

Agricultural Productivity and Farm Size in Latin America

Steven Helfand and Matthew Taylor

5/12/2018

Abstract: The relationship between farm size and productivity has posed a perennial puzzle in development economics. Empirical studies spanning a broad range of contexts have established a stylized fact that the two are inversely related, influencing the debate over land reform in the region and highlighting bottlenecks and opportunities for unleashing productivity gains in the agricultural sectors of developing economies. This report assesses the relationship between farm size and productivity in Latin America by reviewing the existing literature on the farm size – productivity relationship in the region, assessing land productivity data from the Rural Income Generating Activities (RIGA) project for a subset of Latin American countries, synthesizing recent research on the farm size – productivity relationship in Brazil, and estimating total factor productivity and its components with a Mexican panel of agricultural households using data from the Mexican Family Life Survey (MxFLS).

Agricultural Productivity and Farm Size in Latin America

Table of Contents

I. Introduction	1
II. Farm Size and Productivity: Conceptual Issues and Literature Review	2
III. Farm Size and Productivity in Latin America	8
IV. The Brazilian Experience	21
V. Farm Size and Productivity in Mexico	32
VI. Conclusions	81
VII. References	84
Appendix 1: The Distribution of Households and Land Productivity by Farm Size in Selected Latin American Countries	
Appendix 2: Construction of the Output Index	
Appendix 3: Construction of Inputs	
Appendix 4: Regression Results Using Crop Production Only	
Appendix 5: Descriptive Statistics for Household and Community Controls	

I. Introduction

The relationship between farm size and productivity has posed a perennial puzzle in development economics (Barrett et al., 2010; Eastwood et al., 2010). Empirical studies spanning a broad range of contexts have established a stylized fact that the two are inversely related. Leading explanations of this phenomenon include imperfections in labor, land, and credit markets (Sen, 1966; Eswaran and Kotwal, 1986), moral hazard between employers and hired agricultural labor (Feder, 1985), aversion to price risk (Barrett, 1996), and measurement and identification issues (Benjamin, 1985; Assunção and Braido, 2007; Carletto et al., 2013; Bevis and Barrett, 2016; Desiere and Jolliffe, 2017; Gourlay et al., 2017). Despite this theoretical understanding, most empirical studies find that existing theory fails to fully explain the observed inverse relationship and the puzzle remains. This discussion has influenced the debate over land reform in Latin America, highlighting bottlenecks and opportunities for unleashing productivity gains in the agricultural sectors of developing economies.

Recent work on agricultural productivity in Brazil has focused on three features of the farm size-productivity relationship. First, how agricultural productivity is measured. While the development literature has relied heavily on partial measures of productivity, such as land productivity, it is becoming increasingly clear that choice of productivity measure matters for the characterization of the relationship and that total factor productivity (TFP) is the preferred measure (Helfand and Taylor, 2016). Second, how the relationship between farm size and TFP has changed over time and across agricultural regions. In some regions of Brazil the inverse relationship between TFP and farm size has persisted, in some it has become U-shaped, while in other rapidly modernizing regions a direct positive relationship has begun to replace the inverse relationship (Helfand and Taylor, 2016). Third, when heterogeneity of productivity growth is allowed for within farm sizes, there is evidence that suggests strong TFP growth for a portion of the smallest and largest farms, driven largely by rapid technical change (Helfand et al., 2015).

The Brazilian experience points towards a dynamic relationship between farm size and productivity, with the inverse relationship weakening if not reversing with modernization of the agricultural sector. Further, it suggests that a portion of both small, family owned farms and larger, commercially run farms may be best suited to harness technical change and achieve rapid productivity gains. This has important implications for poverty reduction in agriculture. It remains

to be determined, however, just how generalizable the Brazilian experience is to other Latin American economies.

We begin this report with a conceptual discussion of the inverse relationship between farm size and productivity. Section III then uses information from household surveys throughout Latin America—drawn from the Rural Income Generating Activities (RIGA) project—to document the farm size – land productivity relationship in a number of Latin American countries. This is followed, in Section IV, by a summary of the Brazilian experience that demonstrates the importance of moving beyond the land productivity – farm size relationship to analyzing the TFP – farm size relationship, especially in dynamic agricultural regions. In Section V we provide an in depth examination of the farm size – productivity relationship in Mexico. We provide novel estimates of land productivity and total factor productivity by farm size based on data drawn from three waves of the Mexican Family Life Survey (MxFLS), a nationally representative panel of households in Mexico that was collected in the first decade of the 21st century.¹ We conclude with a summary of findings and implications of the research.

II. Farm Size and Productivity: Conceptual Issues and Literature Review

A brief review of the literature

It is important to differentiate between physical yields of individual crops, land productivity, and total factor productivity. Crop yields are measured in physical units such as tons per hectare, and thus do not require aggregating across outputs. Land productivity is a measure of the production of all activities per unit of land, requiring the aggregation of all outputs into a single measure prior to dividing by a measure of land. Aggregation can be accomplished using prices to calculate the value of output. Alternatively, a variety of index methods can be used to construct a quantity index as a measure of output over time. Both crop yields and land productivity are partial measures of productivity, examining the productivity of a single input – land. This is a curious feature of the literature on the inverse relationship in development economics, where “land productivity” has regularly been used synonymously with broader concepts of “productivity.”

¹ We thank Graciela Teruel and the researchers who manage the MxFLS for access to crop data for use in this project.

This association was less problematic prior to the green revolution, and in countries where land is scarce and labor abundant. In 21st century Latin America, however, we believe that total factor productivity is a more appropriate measure of performance.

Throughout the world, the majority of studies conducted in the past fifty years have documented an inverse relationship (IR) between land productivity and farm size (Helfand and Taylor, 2016). An inverse relationship between land productivity and farm size could be observed for a number of reasons. At the most basic level, the production technology could exhibit decreasing returns-to-scale (RTS). However, although not all studies find RTS that are exactly constant, there is a broad consensus that RTS in agriculture are approximately constant (Mundlak, 2001).

When studying the relationship between inputs and outputs as identified in the RTS of a given technology, it is appropriate to measure land as the amount actually utilized in production. Similarly, when estimating the yield of a particular crop, it is appropriate to measure land by the amount utilized (either planted or harvested). But when thinking about productivity more broadly, one would want to allow for the fact that small farm sizes might utilize a higher share of the land they control (own, rent, etc.). Similarly, they might be more likely to harvest the same plot of land more than once per year. For this reason, individual crop yields could provide a misleading indication of the more complete relationship between farm size and productivity. This is also a reason why studies that examine broader measures of productivity, such as total factor productivity (TFP) or “social efficiency,” often choose to examine all of the inputs under the control of a farm, not only those actually used.

Many of the theoretical explanations that have emerged to explain the IR rely on household heterogeneity and/or market failure. Sen (1966) hypothesized that surplus labor generated a wage gap between capitalist and peasant farming, differences in the intensity of labor used, and an inverse relationship. Eswaran and Kotwal (1986), in contrast, highlight the role of endowments and credit constraints in generating distinct farmer classes, land-labor ratios, and an inverse relationship. These results are further strengthened with the consideration of moral hazard and the costs of monitoring hired labor (Feder, 1985). Risk aversion and the differing responses to price risk by net buyers and sellers in the absence of insurance markets provides yet another explanation (Barrett, 1996). To the extent that the IR is caused by high transactions costs that generate market

failures, it might be more appropriate for policy to address these failures directly rather than via land reform.

Measurement error and omitted variables, especially in the form of soil quality, are two additional reasons why we might observe an inverse relationship. If correct, the IR might even be spurious. However, Assunção and Braido (2007) find that the inclusion of soil quality variables reduces but does not eliminate the IR using International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) data from India, while Barrett et al. (2010) find that these variables make no difference in a study of Madagascar. While these are only two studies, they do suggest that there are deeper reasons for the IR beyond unmeasured soil quality. Both studies also provide evidence for the IR at the plot level.

