). However, this measure does not account for the use of inputs other than land. Farms with higher crop yields do not necessarily enjoy higher profitability, higher output per worker, or greater production and cost-efficiency. As a result, the inverse farm size - productivity relationship commonly observed in the literature may be misleading as the higher crop yields observed in small farms may result from extensive family labor participation or other factor inputs.

To account for these shortcomings, we use an agriculture productivity measure called Total (or Multiple) Factor Productivity (TFP/MFP) introduced by O'Donnell (2012); Zhao et al. (2012). The measure can be described as follows:

$$T_i = \frac{Y\left(y_i\right)}{I\left(n_i\right)} \tag{55}$$

$I_i = (n_{1i}, \ldots\ldots n_{Mi})$ and $Y_i = (y_{1i}, \ldots\ldots y_{Mi})$ are sets of input and output vectors, respectively, whereas $Y\left(.\right)$ and $I\left(.\right)$ are scalar aggregator functions. As recommended by Mechri et al. (2017), we include all crops, livestock, dairy production and by-products while aggregating the inputs and outputs.

5.2 Valuation and imputation of family labor

A typical problem in estimating farm size and productivity relationship arises from the cost of family labor which is not reported in any agriculture census or labor force or social and living standards measurement survey, even though family labor accounts for a major proportion of the labor cost. In our sample, about 41% of the surveyed farms report no labor cost, implying that these are fully owner-operated family farms. On average, family labor accounts for 40.6% of the labor cost of the remaining 59% of farms based on market wage rates. On average, small farms employ 1.74 family labor per acre while large farms use only 0.14 family labor per acre. Therefore, valuation and imputation of unreported family labor cost may play a significant role in explaining the IR. We estimate family labor cost by two methods: 1) the marginal product of family labor (MPL) used as shadow wage rate (Jacoby, 1993) and 2) agricultural Market wage rate (MWR). We use both the measures of family labor to assess the significance of family labor imperfection across farm size in explaining the IR. Valuation and imputation of family labor is key to estimating the farm size and productivity relationship. However, the HIES survey does not take into account the number of hours and earnings from family labor. In order to impute the family labor wages at market rate, we link the Labor Force Survey (LFS-2017-18) with the PSLM/HIES survey to estimate the weekly wage rate W_a and the number of work hours for farm labor. The weekly work hours of family labor are estimated from the labor force survey. Then, the following formula is used to calculate the weekly wage rate for family labor and the resulting annual cost at market rate:

$$W_f = \tau W_a * \frac{weekly\ work\ hours\ of\ family\ labour}{weekly\ work\ hours\ of\ agricultural\ labour} \tag{56}$$

$$Annual\ family\ labour\ cost = W_f * 52 \tag{57}$$

Where W_f is the weekly wage rate of family labor and W_a is the market wage rate of agricultural labor.

Another problem in the imputation of family labor is the valuation of self-employed part time labor. To account for this issue, a weight parameter τ is used, which refers to the time allocation by family labor or self-employed worker to the farm. $\tau \in (0.5, 1)$ such as:

$$\tau = \begin{cases} 1 \text{ , if full time farm family worker} \\ \\ 0.5 \text{ , otherwise} \end{cases}$$

On the other hand, the imputed cost of family labor is based on the marginal product of family labor, and is measured as follows:

$$\widehat{W_f} = MPL = \hat{\vartheta}\frac{\hat{y}}{L} \tag{58}$$

Where $\hat{\vartheta}$ is the estimated elasticity of output with respect to family labor, $\hat{y}$ is the predicted output and 'L' is the family labor.

5.3 Variable of Interest

Farm size, the variable of interest, is measured in two ways: first, as a continuous variable in number of acres; second, as a categorical variable with farms categorized as small (up to 2.5 acres), medium (2.5 - 12.5 acres) or large (greater than 12.5 acres). Following Mechri et al. (2017), the measures are based on cultivated rather than available farmland. In our dataset, 43.4% of farms are classified as small, 50.5% are medium, and 6.1% are categorized as large. Small farms cover only 11.6% of the country's farmland while medium and large farms account for 57% and 31% of the country's cultivated land (Figure 9).

6 Results

6.1 Descriptive statistics

Pakistan's average farm size has considerably decreased from 7.6 acres in 2000 to 6.4 acres in 2010 and 5.3 acres in 2018-19. The distribution of farm holdings and agriculture area by farm size shown in Figure-9 points to a sharp increase in land fragmentation happening over the past decade. Large farms accounted for 61% of the total farm area in 2010 compared to 31% in 2019, while their share in the number of landholdings dropped from 13% to 6.1%. On the other hand, area covered by small farms doubled from 6% to 12% while that of medium farms increased from 32% to 57%. Likewise, the corresponding share of these two farm categories in the total number of landholdings increased from 40% to 43% and from 46% to 51%, respectively.

This substantial acceleration in land fragmentation owes to an unending division of family farms through inheritance. The share of farms under contractual cultivation arrangements has also increased over time, as land registration and ownership documentation has improved. This shows in changing farm tenancy patterns. The share of owner-based cultivated farms decreased from 82% in 2010 to 77% in 2019, while the area under cultivation decreased by a similar 5% to 70.3%.The average per acre yield of Pakistani farms (including both crop and livestock) is estimated to be 170,853 Pakistani Rupees (PKR) (Table-1). The yield of small farms is higher (PKR 249,390) than that of medium and large farms (PKR 114,255 and PKR 81,890 respectively). While the mean aggregate output-input ratio of Pakistani farms is estimated to be 2, the ratio for small farms is only 1.5. In comparison, the ratio is 2.1 for medium farms and an even higher 2.9 for large farms. The distribution of geographical and topographical indicators such as farm elevation, land ruggedness, root soil moisture and soil quality are presented in figures 1 to 5.

6.2 Agriculture productivity and farm size

6.2.1 Farm yield and farm size

We begin by studying farm yield in relation to farm size. Table-2 presents OLS results of the estimated model shown in equations 47 and 48. Columns (1) to (5) show the estimates including family labor cost measured in terms of MPL. The corresponding marginal effect, after accounting for all controls, shows that a 1% increase in farm size decreases land productivity by -0.073% (Column-4). After adjusting the size coefficient to returns to scale [8] parameter $\sigma = 0.911$ instead of 1, the negative relationship becomes stronger (marginal effect = -0.08%). The relationship remains negative and significant with and without various groups of controls, including household- and farm-specific controls and those related to climate, geography, soil quality, type of cultivated crop and cropping season. This shows that the yield of small farms is higher than that of large farms when family labor is measured with the MPL. These results are in line with those reported by Helfand & Taylor (2020); Sanchez et al. (2020); Gautam & Ahmed (2018); Kagin et al. (2015); Rosset (2000); Carter (1984); Bardhan (1973); Srinivasan (1972).

In contrast, if we estimate this relationship by measuring family labor wage using hired labor market rate, the negative relationship gets weaker and loses its statistical significance (marginal effect = -0.011%); the coefficient remains identical after adjustment to returns to scale $\sigma = 0.962$ (marginal effect = -0.0114). The farm size and productivity coefficients remain insignificant with and without controls if we impute family labor cost from hired farm-labor i.e. MWR (Column 6-8). When we include the interaction effect of family labor and farm size, the relationship becomes positive and significant. Interestingly, the interaction effect is negative and significant (marginal effect = -0.019 at MPL and -0.030 at MWR), which shows that the increase in family labor increases/decreases the strength of negative/positive association between farm size and productivity (Figure-10). After accounting for the interaction effect of family labor with farm size, the negative association reduced significantly (from -0.73 to -0.43) if family labor is imputed on the MPL basis, while it changes from negative and insignificant (-0.011) to positive and significant (0.035) when family labor is imputed as MWR. This suggests that the inverse association between farm size and land productivity may be attributable to the undervaluation or non-accounting of family labor across farm size.

Moreover, the RTS parameter $\sigma = 0.911$ indicates decreasing returns to scale if family labor is valued on MPL basis, however, reaches close to 1, CRT ($\sigma = 0.962$), when family labor is valued at the given market wage rate.

The yields of small farms could be higher when family labor is measured at MPL and lower when measured at MWR for several reasons.

First, as hypothesized in the theoretical model (Section 3), smaller family farms are endowed with more family labor per acre compared to large farms. Small farms are as much as ten times

[8] The results of returns to scale RTS are presented in Table A1..

more family-labor intensive than large farms. On average, 1.74 small-farm family workers are available per acre compared to 0.448 family workers in medium and 0.14 in large farms (Figure-11). This implies that the higher yield of small farms can be partly attributed to their higher family-labor availability. However, they are less productive given their high labor intensity resulting from the unwillingness of surplus small farm family labor to work on the away-farms. To take an example, as the number of family workers increases from 1 to 11, the farm output per worker decreases from PKR. 225,946 to PKR. 96,306. Correspondingly, small farms' output per worker is just PKR. 193,933, which is 2.6 times lower than large farms that produce PKR. 512,138 per worker (PLSM-2018-19).

The results shown in Table 2 substantiate this assertion: A 1% increase in the family's share in total labor measured as marginal product of family labor increases the farm yield by 0.029%, whereas a corresponding 1% increase in family labor valued at MWR decreases the yield by 0.030% (Table 2, Columns 5 and 8).

Likewise, the overall labor intensity of small farms is 8.2 times greater than that of large farms on market wage basis (Figure 12). The overall input intensity of small farms is 4.2 times higher than that of large farms. However, this intensity is only 2.7 times higher in the case of factors other than labor. Consequently, even though the per acre yield of small farms is higher, their MPL and output per worker is lower than that of large farms (Figure 13).

A second reason for higher yield observed in small and medium farms could be their comparatively-larger per acre livestock holdings, possibly a manifestation of their higher family labor endowment. The per acre livestock yield of medium farms is twice that of large farms while that of small farms is almost ten times as much (Figure-14). Third, the soil quality of small and medium-sized farms is on average better compared to that of large farms. The soil quality of 80% small farms lies in the uppermost soil category compared to 60% of the large farms (Figure 2 and 15).

Fourth, small and medium farms are on average situated closer to rivers and canals, reflecting greater access to irrigation networks. Average distance of small and medium farms from rivers and canals is 34 and 54 km respectively compared to 66 km for large farms. Figure 6 shows farm location around the Indus basin canal irrigation system and water bodies.

6.2.2 TFP and farm size

The aforementioned estimations provide evidence for a negative relationship between farm size and land productivity taken as yield per acre when family labor is measured in terms of the marginal product. However, the relationship turns positive when family labor is measured in terms of the MWR. Smaller farms manage to obtain a somewhat higher yield thanks to their relatively higher labor and input intensity. This latter factor is not considered when calculating the farm yield indicator. We tackle this shortcoming by introducing another measure of farm productivity known as total or multifactor productivity (TFP), which accounts for all the factor inputs. Figure-16 shows a scatter plot of the two productivity measures (yield and TFP T_i) and their linear-fit.

The two productivity measures exhibit contradicting association with farm size. Farm size and yield appear to be inversely related, whereas farm size appears to be directly associated with TFP T_i, supporting our assertion of relatively higher yield but lower TFP of the small farms associated with their higher labor and input intensity. Table 3 presents the results of farm TFP T_i regressed on farm size (i.e. the model corresponding to equations (49) and (50)). The OLS estimation including family labor cost measured by MWR shows a positive and statistically significant relationship (Table 3, Columns 5-8). However, the relationship weakens and loses statistical significance if the family labor cost is measured in terms of the MPL (Columns 1-3, Table 3). The corresponding scatterplot and its linear-fit shown in Figure 17 corroborate the absence of any significant association found in the results on MPL basis. This again highlights the role of wage rate differentials between the small farms which mainly rely on family labor, and the large farms which need to hire labor, in explaining productivity differences across farms of different sizes. The results show that the overall farm productivity increases by 0.207% with a 1% increase in farm size (Column-7). The positive results are in line with the findings of Sheng & Chancellor (2019); Foster & Rosenzweig (2021); Otsuka et al. (2013); Macdonald & Mcbride (2009); Paul et al. (2004). However, this relationship weakens substantially (marginal effect = 0.002%) and loses its statistical significance if family labor is measured using MPL. This supports our theoretical assertion that the farm size - TFP relationship changes when we account for lower implicit wages in the small farms owing to family labor's reluctance to work on away-farms. The results reported in table 3 substantiate the model assertion of lower productivity of family labor valued at market wage compared to hired labor. The TFP decreases by 0.042% with a 1% increase in family labor share in total labor measured by the MWR, but increases by 0.01% with a 1% increase in family labor share measured in terms of its MPL. To account for the actual effects of family labor on farm size in relation to productivity, we add the interaction effect of family labor with farm size in both MPL and MWR models. The interaction effect is negative and statistically significant in both equations (-0.016 and -0.029, respectively) (Columns 4 and 8, Table 3). The positive relationship between farm size and TFP weakens as the proportion of family labor working on farms grows. The strength of this positive association is lower in case when family labor is valued in terms of MPL. Nonetheless, when family labor is valued on MWR basis, the relationship loses its significance or even turns negative (Figure-18). This again substantiates our assertion that imputation of family labor significantly changes the relationship between farm size and productivity.

6.2.3 Technical efficiency and farm size

The higher TFP of large farms found above may result from their greater production efficiency due to better access to technology and quality seeds. In this section we examine how inputs are efficiently utilized by farms of various sizes for a given level of output produced. We compare the efficiency of family labor alternatively measured in terms of MPL and MWR. We examine the possibility of more efficient utilization of inputs by farms of different sizes through Stochastic Frontier Analysis (SFA). The estimates of equation (51), the SFA, presented in table-4, show that large farms are more efficient compared to small farms in terms of yield and TFP estimated with family labor cost measured either at MWR or with adjusted prices (Columns 5-8). Nevertheless, small farms are more efficient in terms of yield while in terms of TFP the association becomes insignificant if MPL is instead taken as the measure of family labor cost (Columns 1-4). The availability of surplus family labor plays a crucial role in this context. More evidence for the above argument comes from Table-4 which shows that farms operating with family labor are technically less efficient, both in yield and TFP, than those without family labor, implying lower productivity of family labor compared to hired labor at MWR (Columns 5-8, Table-4). However, family labor is technically less efficient both in terms of yield and TFP when imputed at MPL instead of MWR, regardless of the farm size (Columns 1-4, Table-4). These findings support our assertion and are consistent with the yield and TFP results discussed earlier.

6.3 Testing for wage differential

The assumption of wage disparities across farms of varying sizes is tested in two ways: first, we test for significant differences in labor productivity across farm sizes. Second, we test for the difference in productivity of family labor relative to hired labor involved in agriculture, as described in equations (27) and (52) of the theoretical model. The output elasticity of labor of the small farms is lower than that of medium and large farms (Table-5). The elasticity of small, medium, and large farms is 0.003, 0.038, and 0.054 respectively. The labor employed by the small farms is therefore less productive than that employed by medium and large farms. Alternatively, we test for labor market imperfections by comparing the productivity of family farm labor with hired labor (corresponding to equations (27), (52), and (54))using the share of family labor in farm labor costs and family worker per acre as proxies. The results reported in Table 5 again show that family labor is relatively less productive than hired labor. This result is consistent across both estimations. Farm output decreases by 0.023% and 0.057% with a 1% increase in the share of family labor in total labor and family labor per acre, respectively.

6.3.1 Testing for non-separability and unwillingness of family labor

We test the practicality of the theoretical assertion of non-separability of production and utility functions, and family labor's unwillingness to work on the away-farms, as asserted in the model equations (11) to (13). We introduce household cultural factors along with production factors to the household labor supply function. Specifically, we included various ethnic groups (Punjabi, Sindhi, Pashtun, Saraiki, Balochi, and Urdu-Speaking) and provinces (Punjab, Sindh, Balochistan, and Khyber Pakhtunkhwa) with different cultural practices. This helps us determine if the household labor supply depends exclusively on production side factors, or jointly with consumer preferences, and whether cultural factors play a significant role in determining family workers' unwillingness to work in away-farms (constraint-regime). We categorize the observations into two groups. The constrained regime, in which households labor supply to the away-farms is influenced by cultural factors, such as ethnicity and regional cultural variations. Households that do not supply labor to the away-farms are included in the constraint regime. In the unconstrained regime, labor supply is determined solely by production factors rather than cultural factors. It includes households that supply their labor to outside farms. We use the switching regression model to separate the observation into two regimes. The division of labor into the two regimes is not observable. The estimates are obtained using MLE via the E-M algorithm [9] . The statistical significance of production and family preferences (cultural factors) is tested for the two regimes. If the cultural preference factors are statistically significant in the constrained regime (self-employed labor in small family farms) but insignificant in the case of unconstrained regime (hired labor working on large farms), the utility and production functions can be considered non-separable. The labor supply of small farms to the away-farms will therefore be considered to also depend on family cultural preferences rather than exclusively on production factors. The results of constrained and unconstrained regimes presented in Table 6 show that 27% of farm workers belong to the unconstrained regime and supply their labor to the market, while 73% belong to the constrained regime, i.e. restricted to family-owned farms. The results show that most indicators of consumer preference (such as labor ethnicity and regional variations) are statistically not different from zero in the unconstrained regime, but are statistically significant in the constrained regime. These findings show that agricultural households' labor supply is influenced not only by production considerations, but also by sociocultural preferences such as the perception of family honor and status.

6.3.2 Discussion

The above findings suggest that the comparative advantage of small farms' productive and technical efficiency is embedded in the commitment of their surplus family labor to work at lower than market wage rates. Accounting for the actual value of family labor plays an essential role in determining the strength and direction of the association between farm size and productivity. The apparently higher productivity and technical efficiency of small farms in terms of yield becomes insignificantly different or even lower than that of large farms when family labor is valued at the prevailing MWR instead of MPL. Likewise, the relatively higher productive and technical efficiency of large farms in terms of TFP becomes insignificant when family labor is valued at MPL instead of MWR. On the one hand, family labor is more productive and technically efficient than hired labor when valued on the MPL basis, both in terms of yield and TFP. On the other hand, it is

[9] For details, see Hartley (1978) and Dempster et al. (1977)

less productive relative to all the measures when valued at the prevailing MWR. This indicates that small farms' relative advantage in productive and technical efficiency disappears due to their inefficient utilization of family labor. Small farms can better utilize their surplus family labor and enhance their productivity by changing their cropping patterns. On average, 77% of small farms produce cereals and fodder, while 4% produce either fruits or vegetables as their main agricultural product (Figures-19 and 20). In contrast, a lower proportion of large farms (58%) is dedicated to cereal crops and only 5% to fodder. Cereals and fodder have some of the lowest yields among all the crops cultivated in Pakistan. Controlling for input cost, the yield of fruits and cash crop-producing farms is higher by PKR. 81,296 and 34,727 compared to that of farms that produce only cereal crops (Table 7). Likewise, the yield of the farms producing vegetables and fodder is PKR. 20,161 and 11,382 higher than that of cereal-producing farms (Figure-21). The relative yield of small farms producing cereals, fodder, and cash crops is similar to those of the large ones, but is as much as three times greater for fruits and 4.5 times greater for vegetables. These two categories of agricultural produce are more labor-intensive, hence their higher per acre yield. In spite of this fact, the TFP of small farms is substantially lower. The TFP of small farms is 37% lower for cereals and 47% lower for fodder compared to large farms (Figure-22). This difference in TFP is much smaller for fruits and vegetables (4% and 11%, respectively). We observe that small farms are relatively more productive in producing fruits and vegetables but nonetheless rely on cereal and fodder production to cover the household's own human and animal consumption requirements. Similarly, a large portion of small farms produce exclusively in one of the two seasons (spring or autumn). Only 84.8 percent of small farms produce in both spring and autumn seasons, compared to 91.5 and 95.1 percent of medium and large farms, respectively (Table 8). Small farms can therefore enhance their farms and family labor productivity by producing perennial labor-intensive crops, such as fruits instead of cereals.

6.4 Robustness Checks

We cross check the results for robustness in two ways: by re-estimating the model with corrected output prices, and accounting for missing land-rental cost. First, we correct for output price differential between the small and the large farms by computing revised output based on the wholesale price of 40 kg bag. Second, we incorporate the effect of missing land-rents by imputing land rents. The results for the yield measure of productivity are robust using these revised prices. The negative association between farm size and yield found with MPL remains negative and significant with similar coefficient values, -.071 for price-corrected and -.064 for imputed-rent estimates (Table 9). Similarly, the insignificant association found with MWR remains as before, with similar coefficient values of -0.01 for price-corrected and 0.006 for imputed-rent estimates. The revised estimates for TFP measures $\hat{T}i$ and T_i are likewise robust in sign and statistical significance, the positive association between farm size and TFP becomes insignificant or even negative when family labor is measured at MPL instead of MWR (Table 10).

7 Conclusion and Recommendations

About 94% of agriculture farms in Pakistan are 12.5 acre or less in size (PSLM-HIES-2018-19). However, large farms (those 12.5 acre or larger) cover 31% of the country's cultivated land. Between 2000 and 2018, average productivity of Pakistani farms grew by 26% while the country's population grew by twice as much, about 57% (PBS, 2018). This underscores the importance of increasing agriculture productivity in improving the food security of Pakistan's majority rural population. In this study, we revisited the inverse farm size and productivity relationship phenomenon, and determined the implications of family labor intensity and labor market distortions, induced by the unwillingness of family labor to work on away farms due to cultural barriers, on the productivity of Pakistani farms. We drew a theoretical model to argue that small farms have lower total factor productivity compared to large farms, even though their yield may be higher. We empirically tested our hypotheses on data from 5,645 agriculture farms in Pakistan. The results of the empirical exercise substantiate the theoretical assertions: first, the higher productive and technical efficiency of small farms and family labor, in terms of yield, compared to that of large farms and hired labor either decreases or turns insignificant when family labor is valued at market wage rates instead of the marginal product. Farm yield decreases by 0.07% with a one percent increase in farm size. However, this negative relationship turns insignificant (-0.02) or even positive when family labor cost is measured at market wage rate rather than the marginal product of labor, with the yield increasing by 0.037% with a 1% increase in farm size. Second, the relatively higher overall and technical efficiency of large farms in terms of TFP becomes insignificant when family labor is valued at marginal product instead of market wage rate. The TFP increases by 0.207% with a one percent increase in land size. The relationship becomes insignificant when family labor cost is measured in terms of marginal product. Third, the divergent behavior of small and large farms in terms of both productivity measures (yield and TFP) depends entirely on whether family labor is accounted for in terms of its marginal product or market wage rate. An increase in family labor employment relative to hired labor increases the farm yield and TFP when family labor cost is computed in terms of the marginal product but decreases the two measures of productivity when it is measured at market rate. This suggests that family labor is the key to understanding the nature and strength of the farm size – productivity relationship. We found higher family labor intensity, labor market distortions and crop selection by the smallholders to be important in this context. Moreover, in spite of the fact that the yield and TFP of small farms are higher for fruits and vegetables, and lower for cereals and fodder, about three-quarter (77%) of small farms still produce cereals and fodder as their principal crop. These findings help better understand the nuances in the farm size – productivity relationship, and challenge the inverse association inferred in the literature based on yield per acre as the indicator of farm productivity. The results also substantiate our model assertions regarding labor market distortions due to unwillingness of family labor to work on away-farms and the non-separability of production and utility functions in agricultural households. Policy interventions therefore need to be devised to reduce labor market imperfections that hamper small farm productivity. Small farmers need to be made aware of opportunities to enhance their productivity through diversification into more lucrative activities such as livestock farming and by switching to more labor-intensive crops such as fruits, vegetables, and flowers, as well as high-value crops like oilseeds. For this to happen, the government will need to improve the availability of a whole set of complementary facilities, including provision of training and information, access to town markets, cold storage, and food-processing plants. Surplus small farm labor can get absorbed in off-farm agriculture-related activities or agro-industrial sector involved in food processing and packaging. This will have the added benefit of raising farm productivity by increasing the demand for farm produce. Finally, administrative and legal measures need to be taken to prevent on-going land fragmentation. Average size of Pakistani farms has shrunk from 7.6 acres in 2000 to 5.3 in 2018. This is putting an upper limit on farm productivity and profitability.

References

Abay, K. A., Abate, G. T., Barrett, C. B., & Bernard, T. (2019). Correlated non-classical measurement errors, 'Second best' policy inference, and the inverse size-productivity relationship in agriculture. *Journal of Development Economics*, *139*(February), 171–184. Retrieved from https://doi.org/10.1016/j.jdeveco.2019.03.008 doi: 10.1016/j.jdeveco.2019.03.008

Aigner, D. J., Amemiya, T., & Poirier, D. J. (1976). On the Estimation of Production Frontiers : Maximum Likelihood Estimation of the Parameters of a Discontinuous Density Function. *International Economic Review*, *17*(2), 377–396. Retrieved from url:http://www.jstor.org/stable/2525708

Akram-Lodhi, A. H. (2001). *Vietnam's Agriculture: Is there an Inverse Relationship?* (No. 348). Netherlands.

Ali, D. A., & Deininger, K. (2015). Is there a farm size-productivity relationship in African agriculture? Evidence from Rwanda. *Land Economics*, *91*(2), 317–343. doi: 10.3368/le.91.2.317

Assunção, J. J., & Braido, L. H. (2007). Testing household-specific explanations for the inverse productivity relationship. *American Journal of Agricultural Economics*, *89*(4), 980–990. doi: 10.1111/j.1467-8276.2007.01032.x

Bardhan, P. K. (1973). Size, Productivity, and Returns to Scale: An Analysis of Farm-Level Data in Indian Agriculture. *Journal of Political Economy*, *81*(6), 1370–1386. doi: 10.1086/260132

Benjamin, D. (1995). Can unobserved land quality explain the inverse productivity relationship? *Journal of Development Economics*, *46*(1), 51–84.