A number of recent papers examine the issue of measurement error in land and output as possible reasons for the IR. Carletto et al. (2013) compare self-reported to GPS generated measures of land and find no support for this type of measurement error as the source of the IR. If anything, the IR is strengthened when GPS measures of land are used. A recent working paper by Desiere and Jolliffe (2017) examines measurement error in self-reported production. The paper compares self-reported production to production estimated by trained technicians using the “crop-cut” method on a sample of subplots. They argue that measurement error on subplots should be independent of plot size because the subplots are all of the same size (often around 4m x 4m). With the crop-cut method, the inverse relationship between plot size and land productivity disappears. This explanation does appear to have promise, but is part of an ongoing debate that includes edge effects as a possible alternative explanation (Bevis and Barrett, 2016). Gourlay et al. (2017), in contrast, support the measurement error in production explanation, ruling out edge effects by applying crop cuts to entire fields. While we think it is possible that measurement error in production could explain a significant portion of the IR for the very smallest farms, we are more skeptical that this could continue to explain the relationship that we observe even as we move to farms of 50, 100 and 1000 hectares.

The papers examined above seek to explain the IR between farm size and land productivity. Helfand and Taylor (2016), in contrast, argue that a focus on land productivity is misplaced. While this might have been acceptable 50 years ago, or even today in countries where little modernization of agriculture has occurred, total factor productivity (TFP) is a more appropriate measure of

performance for comparing farms of different sizes. This argument is not new, but has largely been overlooked by the literature. Berry and Cline (1979) and Binswanger et al. (1995) both pointed out the limitations of using land productivity as a measure of performance, and Bardhan (1973) and Berry and Cline (1979) provided examples where an alternative choice of productivity measure could lead to overturning the IR between farm size and productivity.

Land productivity vs. total factor productivity

In what follows below, we draw from Helfand and Taylor (2016) to explain the connection between land productivity and TFP, and the circumstances under which they provide consistent (or divergent) conclusions about the relationship with farm size.

a) The unconditional relationship between land productivity and farm size

Land productivity is by far the most commonly used measure of productivity in the development literature on the IR. Even when alternative productivity measures are used, the relationship between land productivity and farm size is often a starting point. Land productivity, q , is the partial measure of productivity:

$$\text{Land productivity} = \frac{Q}{A} = q = \psi_u(A) \quad (1)$$

where A is the area of the farm, Q is an index of agricultural output, q is agricultural output per unit of land, and $\psi_u(A)$ connotes that land productivity may be a function of farm size. In a world where farm size and land productivity are unrelated we have $\frac{\partial \psi_u(A)}{\partial A} = 0$. However, the regularity with which empirical work has found $\frac{\partial \psi_u(A)}{\partial A} < 0$ has led to the stylized fact that they are inversely related.

The relationship captured by $\psi_u(A)$ is unconditional (u) in the sense that it is the simple bivariate relationship between land productivity and farm size. Factors that may be causing or influencing this relationship have not been controlled for. Using land productivity as a measure of productivity is inherently limited—as would be any partial measure of productivity—whenever there is more than one factor of production. If other factors vary systematically with farm size then

we might expect to see an IR due to more intensive input use by small farms, implying that our focus on the relationship between land productivity and the size of the farm may be misplaced. Similarly, analysis using different partial productivity measures may result in conflicting policy recommendations.

b) The conditional relationship between land productivity and farm size

It is curious that the unconditional relationship between land productivity and farm size has remained such a focal point in the IR literature. A more appropriate approach is to use a conditional relationship between land productivity and farm size, where the relationship is conditioned on a set of controls, $\mathbf{X}(A)$, that may be correlated with both land productivity and farm size:

$$q = g(\mathbf{X}(A), \psi_c(A)) = \psi_u(A) \quad (2)$$

The conditional (c) relationship, $\psi_c(A)$, should differ from the unconditional relationship to the extent that the conditioning controls explain the unconditional IR. For example, the impact of varying input intensities can be controlled for by including the inputs as controls, household heterogeneity could be controlled for with household fixed effects or household-level variables, market failure controlled for with regional fixed effects or regional-level variables, and omitted variables such as soil quality could be added. When the conditional relationship includes all factors of production as controls, the approach is equivalent to estimating a production function. Helfand and Taylor (2016) show that an unconditional IR is neither a necessary nor a sufficient condition for an inverse relationship between a broader (conditional) measure of productivity and farm size as captured by $\psi_c(A)$.

c) Linking land productivity and total factor productivity

We demonstrate the relationship between land productivity and total factor productivity by way of an example assuming a standard Cobb-Douglas production function. We write production (Q) as a function of labor (L), capital (K) and land (A), where T is the unobserved measure of total factor productivity:

$$Q = f(L, K, A) = TL^\alpha K^\beta A^\gamma \quad (3)$$

If, as is often presumed, constant returns to scale (CRS) holds and the technology is homogenous of degree one ($t=1$), then farm size disappears from the right hand side of (3) after dividing through by farm size. If not, then the natural log of the production function takes the form:

$$\ln q = (t - 1)\ln A + \ln T + \alpha \ln l + \beta \ln k \quad (4)$$

If, there exists a relationship between total factor productivity and size, $\varphi(A)$, we have:

$$\ln q = (t - 1)\ln A + \varphi(A) + \alpha \ln l + \beta \ln k \quad (5)$$

From (5) it is clear that the conditional relationship identified in (2), $\psi_c(A)$, is composed of the relationship between TFP and farm size as well as any deviations from CRS in the production function. Equation (5) highlights three important features of the production function approach that are often overlooked in the literature. First, if CRS holds then the conditional relationship, $\psi_c(A)$, captures the relationship between TFP and farm size, $\varphi(A)$. Second, if CRS does not hold then it will be difficult to empirically differentiate whether a conditional relationship is driven by non-CRS, a relationship between TFP and farm size, or both. Third, if the addition of any controls other than the factor intensities explain differences in the conditional relationship, which is composed of potential non-CRS and the TFP-farm size relationship, they should be interpreted as such. In sum, when empirical research estimates a production function to explore the relationship between farm size and productivity they are, in effect, estimating the relationship between farm size and TFP (with the possible influence of non-CRS) and not farm size and land productivity.

Helfand and Taylor (2016) also explore the conditions under which land productivity and TFP are likely to provide consistent, or divergent, conclusions about their relationship to farm size. We summarize their findings here. The authors derive a relationship between TFP and farm size that depends on a) the elasticity of output per unit of land with respect to farm size and b) the elasticity of costs per unit of land with respect to farm size. This way of framing the relationships leads to two conclusions. First, it provides an alternative demonstration of the fact that an IR between a partial measure of productivity and farm size is neither necessary nor sufficient for the existence of an IR between farm size and a comprehensive measure of productivity such as TFP.

Second, it allows them to discuss how these relationships are likely to evolve with the process of development. In an economy with a relatively undeveloped and traditional agricultural sector where production relies predominantly on land and labor as inputs, the underlying mechanisms for

an IR are likely to be Sen's dual labor market hypothesis and the monitoring costs associated with large farms hiring wage labor. In such a setting, land productivity may be an adequate proxy for measuring the relationship between farm size and TFP. In a second stage of agricultural development, imperfections in credit and labor markets persist, yet advances in agricultural technology lead to mechanization, making capital a more important class of inputs. Larger farms can more readily substitute away from labor, avoiding costly monitoring and moving them towards a more efficient mix of factors of production. In such a scenario, one would not be surprised if a direct relationship between TFP and farm size is observed, even if an IR continues to exist for land productivity. Such an argument mirrors that of Sadoulet and de Janvry (1995), where labor-saving technology at the upper end of the farm size distribution leads to TFP rising with farm size when technology becomes more input intensive. Foster and Rosenzweig (2017) build upon this idea, showing that such a model with economies of scale in labor-saving technology theoretically generates a U-shaped farm size – TFP relationship. Unlike in the first stage, mechanization and the use of other labor saving inputs has rendered land productivity a misleading measure for identifying the relationship between farm size and productivity for this economy, as any inverse farm size – TFP relationship would likely weaken more quickly than the inverse relationship between farm size and land productivity.

Helfand and Taylor discuss a possible third stage of agricultural development that could realign the relationships between TFP, land productivity and farm size. As institutions improve, distortions in the labor and capital markets begin to disappear. Other sources of economies of scale begin to come into play, more related to managerial and technical expertise, and the intensive use of information. These new factors not only contribute to a direct relationship between TFP and farm size, the inverse relationship between land productivity and farm size may wither as well.