Berry, R. A., & Cline, W. R. (1980). Agrarian Structure and Productivity in Developing Countries. *Land Economics*, *56*(4), 495. doi: 10.2307/3146229

Bevis, L. E., & Barrett, C. B. (2020). Close to the edge: High productivity at plot peripheries and the inverse size-productivity relationship. *Journal of Development Economics*, *143*, 102377. Retrieved from https://doi.org/10.1016/j.jdeveco.2019.102377 doi: 10.1016/j.jdeveco.2019.102377

Bhagwati, J. N., & Chakravarty, S. (1969). Contributions to Analysis : Indian Economic Survey. *The American Economic Review*, *59*(2), 1–73.

Bhalla, S. S. (1979). Farm Size, Productivity, and Technical Change in Indian Agriculture. In *Agrarian structure and productivity in developing countries* (pp. 141–193). Baltimore: Johns Hopkins University Press.

Bhalla, S. S. (1988a). Does land inequalty matter? Theory and Measurement. *Journal of Development Economics*, *29*(1), 45–62.

Bhalla, S. S. (1988b). Does land quality matter?. Theory and measurement. *Journal of Development Economics*, *29*(1), 45–62. doi: 10.1016/0304-3878(88)90070-3

Bhalla, S. S., & Roy, P. (1988). Mis-Specification in Farm Productivity Analysis: The Role of Land Quality. *Oxford Economic Papers*, *40*, 55–73. Retrieved from https://academic.oup.com/oep/article-abstract/40/1/55/2361268by

Carter, M. R. (1984). Identification of the inverse relationship between farm size and productivity: an empirical analysis of peasant agricultural production. , *36*, 131–145.

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, *2*(6), 429–444. doi: 10.1016/0377-2217(78)90138-8

Deininger, K., Jin, S., Liu, Y., & Singh, S. K. (2018). Can labor-market imperfections explain changes in the inverse farm size-productivity relationship? Longitudinal evidence from rural India. *Land Economics*, *94*(2), 239–258. doi: 10.3368/le.94.2.239

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of Chemical Information and Modeling*, *39*(1), 1–38. Retrieved from http://www.jstor.org/stable/2984875.

Desiere, S., & Jolliffe, D. (2018). Land productivity and plot size: Is measurement error driving the inverse relationship? *Journal of Development Economics*, *130*, 84–98. Retrieved from https://doi.org/10.1016/j.jdeveco.2017.10.002 doi: 10.1016/j.jdeveco.2017.10.002

Drucza, K., & Peveri, V. (2018). Literature on gendered agriculture in Pakistan: Neglect of women's contributions. *Women's Studies International Forum*, *69*(March), 180–189. Retrieved from https://doi.org/10.1016/j.wsif.2018.02.007 doi: 10.1016/j.wsif.2018.02.007

Färe, R., Grosskopf, S., Lovell, C. A., & Pasurka, C. (1989). 'Multilateral productivity comparisons when some outputs are undesirable: A nonparametric approach'. *Review of Economics and Statistics*, *71*(1), 90–98. doi: 10.2307/1928055

Feder, G. (1985). The relation between farm size and farm productivity. The role of family labor, supervision and credit constraints. *Journal of Development Economics*, *18*(2-3), 297–313. doi: 10.1016/0304-3878(85)90059-8

Fischer, G., Nachtergaele, F. O., Prieler, S., Teixeira, E., Tóth, G., van Velthuizen, H., … Wiberg, D. (2008). *Global Agro-ecological Zones (GAEZ v3.0)* (Tech. Rep.).

Foster, A. D., & Rosenzweig, M. R. (2021). *Are There Too Many Farms in the World? Labor-market Transaction Costs, Machine Capacities and Optimal Rarm Size.* Retrieved from http://www.nber.org/papers/w23909%0ANATIONAL

Garzón Delvaux, P. A., Riesgo, L., & Gomez y Paloma, S. (2020). Are small farms more performant than larger ones in developing countries? *Science Advances*, *6*(41), 1–12. doi: 10.1126/sciadv.abb8235

Gautam, M., & Ahmed, M. (2018). Too small to be beautiful? The farm size and productivity relationship in Bangladesh. *Food Policy*, *84*(xxxx), 165–175. Retrieved from https://doi.org/10.1016/j.foodpol.2018.03.013 doi: 10.1016/j.foodpol.2018.03.013

Gollin, D., Lagakos, D., & Waugh, M. E. (2014). The agricultural productivity gap. *Quarterly Journal of Economics*, *129*(2), 939–993. doi: 10.1093/qje/qjt056

Gourlay, S., Kilic, T., & Lobell, D. B. (2019). A new spin on an old debate: Errors in farmer-reported production and their implications for inverse scale - Productivity relationship in Uganda. *Journal of Development Economics*, *141*(August), 1–35. Retrieved from https://doi.org/10.1016/j.jdeveco.2019.102376 doi: 10.1016/j.jdeveco.2019.102376

Grimard, F. (2000). Rural labor markets, household composition, and rainfall in Cote d'Ivoire. *Review of Development Economics*, *4*(1), 70–86. doi: 10.1111/1467-9361.00079

Grosskopf, S. (2003). Some remarks on productivity and its decompositions. *Journal of Productivity Analysis*, *20*(3), 459–474. doi: 10.1023/A:1027364119672

Hartley, M. J. (1978). Estimating Mixtures of Normal Distributions and Switching Regressions: Comment. *Journal of the American Statistical Association*, *73*(364), 738–741. Retrieved from http://www.jstor.org/stable/2286267. doi: 10.4135/9781412961288.n9

Helfand, S. M., & Taylor, M. P. (2020). The inverse relationship between farm size and productivity: Refocusing the debate. *Food Policy*(June 2019), 101977. Retrieved from https://doi.org/10.1016/j.foodpol.2020.101977 doi: 10.1016/j.foodpol.2020.101977

Herdt, R., & Mandac, A. M. (1981). Modern Technology and Economic Efficiency of Philippine Rice Farmers. *Economic Development and Cultural Change*, *29*(2), 375–399. doi: 10.1086/451253

Jacoby, H. G. (1993). Shadow wages and peasant family labour supply: An econometric application to the peruvian sierra. *Review of Economic Studies*, *60*(4), 903–921. doi: 10.2307/2298105

Jha, R., Chitkara, P., & Gupta, S. (2000). Productivity, technical and allocative efficiency and farm size in wheat farming in India: A DEA approach. *Applied Economics Letters*, *7*(1), 1–5. doi: 10.1080/135048500351997

Julien, J. C., Bravo-Ureta, B. E., & Rada, N. E. (2019). Assessing farm performance by size in Malawi, Tanzania, and Uganda. *Food Policy*, *84*(xxxx), 153–164. Retrieved from https://doi.org/10.1016/j.foodpol.2018.03.016 doi: 10.1016/j.foodpol.2018.03.016

Kagin, J., Taylor, J. E., & Yúnez-Naude, A. (2015). Inverse Productivity or Inverse Efficiency? Evidence from Mexico. *Journal of Development Studies*, *52*(3), 396–411. doi: 10.1080/00220388.2015.1041515

Lamb, R. L. (2003). Inverse productivity: Land quality, labor markets, and measurement error. *Journal of Development Economics*, *71*(1), 71–95. doi: 10.1016/S0304-3878(02)00134-7

Larson, D. F., Otsuka, K., Matsumoto, T., & Kilic, T. (2014). Should African Rural Development Strategies depend on smallholder farms? An exploration of the inverse-productivity hypothesis. *Agricultural Economics*, *45*(3), 355–367. doi: 10.1111/agec.12070

Lipton, M. (2005). *in a Globalizing World The Role of Crop Science in Alleviating Poverty* (No. June 2005).

Long, J. B., Plosser, & I., C. (1983). Real business cycles. *Jounal of Political Economy*, *91*(I), 39–69. doi: 10.4324/9780203443965.ch18

Lopez, R. E. (1984). Estimating labor supply and production decisions of self-employed farm producers. *European Economic Review*, *24*(1), 61–82. doi: 10.1016/0014-2921(84)90013-8

Lowder, S. K., Skoet, J., & Raney, T. (2016). The Number, Size, and Distribution of Farms, Smallholder Farms, and Family Farms Worldwide. *World Development*, *87*, 16–29. Retrieved from http://dx.doi.org/10.1016/j.worlddev.2015.10.041 doi: 10.1016/j.worlddev.2015.10.041

Lyon, S. M. (2005). Culture and information: An anthropological examination of communication in cultural domains in Pakistan. *Cybernetics and Systems*, *36*(8), 919–932. doi: 10.1080/01969720500306410

Lyon, S. M. (2012). Durham Research on-line. In *Consceptual models of nature in pakistan, in cultural models of nature and the environment: self, space and causality.* (pp. 29–34). Illinois, USA: Institute for the Study of the Environment, Sustainability, and Energy.

Macdonald, J. M., & Mcbride, W. D. (2009). The Transformation of U . S . Livestock Agriculture Scale , Efficiency , and Risks Cataloging Record :. *Transformation*(43).

Mazumdar, D. (1965). Size of Farm and Productivity: A Problem of Indian Peasant Agriculture. *Economica, New Series*, *32*(126), 161–173. Retrieved from http://www.jstor.org/stable/2552546

Mechri, A., Lys, P., & Cachia, F. (2017). *Productivity and Efficiency Measurement in Agriculture* (Tech. Rep. No. February). Food and Agriculture Organization of the United Nations. Retrieved from http://gsars.org/wp-content/uploads/2017/02/TR-17.02.2017-Productivity-and-Efficiency-Measurement-in-Agriculture.pdf

Meeusen, W., & van Den Broeck, J. (1977). Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error. *International Economic Review*, *18*(2), 435–444. doi: 10.2307/2525757

Michael Lipton. (2009). *Land Reform in Developing Countries: Property Rights and Property Wrongs.* London: Routledge.

Morten, M. (2019). Temporary migration and endogenous risk sharing in village India. *Journal of Political Economy*, *127*(1), 1–46. doi: 10.1086/700763

Nishimizu, M., & Page, J. M. (1982). Total factor productivity growth, technological progress and technical efficiency change: dimensions of productivity change in Yugoslavia, 1965-78. *Economic Journal*, *92*(368), 920–936. doi: 10.2307/2232675

Nunn, N., & Puga, D. (2012). Ruggedness: The blessing of bad geography in Africa. *Review of Economics and Statistics*, *94*(1), 20–36. doi: 10.1162/REST_a_00161

O'Donnell, C. J. (2012). An aggregate quantity framework for measuring and decomposing productivity change. *Journal of Productivity Analysis*, *38*(3), 255–272. doi: 10.1007/s11123-012-0275-1

Omotilewa, O. J., Jayne, T. S., Muyanga, M., Aromolaran, A. B., Liverpool-Tasie, L. S. O., & Awokuse, T. (2021). A revisit of farm size and productivity: Empirical evidence from a wide range of farm sizes in Nigeria. *World Development*, *146*, 105592. Retrieved from https://doi.org/10.1016/j.worlddev.2021.105592 doi: 10.1016/j.worlddev.2021.105592

Otsuka, K., Liu, Y., & Yamauchi, F. (2016). The future of small farms in Asia. *Development Policy Review*, *34*(3), 441–461. doi: 10.1111/dpr.12159

Otsuka, K., Liu, Y., Yamauchii, & Futoshi. (2013). Factor endowments, wage growth, and changing food self-sufficiency: Evidence from Country-Level Panel Data. *American Journal of Agricultural Economics*, *95*(5), 1252–8. Retrieved from http://ebrary.ifpri.org/cdm/ref/collection/p15738coll2/id/127394

Paul, C. M., Nehring, R., Banker, D., & Somwaru, A. (2004). Scale economies and efficiency in U.S. agriculture: Are traditional farms history? *Journal of Productivity Analysis*, *22*(3), 185–205. doi: 10.1007/s11123-004-7573-1

Rada, N. E., & Fuglie, K. O. (2019). New perspectives on farm size and productivity. *Food Policy*, *84*(xxxx), 147–152. Retrieved from http://dx.doi.org/10.1016/j.foodpol.2018.03.015 doi: 10.1016/j.foodpol.2018.03.015

Reuter, H. I., Nelson, A., & Jarvis, A. (2007). An evaluation of void-filling interpolation methods for SRTM data. *International Journal of Geographical Information Science*, *21*(9), 983–1008. doi: 10.1080/13658810601169899

Rosset, P. (2000). The Multiple Functions and Benefits of Small Farm Agriculture in the Context of Global Trade Negotiations 1. *Development*, *43*(2), 77–82.

Sanchez, F. A., Restuccia, D., Rud, J. P., Aragon, F., Restuccia, D., & Rud, J. P. (2020). *Department of Economics Are Small Farms Really more Productive than Large Farms ?*

Schmidt, P. (1976). On the Statistical Estimation of Parametric Frontier Production Functions. *The Review of Economics and Statistics*, *58*(2), 238–239. doi: 10.2307/1924032

Sen, A. (1962). An aspect of Indian agriculture. *Economic Weekly*, *14*(4-6), 243–246.

Sen, A. (1975). Employment, technology, and development. *Oxford University Press London*.

Sheng, Y., & Chancellor, W. (2019). Exploring the relationship between farm size and productivity: Evidence from the Australian grains industry. *Food Policy*, *84*(xxxx), 196–204. Retrieved from https://doi.org/10.1016/j.foodpol.2018.03.012 doi: 10.1016/j.foodpol.2018.03.012

Skoufias, E. (1994). Using Shadow Wages to Estimate Labor Supply of Agricultural Households. *American Journal of Agricultural Economics*, *76*(2), 215–227. doi: 10.2307/1243623

Srinivasan, T. N. (1972). Farm Size and Productivity Implications of Choice under Uncertainty. *The Indian Journal of Statistics*, *34*(4), 409–420. Retrieved from http://www.jstor.org/stable/25051809

Subal C. Kumbhakar and C. A. Knox Lovell. (2012). *Stochastic Frontier Analysis* (2nd ed.). Cambridge University Press. Retrieved from https://books.google.es/books?hl=es&lr=&id=wrKDztxLWZ8C&oi=fnd&pg=PR9&dq=lovell+2000+stochastic+frontier+analysis&ots=L2OzyZGJ52&sig=lbI0xLQldj4whzRRFZ0r6cAWSQY#v=onepage&q=lovell2000stochasticfrontieranalysis&f=false

Vakis, R., Sadoulet, E., de Janvry, A., & Al, E. (2004). *Testing for Separability in Household Models with Heterogeneous Behavior: A Mixture Model Approach.* California. Retrieved from https://escholarship.org/uc/item/4hs3g5dj

Vollrath, D. (2007). Land distribution and international agricultural productivity. *American Journal of Agricultural Economics*, *89*(1), 202–216. doi: 10.1111/j.1467-8276.2007.00973.x

Yotopoulos, P. A., & Lau, L. J. (1973). A Test for reletive economic efficiency: Some further results. *The American Economic Review*, *63*(1), 214–223.

Zhao, S., Sheng, Y., & Gray, E. M. (2012). Measuring Productivity of the Australian Broadacre and Dairy Industries: Concepts, Methodology and Data. In *Productivity growth in agriculture* (pp. 73–107). Australian Bureau of Agricultural and Resource Economics and Sciences (ABARES).

List of Tables

Table 1: Summary Statistics

Variable	Observations	Mean	Standard. Dev.	Min	Max
Yield Per Acre (PKR.)	5561	170853	177121	1680	3650833
TFP (Ratio)	5561	1.9	1.3	.022	22.8
Farmland (Acres)	5561	5.3	8.8	.06	375
Controls					
Farm Characteristics					
Input cost per acre (PKR.)	5561	125089	144823	666.8	2006833
Distance to Market (Ratio)	5561	.01	.021	0	.444
Land Irrigation (Y/N)	5561	.79	.401	0	1
Crop Choice (Categorical)	5560	1.9	1.2	1	6
Household Characteristics					
Land Ownership (Y=1/N=2)	5561	1.27	.44	1	2
Education Grade (Years)	5561	3.9	4.5	0	20
family Labor per acre	5561	.332	.727	0	16.6
Family Size	5561	7.4	3.64	1	56
Geographical Characteristics					
Elevation (Meters)	5557	396	625	4	4350
Root water (Ratio)	5561	.194	.049	.098	.305
Ruggedness (Index)	5557	14.8	30.3	1.9	157.3
distance from water (KM)	5561	46.4	39.1	.678	313.9
Soil Quality (Rank 1-7)	5561	1.447	.75	1	4

Note:'Land Irrigation' is a binary variable that acquires value 1 if Farm is irrigated, and 0 otherwise.

Source: PSLM/HIES-2018-19, HWSD, NASA Earth data.

Table 2: Yield per acre and farm size relationship (OLS estimates)

Variables	MPL					Market			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Farm Size	-.37***	-.1***	-.07***	-.073***	-.043***	-.005	-.012	-.011	.037***
	(.016)	(.01)	(.008)	(.008)	(.01)	(.013)	(.012)	(.012)	(.014)
Farm size (RTS adj)	-0.41***	-0.11***	-0.077***	-0.08***	-0.047***	-0.0052	-0.0125	-0.0114	0.038***
	(.016)	(.01)	(.008)	(.008)	(.01)	(.013)	(.012)	(.012)	(.014)
Family worker per acre		.019***	.009***	.008***	.029***	-.026***	-.029***	-.03***	.003
		(.003)	(.003)	(.003)	(.004)	(.004)	(.004)	(.004)	(.007)
Farm size*family labor					-.019***				-.03***
					(.003)				(.004)
Input cost per acre		.715***	.788***	.781***	.777***	.615***	.628***	.602***	.6***
		(.014)	(.014)	(.014)	(.014)	(.016)	(.016)	(.016)	(.016)
Distance from Water			-.056***	-.053***	-.052***		-.061***	-.055***	-.054***
			(.011)	(.011)	(.011)		(.015)	(.014)	(.014)
Distance to Market			-1.99***	-2.23***	-2.26***		-1.28***	-1.38***	-1.46***
			(.307)	(.284)	(.283)		(.463)	(.443)	(.44)
Farm Irrigation (Y/N)									
Irrigated land			.104***	.096***	.099***		.249***	.224***	.226***
			(.027)	(.027)	(.026)		(.041)	(.039)	(.038)
Land ownership (Y/N)									
Land Rented			.243***	.247***	.25***		.139***	.145***	.152***
			(.018)	(.017)	(.017)		(.02)	(.02)	(.02)
Education Grade			.003**	.002*	.002		.006***	.005***	.004**
			(.001)	(.001)	(.001)		(.002)	(.002)	(.002)
Family size			.003*	.003	.004**		0	0	.003
			(.002)	(.002)	(.002)		(.003)	(.003)	(.003)
Elevation			-.079***	-.083***	-.083***		-.022	-.033**	-.033**
			(.014)	(.013)	(.012)		(.016)	(.015)	(.015)
Root Water 0-10cm			-.052	.367	.449		-1.38***	-.595	-.457
			(.331)	(.317)	(.315)		(.401)	(.387)	(.381)
Ruggedness			.127***	.114***	.109***		.072***	.059**	.052**
			(.018)	(.018)	(.017)		(.025)	(.024)	(.023)
Soil Quality (Ref. Soil rank1)									
Soil rank2			.133***	.13***	.135***		.023	.034	.043
			(.028)	(.027)	(.027)		(.034)	(.032)	(.031)
Soil rank3			-.073	-.09*	-.082		-.149**	-.154**	-.141**
			(.057)	(.054)	(.054)		(.076)	(.067)	(.065)
Soil rank4			-.019	-.007	0		-.151***	-.103*	-.091*
			(.034)	(.033)	(.032)		(.055)	(.054)	(.052)
Crop Choice (ref: Cereals)									
Cash crops				.102***	.1***			.17***	.167***
				(.018)	(.018)			(.021)	(.021)
Fodder				.052**	.046**			.212***	.201***
				(.02)	(.02)			(.028)	(.027)
Vegetables & Pulses				.077**	.08**			.009	.015
				(.035)	(.034)			(.053)	(.051)
Fruits				.28***	.274***			.405***	.394***
				(.049)	(.046)			(.062)	(.058)
Other Crops				.168***	.157***			.208***	.191***
				(.045)	(.043)			(.066)	(.063)
Crops by seasons (ref. Single)									
Two seasons				.036*	.031			.044	.037
				(.021)	(.021)			(.029)	(.029)
Constant	12.2***	4.02***	3.43***	3.41***	3.41***	4.83***	4.91***	5.01***	4.96***
	(.024)	(.167)	(.172)	(.174)	(.172)	(.189)	(.206)	(.207)	(.206)
Observations	5559	5557	5553	5552	5552	5557	5553	5552	5552
R-squared	.263	.728	.773	.779	.782	.543	.576	.593	.601
Farm & HH Controls	No	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Geo & Soil Controls	No	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Crop Choice	No	No	No	Yes	Yes	No	No	Yes	Yes

Note: Standard errors are in parentheses. *Denote significant at 10%, ** significant at 5%, and *** significant at 1%. Farm yield is the dependent variable. Market and MPL models show the estimates with family labor measured on market wages and marginal product of labor, respectively. RTS adj shows the farm size – productivity estimates adjusted to the returns to scale parameter $\sigma \neq 1$.

Table 3: TFP and farm size relationship (OLS estimates)

Variables	MPL				Market			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Farm Size	.012	.013*	.002	.027***	.229***	.206***	.207***	.253***
	(.008)	(.007)	(.007)	(.009)	(.012)	(.01)	(.011)	(.015)
Family worker per acre	.012***	.001	.001	.018***	-.067***	-.07***	-.072***	-.042***
	(.004)	(.003)	(.003)	(.004)	(.004)	(.004)	(.004)	(.007)
Farm size*family labor				-.016***				-.029***
				(.003)				(.004)
Distance from Water		-.046***	-.045***	-.044***		-.042***	-.037***	-.035**
		(.011)	(.011)	(.011)		(.015)	(.014)	(.014)
Distance to Market		-2.40***	-2.74***	-2.77***		-1.76***	-2.05***	-2.12***
		(.287)	(.252)	(.252)		(.468)	(.432)	(.435)
Farm Irrigation (Y/N)								
Irrigated land		.026	.027	.028		.162***	.145***	.147***
		(.025)	(.026)	(.025)		(.041)	(.041)	(.04)
Land ownership (Y/N)								
Land Rented		.311***	.316***	.32***		.229***	.24***	.247***
		(.019)	(.019)	(.019)		(.023)	(.022)	(.022)
Education Grade		.003*	.002	.002		.007***	.006***	.005***
		(.001)	(.001)	(.001)		(.002)	(.002)	(.002)
Family size		.001	0	.002		-.007**	-.007***	-.005*
		(.002)	(.002)	(.002)		(.003)	(.003)	(.003)
Elevation		-.118***	-.118***	-.119***		-.073***	-.081***	-.082***
		(.015)	(.014)	(.014)		(.017)	(.016)	(.016)
Root Water 0-10cm		.513	.766**	.838**		-.926**	-.292	-.159
		(.376)	(.357)	(.356)		(.414)	(.398)	(.394)
Ruggedness		.155***	.141***	.138***		.102***	.087***	.081***
		(.018)	(.018)	(.017)		(.025)	(.025)	(.024)
Soil Quality (Ref. Soil rank1)								
Soil rank2		.215***	.203***	.208***		.139***	.141***	.15***
		(.031)	(.031)	(.031)		(.036)	(.035)	(.034)
Soil rank3		-.039	-.065	-.058		-.119	-.137*	-.124
		(.06)	(.059)	(.059)		(.08)	(.076)	(.076)
Soil rank4		.061*	.052	.058*		-.053	-.028	-.017
		(.036)	(.033)	(.033)		(.05)	(.051)	(.05)
Crop Choice (ref: Cereals)								
Cash crops			.07***	.068***			.14***	.137***
			(.021)	(.021)			(.024)	(.024)
Fodder			-.045**	-.052**			.089***	.077***
			(.021)	(.021)			(.028)	(.027)
Vegetables & Pulses			.127***	.131***			.083	.089*
			(.035)	(.034)			(.05)	(.05)
Fruits			.239***	.233***			.391***	.38***

continued

Table 3: TFP and farm size relationship (OLS estimates)

Variables	MPL				Market			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
			(.054)	(.052)			(.063)	(.058)
Other Crops			.149***	.14***			.189***	.173***
			(.043)	(.042)			(.064)	(.061)
Crops by seasons (ref. Single)								
Two seasons			.043*	.039*			.065**	.058**
			(.023)	(.023)			(.029)	(.029)
Constant	.803***	1.106***	1.051***	1.017***	.297***	.675***	.512***	.451***
	(.014)	(.087)	(.089)	(.089)	(.023)	(.114)	(.116)	(.116)
Observations	5557	5553	5552	5552	5557	5553	5552	5552
R-squared	.005	.256	.274	.28	.18	.261	.276	.286
Farm & HH Controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Geo & Soil Controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Crop Choice	No	No	Yes	Yes	No	No	Yes	Yes

Note: Standard errors are in parentheses. *Denote significant at 10%, ** significant at 5%, and *** significant at 1%. Farm TFP ($\hat{T}_i or T_i$) is the dependent variable. Market and MPL models show the estimates with family labor measured on market wages and marginal product of labor, respectively.