III. Farm Size and Productivity in Latin America

In this section we present empirical evidence on farm size and productivity in a number of Latin American countries. We begin by providing a brief summary of the literature on agricultural TFP in Latin American and Caribbean (LAC) countries. We then describe data from the 1960s and 1970s drawn from Berry and Cline (1979) on the Northeast of Brazil, Colombia, and to a lesser extent Mexico. In the first two cases they have estimates on both land productivity and TFP. We

then present estimates of land productivity from the early 21st century from a half dozen countries in the region. With the exception of Paraguay, these were calculated from the Living Standards Measurement Study (LSMS) surveys that were systematized across countries by the Rural Income Generating Activities (RIGA) project.²

a) Agricultural TFP in Latin America

A number of studies provide estimates of agricultural productivity in LAC countries and the region overall. While these studies vary in the countries and time periods studied, the source of data, and the approach to measuring TFP, they are overwhelmingly cross-country, national-level studies to estimate long-run, average, annual growth in agricultural TFP. Recent examples include Headey et al. (2010) who use stochastic production frontier analysis and data envelopment analysis to estimate TFP over the 1970-2001 period, finding mean average annual growth rates of 2.2-2.3% for LAC countries. In contrast, Fuglie (2008) uses an index number approach to assess agricultural productivity over the 1961-2006 period, finding that growth in LAC agricultural TFP has accelerated in that period, from an average annual rate of 0.61% from 1970 to 1979 to 2.48% from 2000 to 2006. Bravo-Ortega and Lederman (2004) estimate an average production function, finding that, on average, agricultural TFP grew at a rate of 1.2% between 1960 and 2000 for LAC countries. More recently, Trindade and Fulginiti (2015) use both a stochastic production frontier and a Malmquist Index to estimate growth in agricultural TFP for a subset of South American countries over the 1969-2009 period. They find roughly similar results from the two approaches, and also find accelerating TFP growth that results in an average annual rate of 2.29% in the first decade of the 21st century. Common across these more recent studies and those that preceded them is a finding of positive average annual growth in agricultural TFP in the LAC region.

Similarly, a consistent finding in the literature is heterogeneity in agricultural productivity growth in the region. Headey et al. (2010) identify a subset of countries with fast growing agricultural TFP such as Costa Rica (3.3% annual growth rate over 1970-2001), a group of countries with average growth rates (estimated to be 2.2-2.3% per year), and a group of countries with slow growing agricultural TFP, including some South American countries such as Paraguay,

² See <http://www.fao.org/economic/riga/riga-about/en/> .

Uruguay, and Argentina, and Caribbean nations such as Haiti. Bravo-Ortega and Lederman (2004) find Brazil, Mexico, and Argentina having the highest growth rates (near 2% per year on average), with Paraguay and El Salvador having the lowest average agricultural productivity growth over the period (0.74% and 0.53%, respectively). Trindade and Fulginiti (2015) find that Chile and Brazil grow most rapidly over the 1969-2009 period (2.55% and 2.03%, respectively), with Colombia and Bolivia growing the slowest (0.74% and 0.71%, respectively).

While informative regarding average growth in agricultural TFP, the above studies do not speak to any relationship with farm size. The sole exception in the literature is Vollrath (2007), who includes a land Gini in a national-level, cross-country analysis. While the analysis does not speak directly to the farm size – productivity relationship, the finding of a negative relationship between land inequality and agricultural TFP is consistent with an inverse relationship between farm size and TFP.

b) Estimates of the inverse relationship in the 1960s and 1970s

Brazil

Berry and Cline report estimates of land productivity and TFP calculated from a 1973 survey of approximately 8,000 farms in the Northeast of Brazil. The data are classified in seven separate zones to control for differences in soils, climate and agricultural potential. Table 1 shows average farm size and land productivity for six size classes in each of the seven regions. There is an IR in six of the seven zones, with the exception of zone F where cocoa production is located and land productivity rises and then falls. Across the six zones with an IR, land productivity in the largest farm size class is only 13% of land productivity in the smallest farm size class, and in four of the six zones it is less than 6%. When the authors attempt to control for land quality by using the value of land in the denominator, rather than the quantity of land, the qualitative results remain. Results from a number of different models that regress the value of output on farm size, zone dummies, and the value of land all confirm a negative and statistically significant coefficient on farm size. Similarly, when they seek to control for product mix by classifying producers into one of six crop

groups³ and running the regressions separately for each group, they find a significant negative coefficient in all cases other than cocoa.

Berry and Cline also estimate “total social factor productivity,” which is a measure of TFP at hypothesized “social” prices. TFP is measured as the value of output divided by the sum of land, capital and labor costs. The opportunity cost of land and capital is assumed to be 15% of the value of the stocks, while the assumed social opportunity cost of labor is between zero and one half of the regional minimum wage. The authors conclude: “In general, total social factor productivity declines as farm size rises.”⁴ The decline is more rapid when labor is valued at zero. When labor is valued at half the minimum wage, “there is a tendency for TFP to be higher in the second and third size groups than the first,” suggesting that the smallest of farms use too much labor. The sensitivity of the relationship to the assignment of values to labor is indicative of a more general problem with index approaches to measuring TFP, namely the assignment of weights and/or values to aggregate inputs. TFP does not decline as rapidly with size as land productivity does. Whereas land productivity in the largest size class was 13% of the smallest, TFP in the largest size class is 52% of the size class with the highest TFP (which

³ Livestock, rice, cocoa, sugar, subsistence, and other.

⁴ The quotations in this paragraph are from Berry and Cline (1979), p. 56.

Table 1: Farm Size and Productivity in the Northeast of Brazil, 1973

Zone	Size group	Avg. farm size (ha.)	Avg. gross receipts/area
A	1	3.7	85.92
	2	25.5	30.73
	3	71.9	16.19
	4	138.9	8.80
	5	313.2	5.00
	6	1,178.0	2.20
B	1	3.7	78.25
	2	26.6	33.14
	3	68.8	15.03
	4	138.9	8.45
	5	317.2	5.99
	6	1,396.9	4.40
C	1	4.9	60.41
	2	27.5	35.54
	3	72.4	28.19
	4	143.3	18.33
	5	288.1	15.87
	6	1,059.2	11.14
D	1	4.8	21.00
	2	24.2	52.16
	3	71.8	47.06
	4	138.4	13.91
	5	282.5	14.17
	6	1,210.6	9.80
E	1	3.7	353.03
	2	26.1	63.88
	3	72.8	46.64
	4	143.6	35.24
	5	283.5	45.81
	6	2,303.6	7.76
F	1	5.6	173.13
	2	27.1	243.65
	3	70.6	268.71
	4	142.3	314.40
	5	294.6	253.46
	6	620.0	227.49
G	1	4.0	197.11
	2	26.5	54.49
	3	73.3	51.34
	4	143.4	30.49
	5	299.1	28.88
	6	1,135.3	10.17

Source: *Berry and Cline (1979), p. 46.*

includes any of the smallest three size classes). Given that the data are from the early 1970s, prior to a period of rapid technological change in Brazilian agriculture, it is not surprising that the TFP – size relationship was still largely negative.

Colombia

Berry and Cline analyze data on Colombia drawn from over one million farms surveyed in the 1960 Agricultural Census. Table 2 reproduces their key results on land productivity. They show a strong negative relationship between value added per hectare and farm size, with farms over 500ha exhibiting only about 10% of the land productivity of farms under 3 ha. After controlling for differences in land prices across farm size as a proxy for difference in land quality, value added per “effective hectare” of large farms rises to approximately 47% of what it is on the smallest farms. This is among the earliest evidence that controlling for soil quality is important in explaining the IR between farm size and productivity. The authors explain that the reasons for the IR relate to the use of less labor per hectare on large farms, a lower share of owned land utilized, and a lower intensity of cropping on the land that is put into production. The one countervailing force was the existence of higher crop yields for some crops on the larger farms.⁵

Table 2: Farm Size and Productivity in the Colombia, 1960-61

Farm size	Number of farms (1,000)	Value-added per:	
		hectare	effective hectare
0-3	606.4	1.37	0.75
3-5	150.2	0.86	0.79
5-10	169.2	0.73	0.73
10-50	201.0	0.44	0.57
50-500	76.0	0.23	0.38
over 500	6.9	0.13	0.35
All farms	1,209.7	0.29	0.46

Source: *Berry and Cline (1979), p. 59.*

Berry and Cline also present estimates of total factor productivity by farm size that are constructed in an analogous fashion to those of Brazil. TFP is calculated as the value of output divided by the opportunity cost of inputs under various assumptions about the value of social

⁵ Berry and Cline (1979), p. 58.

opportunity costs, reflecting market failures, minimum wage policies, surplus labor, etc. The results are similar to what was found in Brazil in that the smallest farms are shown to overuse labor and have TFP approximately 15% lower than farms in the 5-10 ha class (except when surplus labor is assumed to drive social wages all the way to zero). When the opportunity cost of labor is assumed to be half of the minimum wage, TFP is highest for farms in the 5-10 ha class, with TFP on farms over 500 ha having between 55% and 88% of this level (depending on the assumptions adopted). Thus, as in Brazil in 1973, the relationship between TFP and farm size was still largely negative in Colombia in 1960, but the relationship was not as strong as when the IR is calculated with land productivity.