Table 4: Farm size and technical efficiency estimates - Stochastic Frontier Analysis

Variables	MPL				Market			
	(1) Yield	(2) Yield-P	(3) TFP	(4) TFP-P	(5) Yield	(6) Yield-P	(7) TFP	(8) TFP-P
Input cost per acre	.713***	.723***			.606***	.61***		
	(.007)	(.007)			(.008)	(.009)		
Distance from Water	-.061***	-.063***	-.046***	-.022*	-.065***	-.068***	-.036***	-.039***
	(.006)	(.007)	(.007)	(.012)	(.008)	(.008)	(.009)	(.01)
Distance to Market	-3.11***	-3.762***	-3.796***	1.111**	-1.953***	-2.256***	-1.955***	-2.362***
	(.25)	(.256)	(.287)	(.435)	(.292)	(.295)	(.359)	(.363)
Farm Irrigation (Y/N)								
Irrigated land	.078***	.062***	-.009	.221***	.197***	.184***	.148***	.135***
	(.014)	(.015)	(.017)	(.028)	(.019)	(.019)	(.022)	(.023)
Soil Quality (Ref. Soil rank1)								
Soil rank2	.099***	.11***	.189***	-.054**	.013	.022	.114***	.125***
	(.014)	(.015)	(.016)	(.025)	(.018)	(.018)	(.021)	(.021)
Soil rank3	-.045	-.045	.007	-.064	-.118***	-.115***	-.094**	-.088*
	(.029)	(.03)	(.034)	(.054)	(.037)	(.038)	(.044)	(.045)
Soil rank4	.002	.006	.076***	-.105**	-.066**	-.055*	-.024	-.016
	(.026)	(.027)	(.029)	(.046)	(.032)	(.033)	(.038)	(.038)
Crop Choice (ref: Cereals)								
Cash crops	.107***	.065***	.036**	.135***	.173***	.147***	.12***	.089***
	(.014)	(.014)	(.016)	(.024)	(.017)	(.018)	(.02)	(.02)
Fodder	.064***	.058***	-.096***	-.003	.17***	.162***	-.046**	-.055**
	(.014)	(.014)	(.016)	(.027)	(.018)	(.019)	(.021)	(.022)
Vegetables & Pulses	.085***	.113***	.151***	-.016	.031	.058*	.077**	.104***
	(.024)	(.025)	(.028)	(.045)	(.031)	(.032)	(.037)	(.038)
Fruits	.23***	.248***	.131***	.298***	.322***	.343***	.247***	.269***
		(.03)	(.033)	(.055)	(.038)	(.039)	(.046)	(.047)
Other Crops	.163***	.181***	.09**	.113*	.204***	.217***	.143***	.158***
	(.032)	(.034)	(.037)	(.06)	(.042)	(.043)	(.051)	(.052)
Land ownership (Y/N)								
Land Rented	.254***	.254***	.371***	.042*	.133***	.135***	.256***	.258***
	(.012)	(.013)	(.014)	(.022)	(.015)	(.016)	(.018)	(.018)
Education Grade	.002	.002**	.003**	.01***	.006***	.007***	.013***	.013***
		(.001)	(.001)	(.002)	(.001)	(.001)	(.002)	(.002)
Family size	.001	0	-.001	.03***	-.007***	-.007***	-.014***	-.014***
	(.001)	(.001)	(.002)	(.002)	(.002)	(.002)	(.002)	(.002)
Elevation	-.09***	-.099***	-.133***	.029**	-.033***	-.042***	-.067***	-.078***
	(.007)	(.008)	(.008)	(.012)	(.009)	(.009)	(.011)	(.011)
Root Water 0-10cm	.398**	.2	.721***	-.995***	-.196	-.311	-.641**	-.78***
	(.192)	(.204)	(.223)	(.366)	(.248)	(.255)	(.296)	(.303)
Land slope	-.977***	-1.06***	-.412***	-.843***	-1.53***	-1.55***	-1.01***	-1.04***
	(.117)	(.123)	(.136)	(.234)	(.156)	(.16)	(.188)	(.191)
Land slope square	.04***	.043***	.02***	.027***	.061***	.062***	.041***	.042***
	(.005)	(.005)	(.005)	(.009)	(.006)	(.006)	(.007)	(.008)
Ruggedness .003***		.003***	.003***	.001**	0	0	0	0
	(0)	(0)	(0)	(.001)	(0)	(0)	(.001)	(.001)
Constant	10.43***	11.06***	3.69***	16.62***	15.03***	15.25***	7.31***	7.67***
	(.759)	(.8)	(.87)	(1.496)	(1.009)	(1.036)	(1.201)	(1.224)
Technical efficiency								
lnsig2v: Constant	-2.17***	-2.25***	-1.84***	-1.13***	-2.29***	-2.32***	-1.71***	-1.73***
	(.029)	(.046)	(.025)	(.023)	(.046)	(.047)	(.039)	(.04)
lnsig2u: Large farms	2.21**	.582***	-.378	-35.4	-.704***	-.686***	-1.67***	-1.57***
	(1.002)	(.107)	(.283)	(805.5)	(.063)	(.06)	(.091)	(.082)
lnsig2u:Family labor (Y/N)								
lnsig2u:Family labor (yes)	-1.22***	-.634***	-2.01**	-1.03***	.231***	.225***	.691***	.66***
	(.243)	(.102)	(.925)	(.089)	(.055)	(.053)	(.066)	(.063)
lnsig2u: Constant	-5.33***	-2.61***	-3.71***	.423***	-.871***	-.718***	-.668***	-.536***
	(1.291)	(.184)	(.488)	(.048)	(.053)	(.049)	(.061)	(.057)
Observations	5550	5550	5550	5550	5556	5556	5556	5556
Pseudo R2	.z	.z	.z	.z	.z	.z	.z	.z
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: Standard errors are in parentheses. *Denote significant at 10%, ** significant at 5%, and *** significant at 1%. All equations in the table are estimated using MLE. Farm technical efficiency: Yield and TFP are the dependent variables. Yield-P and TFP-P indicate the corresponding price corrected estimates. Market and MPL models show estimates with family labor based on market wages and Marginal Product of Labor, respectively. Large farms technical efficiency is compared with that of small farms (less than 2.5 acres). Farms' family-labor technical efficiency is compared to that of without family labor.

Table 5: Testing Labor Market Imperfections - Productivity of family vs Hired Labor.

	(1) Output	(2) Log of Family Worker Share in Total Labor	(3) Log of No. of Family Worker per Acre
Small Farm*Labor	.003		
	(.004)		
Medium Farm*Labor	.038***		
	(.004)		
Large Farm*Labor	.054***		
	(.004)		
Log of other inputs	.661***		
	(.007)		
Log of input		.587***	.57***
		(.011)	(.011)
Log of Farm Size		.377***	.371***
		(.008)	(.009)
Log of Family Labor cost share		-.023***	
		(.003)	
Log Family labor per acre			-.057**
			(.023)
Constant	4.788***	5.196***	5.374***
	(.083)	(.13)	(.133)
Observations	5553	5559	5561
R-squared	.789	.671	.668

Note: Standard errors are in parentheses. *denote significant at 10%, ** significant at 5%, and *** significant at 1%. All models are estimated using OLS.

Table 6: Testing for Non-separability of Production and Utility Functions – E-M Algorithm

	Variables Type	Constrained (Owned-farm)	Unconstrained (Away- or off-farm)	Switching Regression
Market Wage	Q	2.81e-06 ***	4.75e-07 ***	-3.56E-06***
		(4.56e-07)	(1.59e-07)	(5.13E-07)
Output Produced	Q	5.79e-07 ***	-2.82e-08 **	-8.99E-07***
		(3.61e-08)	(1.32e-08)	(2.48E-08)
All Inputs Cost	Q	-7.23e*10-07 ***	5.29e-08 *	1.42E-06***
		(.0021)	(2.93e-08)	(5.01E-08)
Farm size	Q	.002	.000346	-0.012***
		(.0016)	(.000336)	(0.00052)
Livestock holding	Q	-.001	.075***	0.054*
		(.017)	(.021)	(0.028)
Agri Primary Occupation	Q	.391***	.556***	0.132***
		(.011)	(.018)	(0.021)
Land ownership	Q	.094***	.051***	-0.21***
		(.01)	(.013)	(0.016)
Education	Q	-.019***	-.009***	0.038***
		(.001)	(.002)	(0.0015)
Age	H	.002***	.00021	-0.0038***
		(.00026)	(.00031)	(0.00041)
Regional Preferences				
KPK	H	-.189***	-.039	0.625***
		(.042)	(.033)	(0.041)
Punjab	H	-.288***	.027	0.676***
		(.044)	(.028)	(0.041)
Baluchistan	H	-.042**	.06***	0.214***
		(.018)	(.018)	(0.027)
Rural	H	.075***	.03	-0.168***
		(.015)	(.028)	(0.028)
Ethnic Group Preferences				
Urdu-speaking	H	.023	.046	0.11**
		(.031)	(.035)	(0.051)
Punjabi	H	.116***	.002	-0.14**
		(.034)	(.041)	(0.056)
Sindhi	H	-.304***	-.022	0.73***
		(.05)	(.038)	(0.058)
Pashtun	H	-.068**	.008	0.19***
		(.029)	(.043)	(0.049)
Saraiki	H	.014	-.019	0.065
		(.034)	(.043)	(0.057)
Male/Total labor Ratio	H	-.294***	-.162***	0.52***
		(.027)	(.036)	(0.044)
Constant		.436***	.265***	-1.06

continued

Table 6: Testing for Non-separability of Production and Utility Functions – E-M Algorithm

	Variables Type	Constrained (Owned-farm)	Unconstrained (Away- or off-farm)	Switching Regression
		(.057)	(.055)	(0.074)
Observations		10572	4084	14656
R-squared		.219	.223	.297

Note: Standard errors are in parentheses. *Denote significant at 10%, ** significant at 5%, and *** significant at 1%. Variable type 'Q' denotes the Production variables and 'H' the family preferences. The dependent variable is binary in nature, which take the value of 1 if an individual is working on owned-farm, 0 otherwise. Switching regression divides the sample into two regimes: Constrained and Un-constrained based on the factors mentioned.

Table 7: Farm Yield by Crop Type

Yield	Coefficient.	S.E.	p-value	Significance
Cash Crops	34727	(2092)	0	***
Fodder	11382	(2164)	0	***
Veg & Pulses	20161	(3930)	0	***
Fruits	81296	(4576)	0	***
Other Crops	5542	(5414)	.31	
Per Acre Input Cost	.21	(.008)	0	***
Constant	58441	(1262)	0	***
R-squared	0.188	Observations	5560	

Note: Standard errors are in parentheses. *Denote significant at 10%, ** significant at 5%, and *** significant at 1%. The dependent Variable is Farm Yield. Cereals are taken as the reference (omitted) category. All estimates are based on OLS.

Table 8: Distribution of farms by season and primary crop produced.

Crop Category	Crop Type	% of farms by crop choice			
		Small	Medium	Large	Total
Cereals	Wheat & Rice	45	55	58	51
Cash crops	Cotton & sugar	11	28	21	20
Fodder	Fodder & Maize	32	10	5	20
Vegetables and pulses	Vegetables and pulses	4	4	10	4
Fruits	All Fruits	4	2	3	3
Other crops	All other crops	3	2	3	2
Total		100	100	100	100
Crops raised in two seasons Rabi (spring) and Kharif (winter)					
Farms raised only Rabi (spring) or Kharif (winter) crop		15.2	8.5	4.9	11.3
Farms raised both Rabi (spring) and Kharif (winter) crop		84.8	91.5	95.1	88.7
Total		100	100	100	100

Note: The primary Crop is determined based on the maximum share in output across all crops raised by the farm.
Source: Authors' estimates using PSLM-HIES-2018-19 data.

Table 9: Yield per acre and farm size relationship (OLS estimates)

	MPL		Market	
	(1) Price-corrected	(2) Imputed land-rent	(3) Price-corrected	(4) Imputed land-rent
Farm Size	-.071***	-.064***	-.01	.006
	(.01)	(.009)	(.013)	(.013)
Family worker per acre	.007**	.012***	-.031***	-.025***
	(.003)	(.003)	(.004)	(.004)
Input cost per acre	.792***	1.186***	.607***	.854***
	(.015)	(.022)	(.016)	(.022)
Distance from Water	-.055***	-.061***	-.057***	-.058***
	(.011)	(.012)	(.014)	(.014)
Distance to Market	-2.943***	-2.05***	-2.087***	-1.301***
	(.514)	(.293)	(.684)	(.449)
Farm Irrigation (Y/N)				
Irrigated land	.085***	.151***	.215***	.248***
	(.027)	(.028)	(.039)	(.04)
Land ownership (Y/N)				
Land Rented	.246***	.231***	.142***	.14***
	(.019)	(.017)	(.021)	(.019)
Education Grade	.003**	.002	.006***	.005***
	(.001)	(.001)	(.002)	(.002)
Family size	.003	.003*	0	.001
	(.002)	(.002)	(.003)	(.002)
Elevation	-.101***	-.072***	-.049***	-.025
	(.014)	(.013)	(.017)	(.016)
Root Water 0-10cm	.068	-.198	-.91**	-.888**
	(.362)	(.331)	(.434)	(.392)
Ruggedness	.126***	.105***	.07***	.048**
	(.02)	(.018)	(.026)	(.024)
Soil Quality (Ref. Soil rank1)				
Soil rank2	.14***	.076***	.041	.007
	(.029)	(.028)	(.033)	(.032)
Soil rank3	-.092	-.137**	-.157**	-.175***
	(.058)	(.056)	(.07)	(.066)
Soil rank4	-.003	-.068	-.102*	-.136**
	(.035)	(.041)	(.055)	(.057)
Crop Choice (ref: Cereals)				
Cash crops	.048*	.131***	.117***	.183***
	(.026)	(.018)	(.03)	(.021)
Fodder	.047**	.033	.21***	.194***
	(.021)	(.021)	(.029)	(.028)
Vegetables & Pulses	.102***	-.002	.033	-.042
	(.036)	(.048)	(.055)	(.062)

continued

Table 9: Yield per acre and farm size relationship (OLS estimates)

	MPL		Market	
	(1) Price-corrected	(2) Imputed land-rent	(3) Price-corrected	(4) Imputed land-rent
Fruits	.309***	.269***	.436***	.401***
	(.054)	(.04)	(.066)	(.059)
Other Crops	.186***	.12***	.227***	.17***
	(.045)	(.044)	(.066)	(.064)
Crops by seasons (ref. Single)				
Two seasons	.05**	.095***	.058*	.086***
	(.025)	(.023)	(.033)	(.03)
Constant	3.406***	-1.664***	5.071***	1.776***
	(.179)	(.269)	(.216)	(.284)
Observations	5552	5552	5552	5552
R-squared	.759	.755	.572	.587
Farm & HH Controls	Yes	Yes	Yes	Yes
Geo & Soil Controls	Yes	Yes	Yes	Yes
Crop Choice	Yes	Yes	Yes	Yes

Note: Standard errors are in parentheses. *Denote significant at 10%, ** significant at 5%, and *** significant at 1%. Farm yield is the dependent variable. Market and MPL models show the estimates with family labor measured on market wages and marginal product of labor, respectively.

Table 10: Modified TFP and farm size relationship (OLS estimates)

	MPL		Market	
	(1) Price-corrected	(2) Imputed land-rent	(3) Price-corrected	(4) Imputed land-rent
Farm Size	0	-.108***	.204***	.068***
	(.008)	(.009)	(.012)	(.012)
Family worker per acre	0	.015***	-.073***	-.035***
	(.003)	(.003)	(.004)	(.004)
Distance from Water	-.047***	-.065***	-.039***	-.051***
	(.011)	(.012)	(.014)	(.014)
Distance to Market	-3.432***	-1.789***	-2.739***	-1.393***
	(.448)	(.339)	(.639)	(.431)
Farm Irrigation (Y/N)				
Irrigated land	.019	.181***	.137***	.23***
	(.025)	(.029)	(.041)	(.04)
Land ownership (Y/N)				
Land Rented	.312***	.195***	.236***	.153***
	(.02)	(.016)	(.023)	(.019)
Education Grade	.003*	.002	.007***	.005***
	(.001)	(.001)	(.002)	(.002)
Family size	0	.005**	-.008***	-.001
	(.002)	(.002)	(.003)	(.002)
Elevation	-.135***	-.054***	-.098***	-.039***
	(.015)	(.013)	(.017)	(.015)
Root Water 0-10cm	.447	-.333	-.61	-.746*
	(.395)	(.336)	(.44)	(.387)
Ruggedness	.152***	.091***	.098***	.06**
	(.02)	(.018)	(.026)	(.024)
Soil Quality (Ref. Soil rank1)				
Soil rank2	.208***	.044	.146***	.035
		(.028)	(.036)	(.032)
Soil rank3	-.068	-.144**	-.139*	-.162**
	(.061)	(.057)	(.078)	(.067)
Soil rank4	.052	-.091**	-.027	-.111**
	(.035)	(.044)	(.052)	(.055)
Crop Choice (ref: Cereals)				
Cash crops	.018	.144***	.088***	.181***
		(.018)	(.031)	(.021)
Fodder	-.046**	.09***	.088***	.161***
	(.022)	(.021)	(.028)	(.027)
Vegetables & Pulses	.15***	-.018	.106**	-.022
	(.035)	(.052)	(.051)	(.059)
Fruits	.27***	.294***	.421***	.407***
	(.059)	(.042)	(.067)	(.059)

continued

Table 10: Modified TFP and farm size relationship (OLS estimates)

	MPL		Market	
	(1) Price-corrected	(2) Imputed land-rent	(3) Price-corrected	(4) Imputed land-rent
Other Crops	.167***	.138***	.208***	.178***
	(.043)	(.046)	(.064)	(.063)
Crops by seasons (ref. Single)				
Two seasons	.056**	.082***	.079**	.096***
		(.023)	(.033)	(.029)
Constant	1.166***	.452***	.628***	.041
	(.096)	(.098)	(.122)	(.114)
Observations	5552	5552	5552	5552
R-squared	.268	.212	.267	.146
Farm & HH Controls	Yes	Yes	Yes	Yes
Geo & Soil Controls	Yes	Yes	Yes	Yes
Crop Choice	Yes	Yes	Yes	Yes

Note: Standard errors are in parentheses. *Denote significant at 10%, ** significant at 5%, and *** significant at 1%. Farm TFP ($\hat{T}_i$ or T_i) is the dependent variable. Market and MPL models show the estimates with family labor measured on market wages and marginal product of labor, respectively.

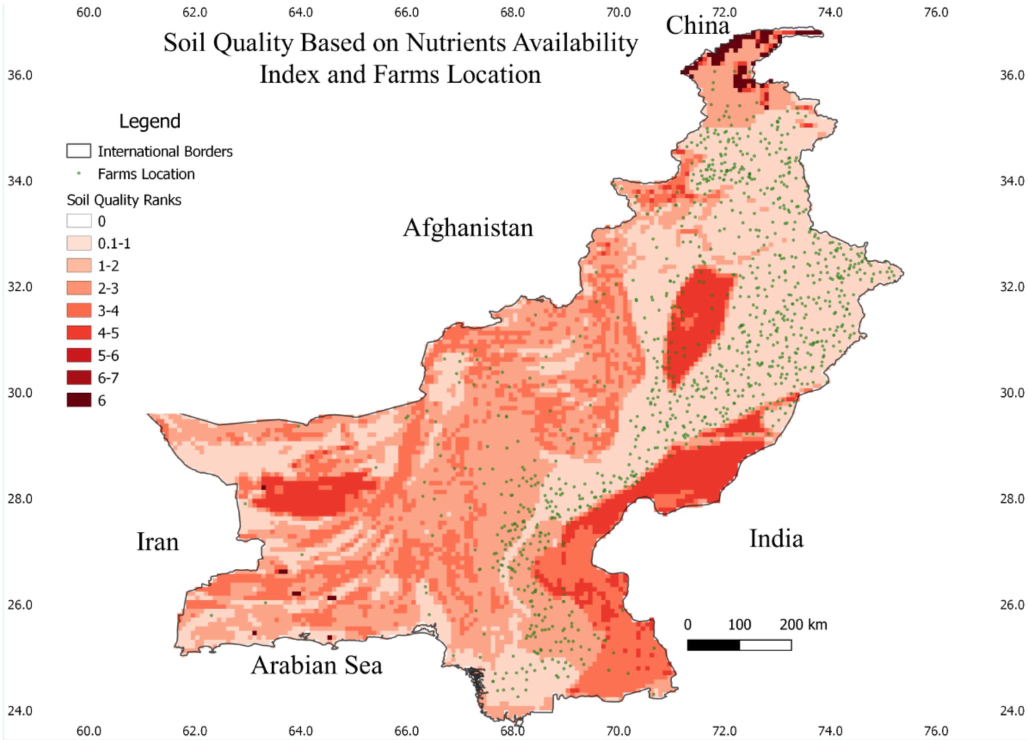
Table A1: Testing for constant return to scale

Log of output	Market	MPL
Log of input	.557***	.702***
	(.01)	(.008)
Log of land rental	.405***	.209***
	(.008)	(.007)
Constant	1.269***	2.053***
	(.113)	(.075)
Observations	5561	5561
R-square	0.664	0.793
RTS σ	0.962	0.911
F-test $\sigma = 1$	16.49	200.4

Note: Standard errors are in parentheses. *Denote significant at 10%, ** significant at 5%, and *** significant at 1%. The dependent variable is farm yield. Market and MPL models show estimates with family labor based on market wages and marginal product of labor, respectively. Estimates are based on OLS.

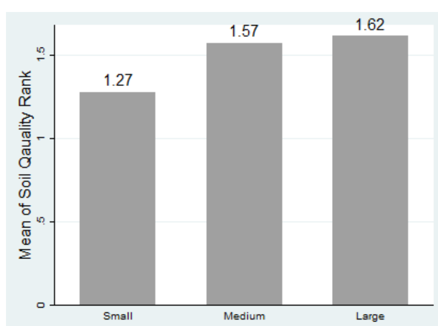
List of Figures

Figure 1: Geographical distribution of farms and corresponding soil quality.



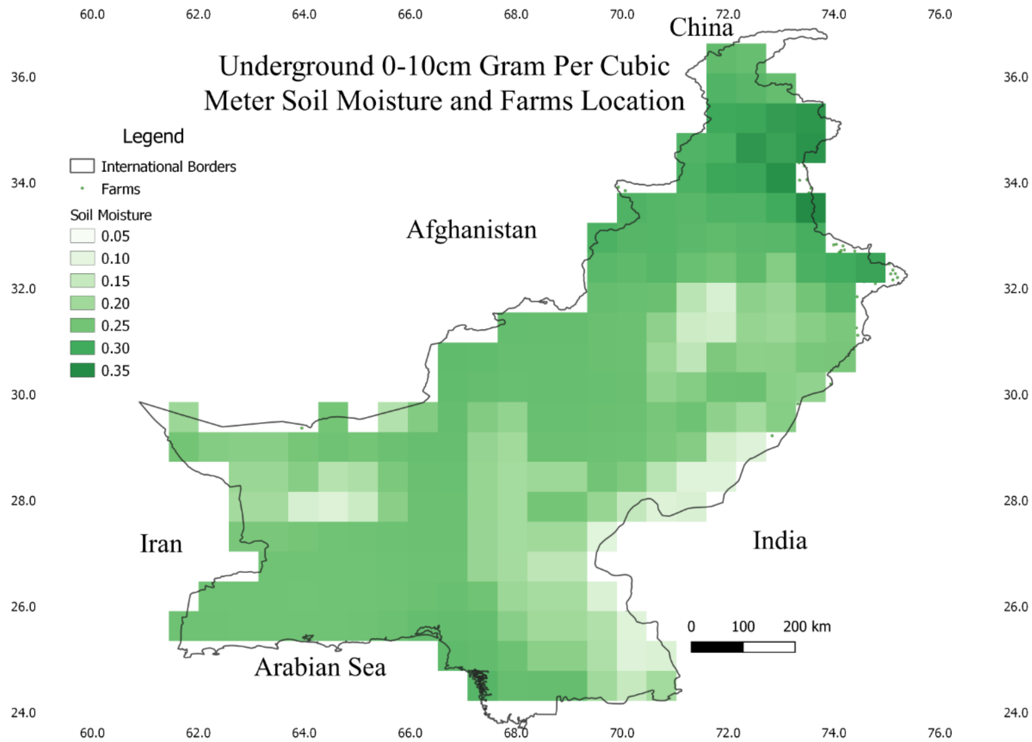
Source: Authors' calculations using PSLM/HIES-2018-19 and Harmonized World Soil Database (HWSD).

Figure 2: Mean soil Quality by farm size.



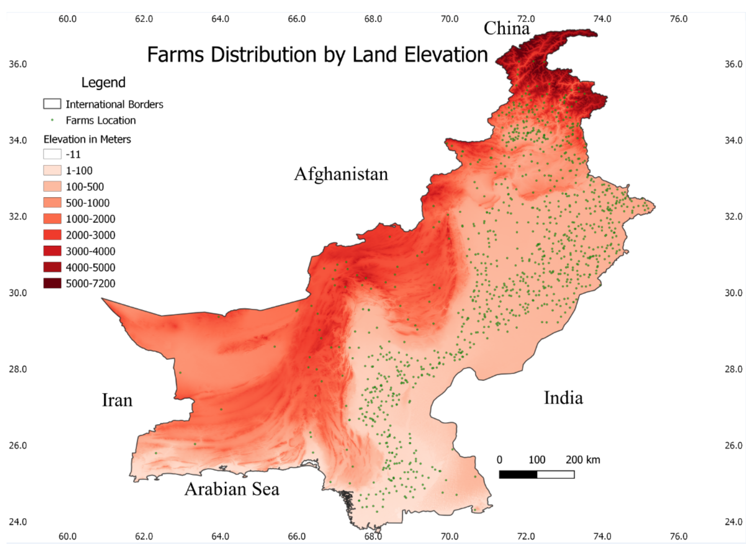
Source: Author estimates from PSLM/HIES-2018-19.