Mexico

Mexico was not one of the case studies analyzed by Berry and Cline, but they do present some information in an appendix that is drawn from the Mexican Agricultural Census. Table 3 shows land productivity in 1940 and 1960 for small and large private farms, and for *ejidos*.⁶ Land is measured in value terms rather than by area. The table shows that small farms had considerably higher land productivity than large farms, but that the gap was cut in half during this period. Land productivity on small farms was about 6.5 times larger in 1940, but this ratio declined to about 3.5 in 1960. Berry and Cline explain that this narrowing was partly a reflection of a shift from livestock to crops on large farms, facilitated by government investment in infrastructure, provision of credit, and other supportive policies. We hypothesize that there was a similar—perhaps even more accentuated—trend in the relationship between TFP and farm size. Land productivity on small farms was also about three times as high as on *ejidos*, with little change in this ratio over time.

Table 3: Farm Size and Productivity in the Mexico, 1940 and 1960

Land productivity	1940			1960		
	Large	Small	Ejidos	Large	Small	Ejidos
Output per value of land	0.24	1.55	0.45	0.33	1.17	0.36

Source: Berry and Cline (1979), p. 198.

⁶ An *ejido* is a communally held farm operating as part of Mexico's *ejido* system, which emerged as part of the Mexican agrarian land reform in the early 20th century.

c) Estimates of the inverse relationship in the early 21st century

We attempted to document the land productivity – farm size relationship across Latin American countries based on Agricultural Census data. Unfortunately, most countries do not have sufficient information by farm size that is readily available. As an alternative, we turned to household level data made available by the Rural Income Generating Activities (RIGA) project.⁷ RIGA was a joint project conducted by the FAO, World Bank, and American University. Drawing largely from the Living Standards Measurement Study (LSMS) surveys conducted by national statistical offices together with the World Bank, the RIGA project sought to create consistently defined variables across countries to study the income generating activities of rural households. Part of the motivation for the RIGA project was to document the importance of non-farm activities in rural areas of developing countries. The LSMS are an excellent source of information on agricultural activities because the surveys have well designed agricultural modules. The RIGA database includes over 45 household surveys drawn from more than twenty countries. Five countries from Latin America are included—Bolivia, Ecuador, Guatemala, Nicaragua, and Panama. We restrict attention to the most recent year with usable data because the time span covered for countries with multiple surveys is not that large—for example 1995 and 1998 in Ecuador, or 2000 and 2006 in Guatemala. We complement the RIGA data from these five countries with information on Paraguay, drawn from a survey conducted in 2000-01.

One limitation of the RIGA data that is particularly relevant for the study of the IR is that they are based on surveys of households, not farms. They are representative of the population—and rural population—of each country, but not necessarily of the farms. Thus, large-scale corporate agriculture is not represented in this data, and we are unlikely to be capturing the full extent of the farm size – productivity relationship at the upper end of the farm size distribution. Some light can be shed on this issue when we discuss the IR for “family” vs “non-family” farms in Brazil.

For each Latin American country in the RIGA dataset we show scatter plots of the log of land productivity and the log of farm size for rural households with agricultural production. The figures also include the estimated kernel-weighted non-parametric relationship between these two variables.⁸ Land productivity is calculated as the value of output of all agricultural activities (gross

⁷ See <http://www.fao.org/economic/riga/riga-about/en/>.

⁸ This analysis uses the `lpoly` command in Stata.

crop income plus gross livestock income) divided by land owned and operated (which includes land rented in).

Figures 1 through 5 show the scatter plots and non-parametric relationships between land productivity and farm size.⁹ Appendix 1, Tables 1 through 6, complement the scatter plots with data by size class on the number of households and land productivity. There are six size classes for each country, ranging from 0-0.5 ha to 10+ ha. In some countries the share of households with more than 10 ha of land is less than 3% (Bolivia and Guatemala), while in others it is between 20% and 40% (Nicaragua, Panama and Ecuador).

Figures 1 through 5 show a strong inverse relationship between land productivity and farm size in all five countries. The appendix tables help provide a sense of the magnitude. In most cases the farms in the smallest size class (0-0.5 ha) have dramatically higher land productivity than all other size classes. In Bolivia and Panama it is ten times larger than the next size class, while in Ecuador it is four times larger, and in Guatemala it is more than double. Even if we examine the IR for the remainder of the size distribution, we see that households with 1-2 ha have considerably higher land productivity than households that operate more than 10ha. In Bolivia and Guatemala the productivity of these smaller farms with 1-2 ha is at least triple the productivity of the larger farms, and in the other countries it is at least five times larger. Based on this household data, there is no doubt that an IR between land productivity and farm size was alive and well at the turn of the 21st century.

The RIGA data for the five Latin American countries is complemented with data on Paraguay drawn from Toledo (2010). The data come from the 2000/01 National Household Survey.¹⁰ The table shows that 43% of the households with land in Paraguay had between zero and 1 ha, and around 27% of households had more than 10 ha. Figure 6 Shows that the IR between land productivity and farm size is similar to what was presented in this section for other Latin American countries. The relationship is negative, although perhaps more stepwise than monotonic. Land productivity falls by more than half as we move from households with 0-1 ha to households with between 1 ha and 5 ha, and again falls by about half for households with between 5 ha and 100 ha.

⁹ In all cases the top 1% of observations on land productivity were dropped. In Guatemala, households with less than .01ha of land were also excluded.

¹⁰ Notes on the data and variable definitions can be found together with Appendix 1 Table 6.

Households with 1-2 ha have more than six times the land productivity of households with over 100 ha.

Figure 1: The relationship between land productivity and farm size in Bolivia, 2005

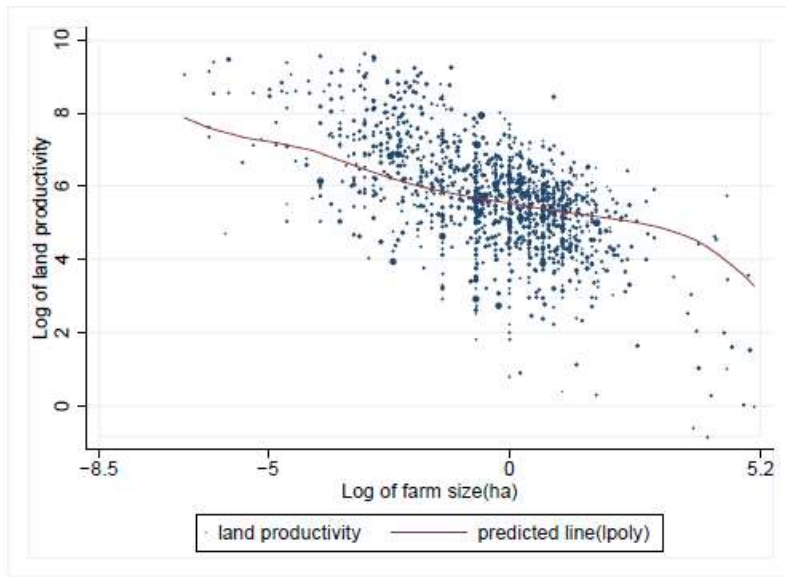


Figure 2: The relationship between land productivity and farm size in Guatemala, 2006