Figure 3: Geographical distribution of farms and corresponding soil moisture.



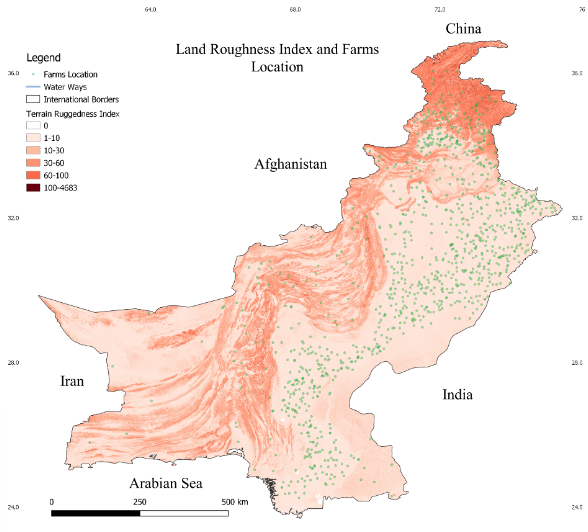
Source: Authors' calculations using PSLM/HIES-2018-19 and NASA (GLDAS-2.1).

Figure 4: Geographical distribution of farms and corresponding land elevation.



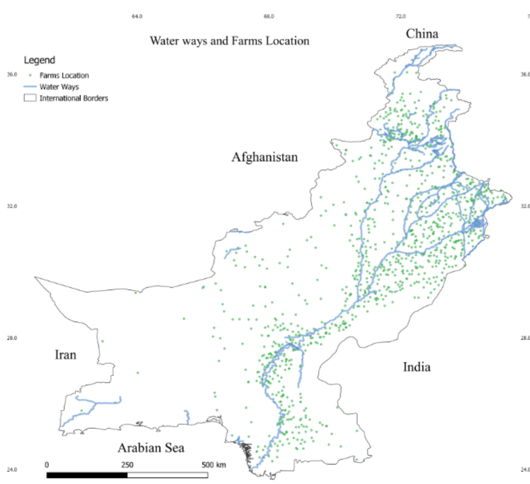
Source: Authors' calculations using PSLM/HIES-2018-19 and USGS/NASA SRTM Version-1.

Figure 5: Geographical distribution of farm land ruggedness.



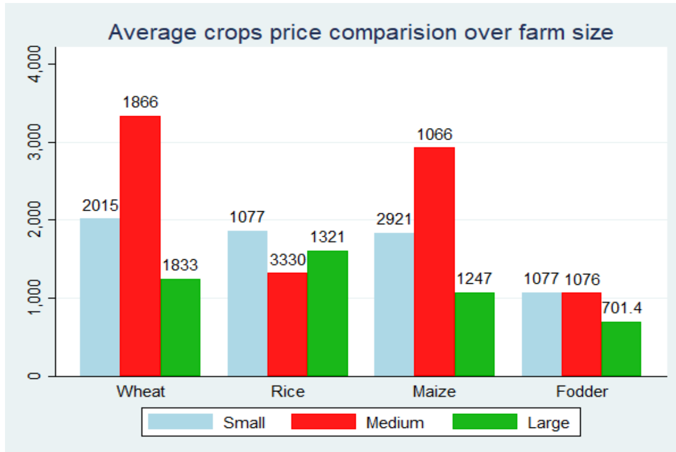
Source: Authors' calculations using PSLM/HIES-2018-19 and USGS/NASA SRTM Version-1.

Figure 6: Farm location by distance from water bodies.



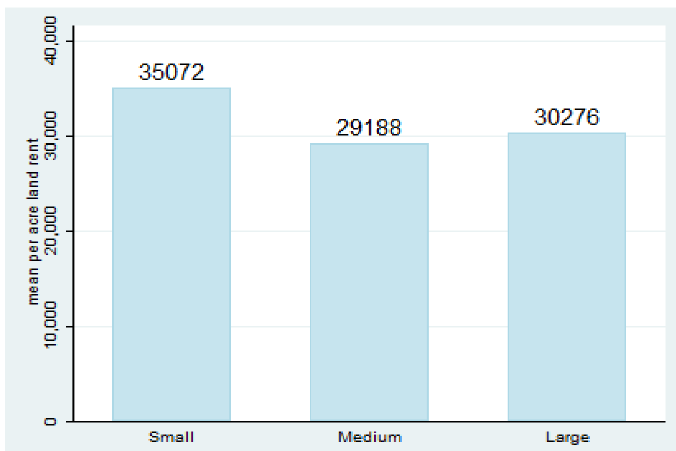
Source: Author estimates from PSLM/HIES-2018-19.

Figure 7: Crop price differential by farm size.



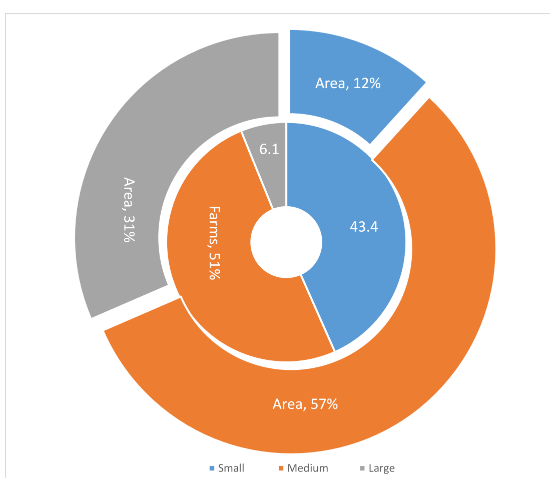
Source: Authors' calculations using PSLM/HIES-2018-19.

Figure 8: Mean land rent by farm size.



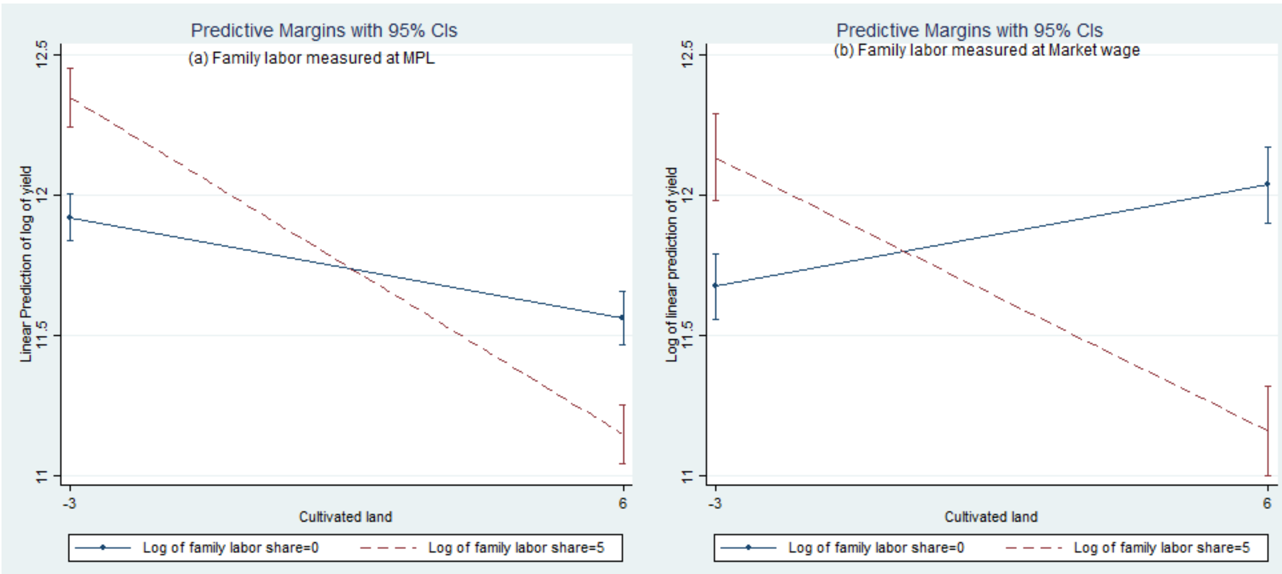
Source: Authors' calculations using PSLM/HIES-2018-19.

Figure 9: Number of farms and corresponding area share by farm size.



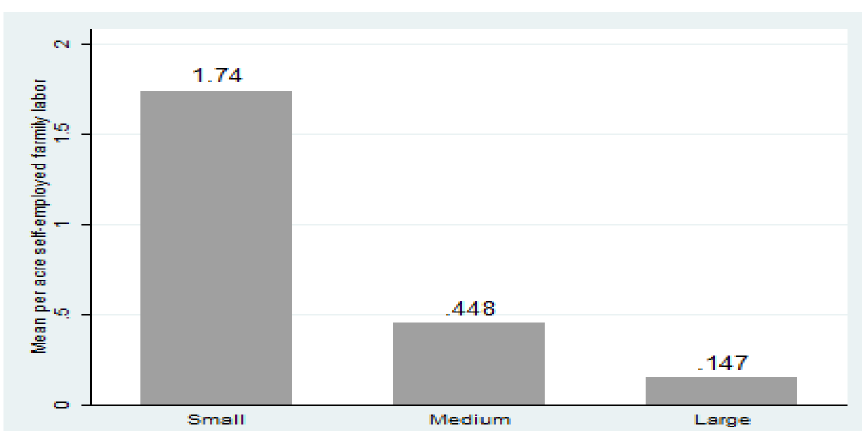
Source: Author estimates from PSLM/HIES-2018-19.

Figure 10: Yield and farm size interaction with family labor.



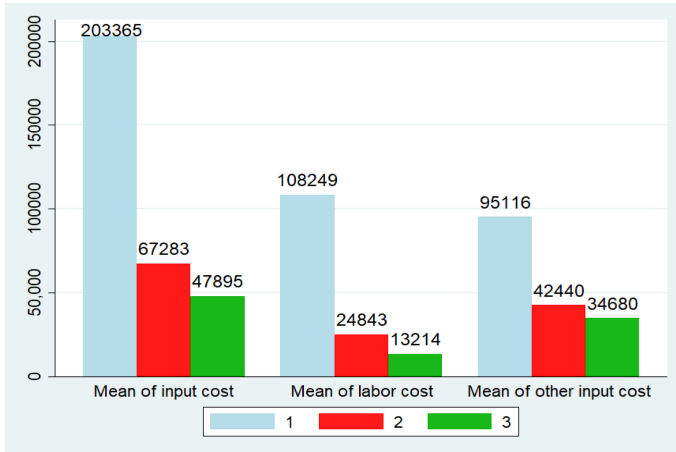
Source: Author estimates from PSLM/HIES-2018-19.

Figure 11: Mean farm family worker per acre.



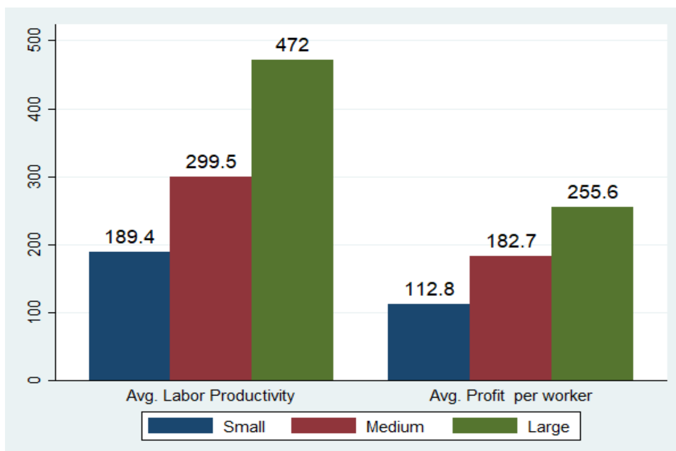
Source: Authors' calculations using PSLM/HIES-2018-19.

Figure 12: Per acre inputs and labor cost by farm size (in PKR)



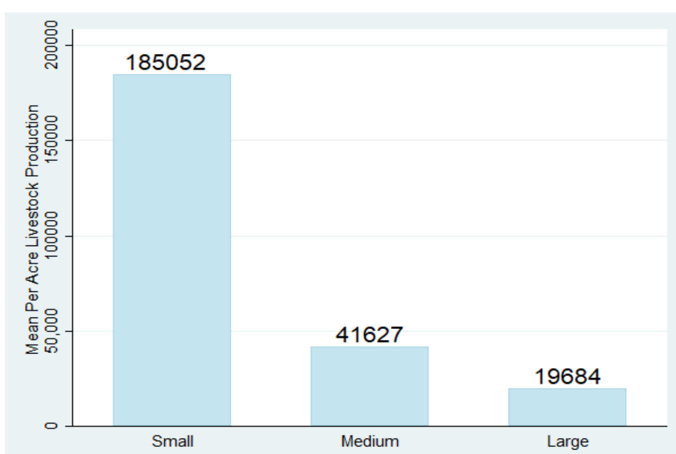
Source: Authors' calculations using PSLM/HIES-2018-19.