Figure 2: Land productivity by farm size in Guatemala, 2006

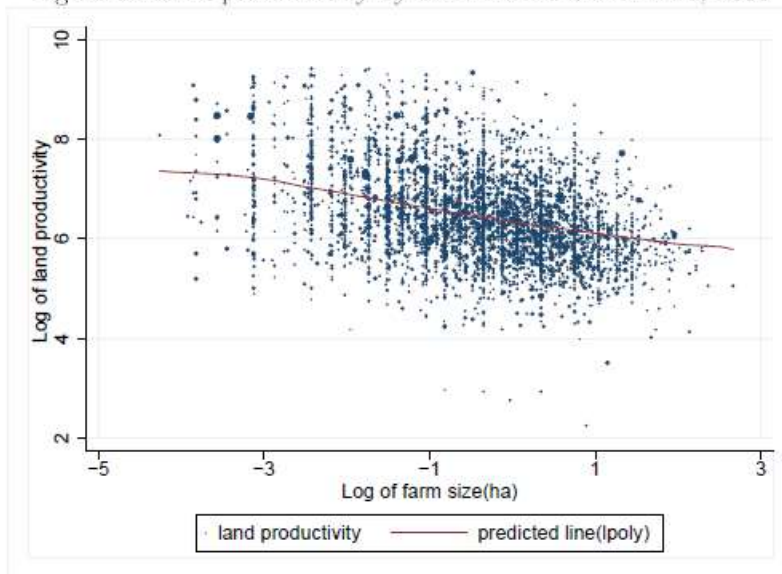


Figure 3: The relationship between land productivity and farm size in Nicaragua, 2005

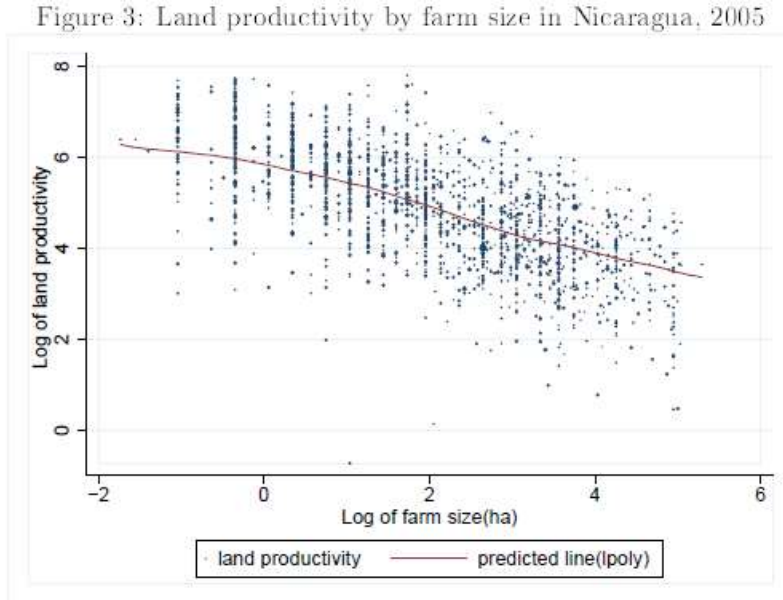


Figure 4: The relationship between land productivity and farm size in Panama, 2003

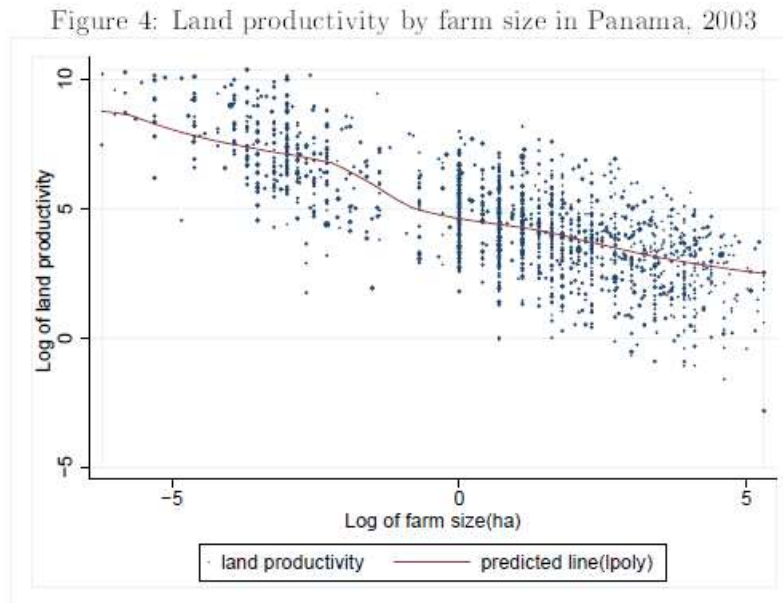


Figure 5: The relationship between land productivity and farm size in Ecuador, 1995

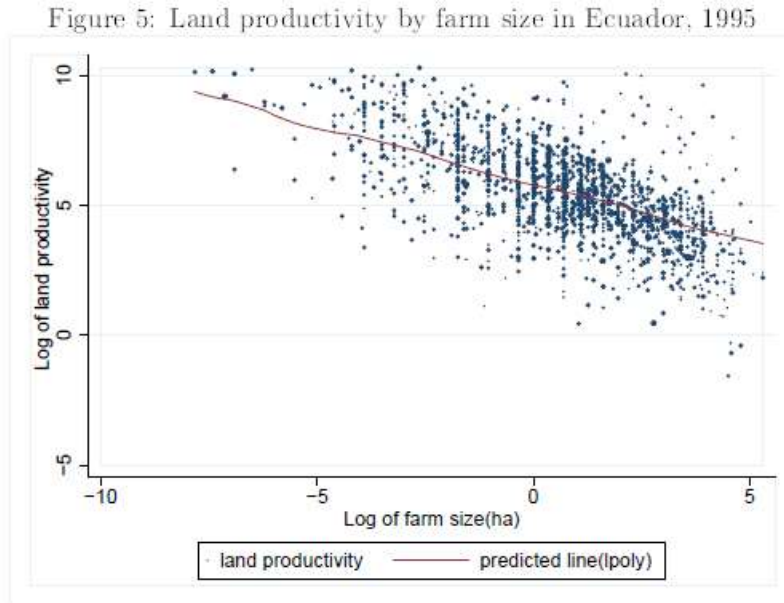
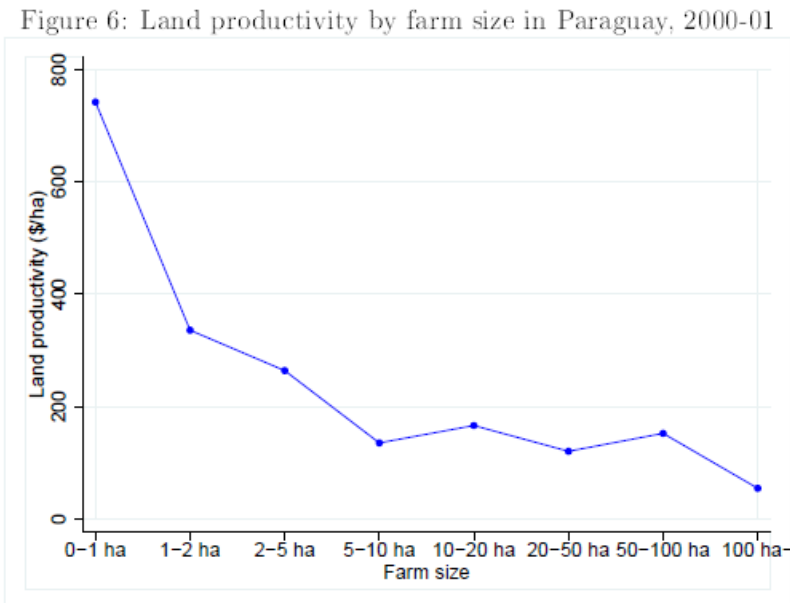


Figure 6: The relationship between land productivity and farm size in Paraguay, 2000-01



Source: Calculated from data in Toledo (2010), p. 92.

IV. The Brazilian Experience

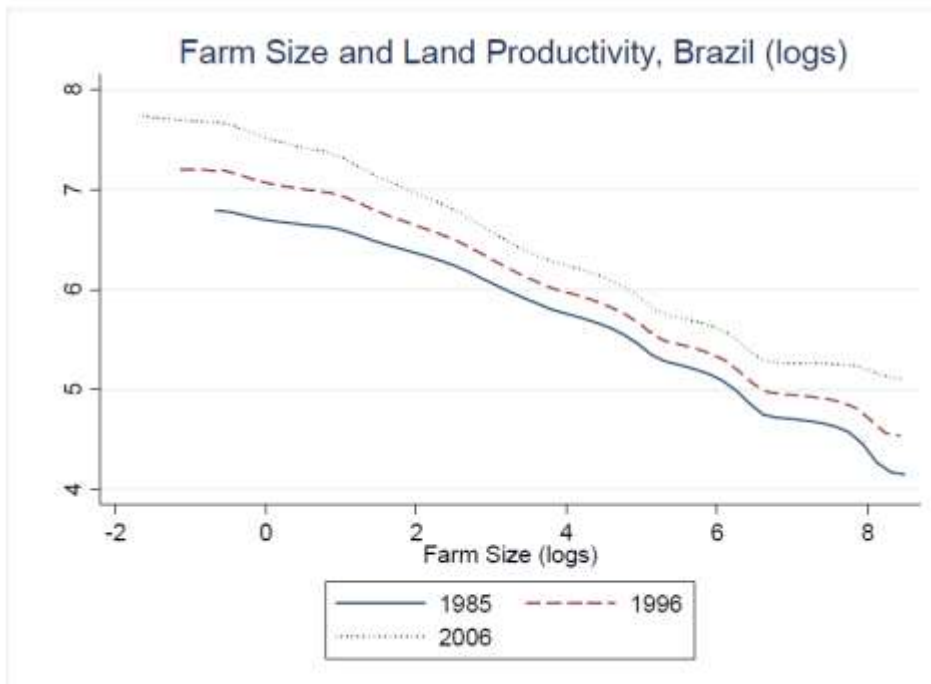
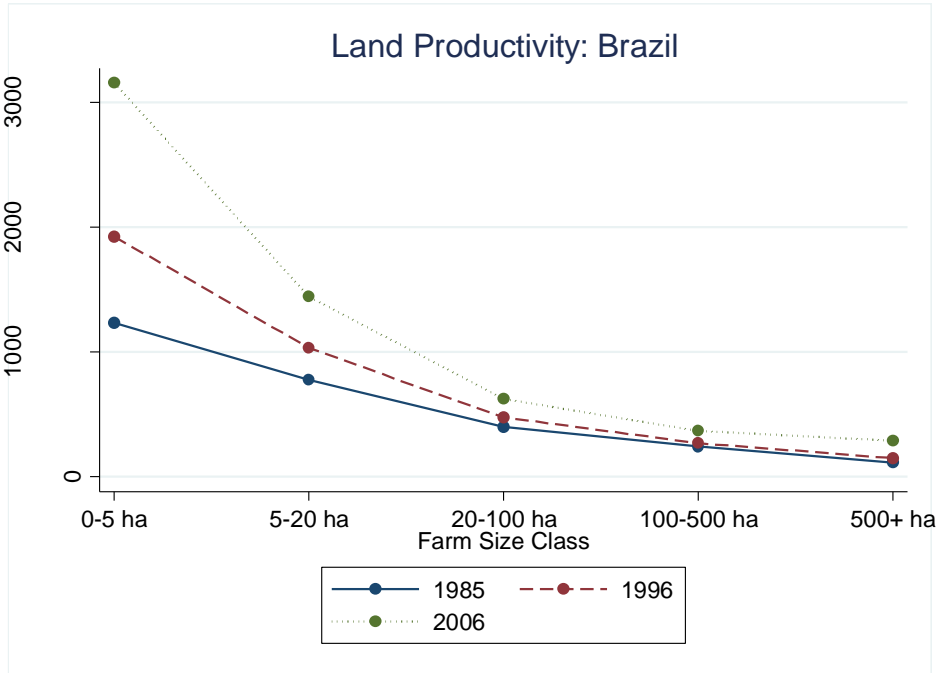
The experience of Brazilian agriculture between 1985 and 2006 highlights the need to 1) move beyond land productivity to a focus on total factor productivity when assessing the relationship between productivity and farm size, 2) consider as much of the farm size distribution as possible, and 3) allow for regional differences when feasible. As discussed in previous sections, as agricultural sectors modernize, technology changes and the production process incorporates more capital and intermediate inputs. In such settings, land productivity is no longer an adequate measure of performance and can be an increasingly misleading proxy for TFP. Thus, an explicit focus on TFP becomes more important. Second, in many parts of the developing world the majority of farms are extremely small. Using FAO data from the 1990s, Eastwood et al. (2010) report that 92% of farms in East Asia were smaller than 2 ha. The situation is similar in South Asia (78%) and Sub-Saharan Africa (69%). Brazil is unusual in this regard, combining the small farms common in many developing countries with much larger farms more typical of the U.S. or Australia. Brazil had over 100,000 farms larger than 500 ha—or 2% of the total—at the time of the last Census in 2006. Some of these farms had tens of thousands of hectares each. To the extent that rapid modernization is happening on the largest farms, it is important that these be included in the analysis. Finally, there can be important difference across regions in topography, farm size, product mix, agricultural potential, technology, and whenever possible it is helpful to disaggregate the analysis of productivity to allow these differences to appear more clearly. All three of these issues—land productivity vs. TFP, the importance of very large farms, and regional heterogeneity—have been addressed by Helfand and co-authors in a series of papers spanning more than a decade. We summarize here some of the most important findings and lessons from those studies.

1. There still is an IR between farm size and land productivity in Brazil

Helfand and Taylor (2016) use Agricultural Census data from 1985, 1995/1996, and 2006 to estimate the farm size – productivity relationship using both land productivity and TFP. The data are structured as a pseudo-panel, with the farm level information aggregated within each municipality by farm size bin in each census year, creating a panel of representative farms for each farm size bin in each municipality. Using this data, they calculate the relationship between land

productivity and farm size both in discrete bins, and with a non-parametric local linear regression. Both sets of results are shown below in Figure 7a and 7b. The Figures show an

Figure 7a and 7b: The Land Productivity – Farm Size Relationship in Brazil



Note: Smoothed as a local polynomial regression with bandwidth of 1.25 and Epanechnikov kernel.

Source: Helfand and Taylor (2016).

IR over the full range of farm sizes, with land productivity increasing each decade as a result of technical change and the increased use of capital and intermediate inputs.

2) The existence of a direct relationship between crop yields and farm size has not caused the IR between land productivity and farm size to disappear

As discussed in Section II, crop specific yields can be a misleading indicator of the relationship between overall land productivity and farm size for a number of reasons. These include the fact that large farms might a) utilize a smaller share of their land for production, b) utilize land for production less intensively, c) have fewer harvests per year, and d) have a product mix that includes lower value products. These forces appear to be at play in Brazil, as Ferreira Filho and Vian (2015) document physical yields that rise with farm size for a variety of products in 2006. Comparing farms in broad size groups, they show that yields for many crops rise up to the 500-1000 ha size group, at which point they tend to stabilize (with cassava and rice falling beyond that point, and peanuts continuing to rise). In the case of cotton, peanuts, rice and beans, yields are more than double in the 500-1000 ha class compared to the farms with less than 100 ha. For soybeans and milk, in contrast, productivity differs by less than 15% across all size groups. These results are not inconsistent with an IR between overall land productivity and farm size and, importantly, they exclude livestock production.

3) There is an inverse relationship between farm size and land productivity within both “commercial” farms and family farms

It has become common in Brazil to distinguish between family farms and non-family farms. After briefly defining family farms, we show that the IR holds within both groups. Federal Law 11.326 of 2006 identifies four characteristics that define family farms, including farm size limits that vary by municipality, using a preponderance of family labor, deriving a majority of household income from the farm, and not using hired farm managers. Based on the legal definition, 84% of farms in 2006 were identified as family farms. There is a high degree of overlap of this group with the 90% of farms that had less than 100 ha. While some family farms are modern and market

oriented, many are not. Thus, although somewhat imprecise, non-family farms are often described as “commercial” farms.