Figure 13: Output and profit per worker across farm sizes (in thousands of PKR).



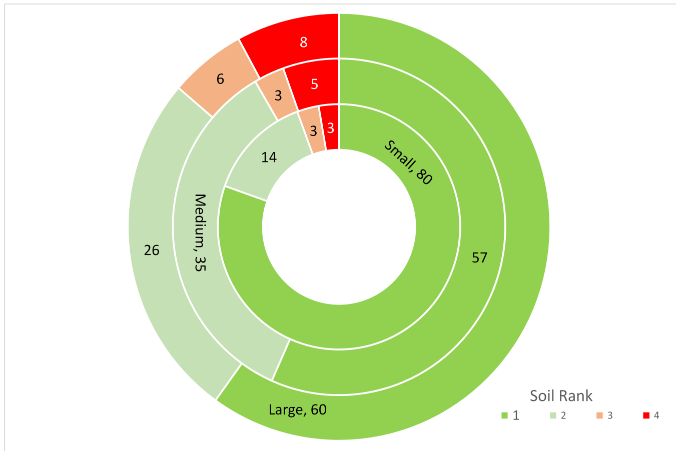
Source: Authors' calculations using PSLM/HIES-2018-19.

Figure 14: Average per acre livestock production by farm size.



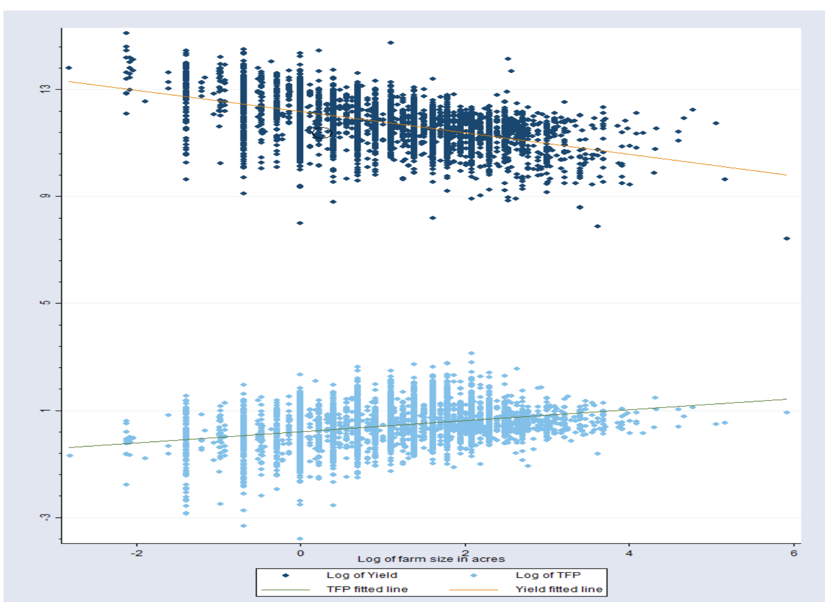
Source: Authors' calculations using PSLM/HIES-2018-19.

Figure 15: Soil quality distribution by farm size.



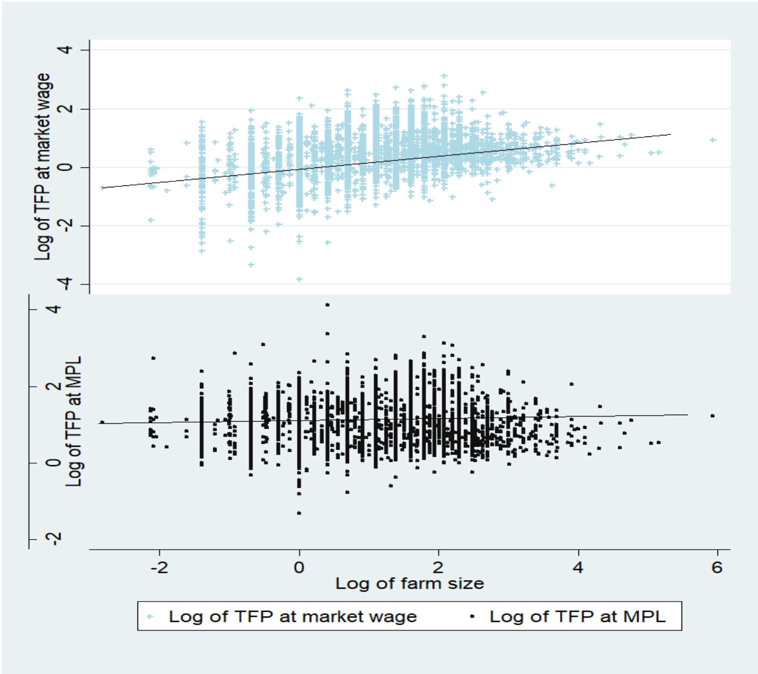
Source: Authors' calculations using PSLM/HIES-2018-19
and Harmonized World Soil Database (HWSD).

Figure 16: Farm yield and TFP at MWR and farm size – linear fit.



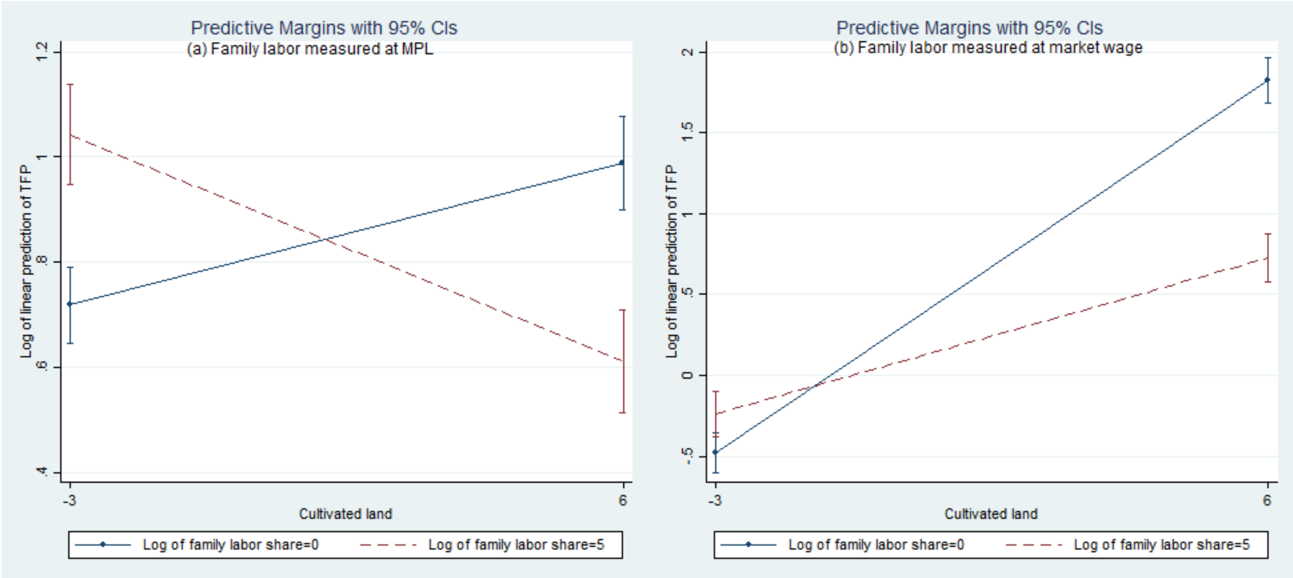
Source: Author estimates from PSLM/HIES-2018-19.

Figure 17: TFP at MPL and market wage and farm size – linear fit



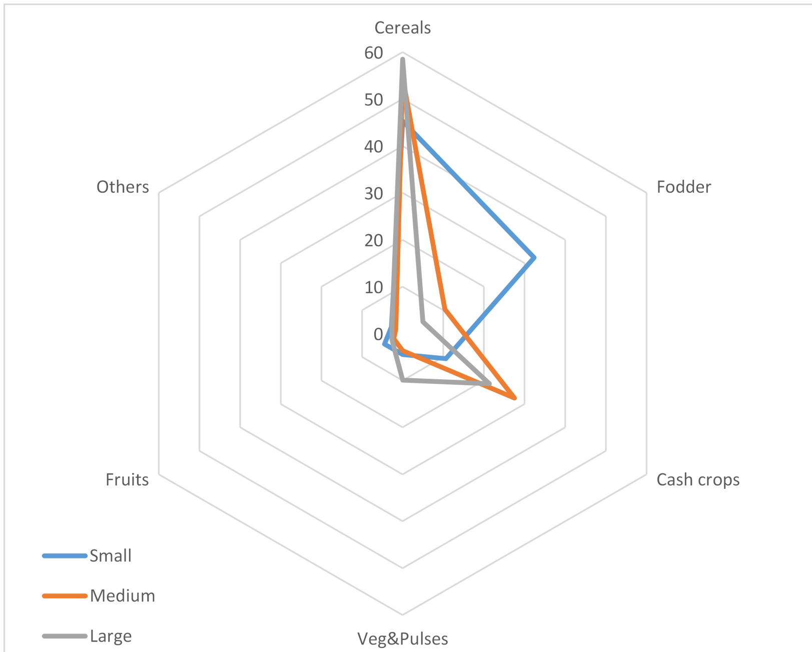
Source: Author estimates from PSLM/HIES-2018-19.

Figure 18: TFP and farm size interaction with family labor.



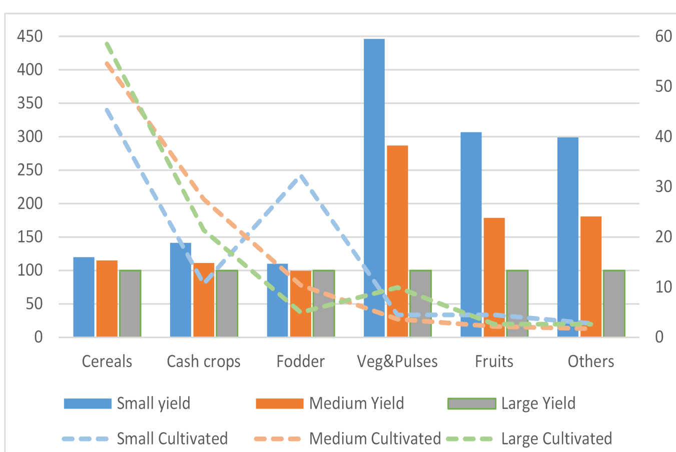
Source: Author estimates from PSLM/HIES-2018-19.

Figure 19: Distribution of primary crops cultivated by farm size.



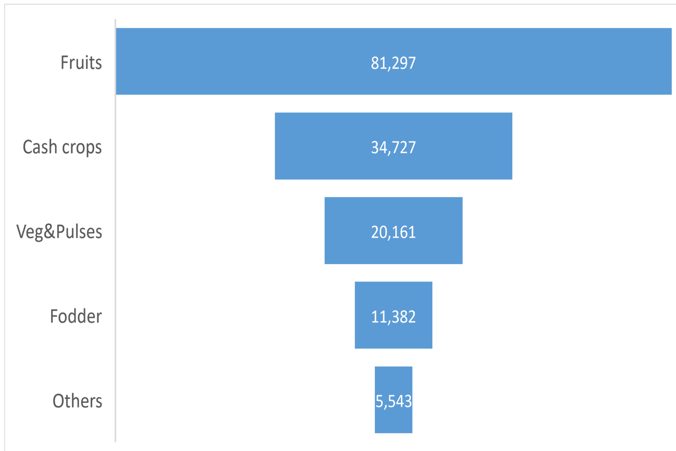
Source: Author estimates from PSLM/HIES-2018-19.

Figure 20: Crops Cultivated and Relative Yield By Farm Size.



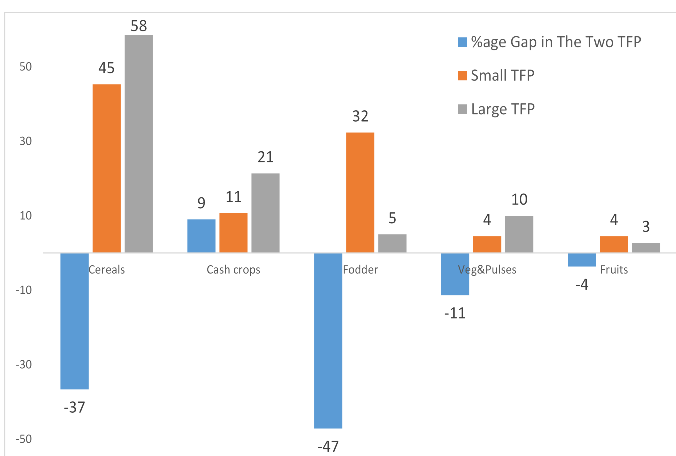
Source: Authors' calculations using PSLM/HIES-2018-19.

Figure 21: Difference in farm yield by crops raised relative to Cereals.



: Authors' calculations using PSLM/HIES-2018-19.

Figure 22: Difference in crop TFP by farm size.



Source: Authors' calculations using PSLM/HIES-2018-19.