Helfand et al. (2014) use the 2006 Agricultural Census to analyze land productivity by farm size—for both family and non-family farms—to dispel the common misconception in Brazil that family farms are more productive. The data are presented in Table 4. On average, it is true that family farms are more productive. As a group, land productivity of family farms (R\$648) is 83% higher than the land productivity of non-family farms (R\$354). This, however, is due to the fact that family farms tend to be smaller, and there is a strong monotonic IR within each group. Within family farms, farms with 10-20 ha have land productivity

Table 4: Value of Output per Hectare by Farm Type and Farm Size (Brazil)

Type / Size(ha)	Land Productivity	Ratio: F/NF
Family (F)	648	1.83
0-1	7,192	0.54
1-5	2,858	0.39
5-10	1,867	0.44
10-20	1,198	0.43
20-50	668	0.36
50-100	308	0.30
100-200	119	0.17
200-500	82	0.23
Non-Family (NF)	354	
0-1	13,263	
1-5	7,315	
5-10	4,224	
10-20	2,794	
20-50	1,877	
50-100	1,019	
100-200	685	
200-500	352	
500-1000	260	

Source: Helfand et al. (2014) based on data from the 2006 Agricultural Census.

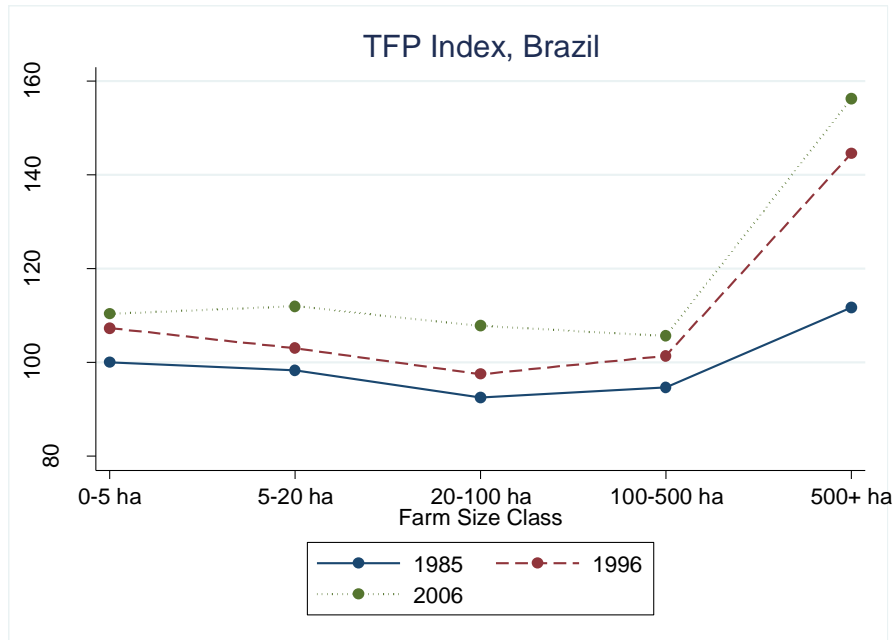
Note: the mean for each type excludes farms with zero land.

that is roughly 10 times larger than farms with 200-500ha, and farms that have 0-1ha have land productivity that is nearly 100 times larger. The ratios are similar for non-family farms: farms in the 0-1ha and 10-20ha classes have land productivity that is roughly 10 and 50 times larger than the farms in the 500-1000 class. But when we hold farm size constant, non-family farms have higher land productivity in every size class (see the column “ratio: F/NF” in Table 4). In the 0-1ha class, land productivity of family farms is only 54% of the land productivity of non-family farms. This falls to around 20% for the larger size classes. Holding farm size constant, the higher land productivity of non-family farms is a result of the technology used in production, and greater use of purchased inputs.

4) The relationship between farm size and total factor productivity is much flatter than the relationship between land productivity and farm size, and it becomes positive for the largest farms

In contrast to the IR between land productivity and farm size, Figure 8 displays the relationship between farm size and TFP. Helfand and Taylor (2016) estimate TFP by farm size using a production function where the log of output per unit of land is regressed against the log of family labor, purchased inputs and a capital index (all per unit of land). The models include municipal fixed effects and time varying size dummies that capture TFP by farm size class in each census year. When using TFP as the measure of productivity, we see a much flatter relationship between farm size and productivity over farms from 0 to 500 ha, with the farms above this size becoming significantly more productive. Relative to farms with 0-5 ha, farms in the 500+ ha class are estimated to have between 35% and 40% higher TFP in 1996 and 2006. Further, by 2006 we are unable to reject a horizontal TFP relationship for farms between 0 and 500 ha. Thus, over time, the relationship has flattened for small and medium sized farms and become positive for the largest farms.

Figure 8: The TFP – Farm Size Relationship in Brazil



Source: Helfand and Taylor (2016).

5) *There are important differences in the farm size – productivity relationship across regions*

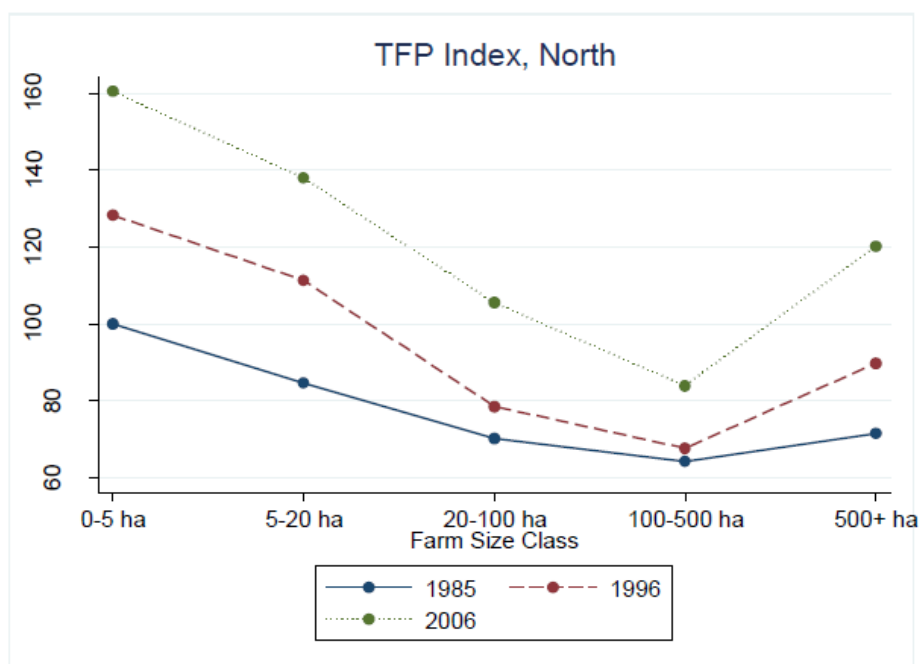
Brazil is a huge country, and its five regions have different agricultural potential, product mixes, and histories. This suggests the importance of conducting a regional analysis. To provide context, the North is where the Amazon rain forest is located. The Northeast is where colonization began, and is home to a large semi-arid zone. Although historically a less developed, low technology region, it contains pockets of recent rapid modernization in both irrigated fruits, as well as grains that have followed the cerrado up from the Center-West. The Southeast is where the largest cities and the country’s industrial base are located. This too is where, among many other products, most sugarcane, oranges and coffee are produced. The South is a region that was influenced by considerable European immigration. It is dominated by family farms that produce a wide variety of products, including grains, chickens and pigs, dairy, irrigated rice, and perennial fruit suited to a more temperate climate. Finally, the Center-West is where Brazil’s agricultural boom has taken place since the 1970s. This is where the cerrado (or Brazilian savannah) is located, extending up into the Northeast. Huge mechanized farms have taken over in recent decades, producing soybeans, corn, cotton and other products, often as a substitute for extensive cattle

raising. With these differences, it is natural to expect that the farm size productivity relationships could differ across regions.

Although not shown here, Helfand and Taylor (2016) present land productivity figures for each of the five regions of the country. All regions show a strong IR between zero and the 100-500 ha class, while in a couple of the regions the relationship appears to become horizontal, or even slightly positive above this size. This suggests that the rising productivity of the largest farms has not been spread equally across the country.

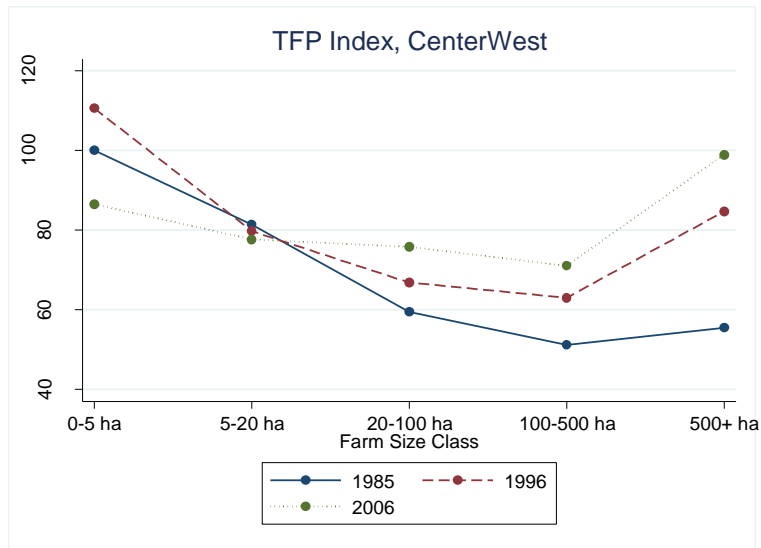
The most interesting results relate to TFP and farms size across regions. In the less advanced agricultural regions, such as the North and Northeast, there continues to be an IR up to 500 ha, but the largest size class has had rapid productivity growth and is starting to transform the relationship into more of a U-shape (see Figure 9 on the North). In the more modern regions, the TFP – size relationship is dynamic, becoming increasing flat up to 500 ha and then positive beyond that size (see Figure 10 on the Center-West). These results for the Center-West are broadly consistent with those presented in Helfand and Levine (2004), which show a U-shaped relationship between farm size and technical efficiency in the Center-West in 1996.

Figure 9: The TFP – Farm Size Relationship in Brazil



Source: Helfand and Taylor (2016).

Figure 10: The TFP – Farm Size Relationship in Brazil’s Center West



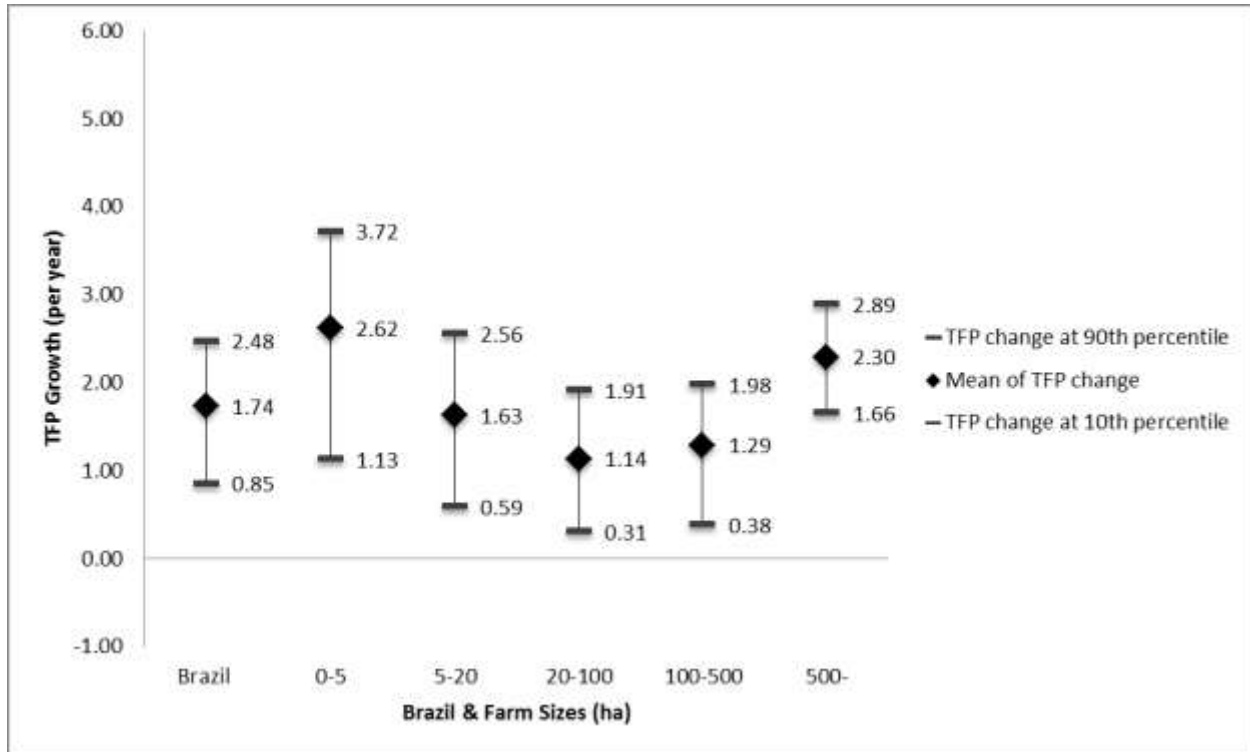
Source: Helfand and Taylor (2016)

6) Heterogeneity of TFP growth within size classes

Both Helfand et al. (2015) and Helfand et al. (2017) explore the heterogeneity of TFP growth within farm size classes over the period 1985-2006. The first paper does so with a stochastic frontier production function, while the second paper uses an average translog production function with fixed effects and recovers the entire distribution of TFP changes within each size class. Both approaches suggest significant dispersion in performance. Thus, while there are important differences across farm sizes, which is what we have focused on so far in this paper, farm size is just one of many potential variables that could explain differences in productivity growth. There are dynamic farms of all sizes, and it is important to identify what differentiates these farms from the rest.

Figure 11 provides an indication of the dispersion in TFP growth for each size class. It shows that median TFP growth of the smallest and largest farms was often higher than TFP growth at the 90th percentile for farms in the middle of the size distribution. Alternatively, TFP growth at the 10th percentile for farms over 500ha was estimated to be faster than at the median for farms

Figure 11: Dispersion of TFP Growth by Farm Size, 1985-2006



Source: Helfand et al. (2015).

between 5ha and 500ha. Helfand et al. (2017) explore some of the reasons for differential performance in TFP growth. Education of the farmers and their employees is found to boost TFP growth across most farm sizes. Differences in specialization also seem to matter – for farms over 500 ha, specialization in annual crops (such as soybeans, corn and cotton) was associated with faster than average TFP growth, while in contrast, farms under 5 ha that specialized in perennials were the ones that performed best. Clearly, a focus on the many determinants of TFP growth—including but not limited to farm size—warrants additional attention.

7) Productivity and poverty reduction

One way to think about TFP growth is in terms of the real value of output rising more rapidly than the real value of inputs. As a result, profits rise leading to an increase in income for farmers. Thus, TFP growth could be a powerful mechanism for poverty reduction. Even with higher levels of productivity, however, it might be that some farms are simply too small to generate sufficient

income to bring their household members above the poverty line. Helfand et al. (2011) seek to tackle this question. They estimate the extent to which the distributions of land and productivity explain poverty levels among agricultural producers in Brazil. They first use farm level data from the 1995-96 Agricultural Census to estimate a stochastic frontier profit function. They then use the estimated model to simulate the impacts on poverty of possible scenarios. The results of some of the most interesting counterfactual simulations are described below.

When comparing the differences in poverty among farmers across regions, Helfand et al. (2011) find that the distribution of land is almost always the most important explanatory factor. For example, poverty among farmers was 54 percentage points (pp) higher in the Northeast than in the Southeast (when only agricultural income is considered). If the farmers in the Northeast had the distribution of land that farmers in the Southeast had, poverty was simulated to be 36pp lower. Differences in productivity were normally the second most important factor. Poverty would be 28pp lower in the Northeast if the farmers there had the distribution of productivity observed among the farmers in the South.

When examining the differences between the poor and non-poor farmers within each region, differences in productivity are shown to be more important—although of a similar magnitude—than differences in the distribution of land. For example, in the Northeast, 76% of farmers were poor. If the poor farmers in the Northeast had the distribution of productivity that the non-poor farmers had, poverty would fall by 44pp. If they had the distribution of land, poverty would fall by 40pp. In the case of the South where poverty among farmers was 22%, productivity leads to an 18pp reduction in poverty while the effect of the distribution of land was only 12pp.

Based on these and other simulations, the authors conclude that poverty among agricultural producers is a result of both the distribution of land and low levels of productivity among the poor. Holding the distribution of land constant, income can increase as a result of rising productivity. But many farms—especially in the Northeast of Brazil—have too little land to generate a non-poor level of income even with higher levels of productivity. Thus, attacking both fronts is likely to be necessary to reduce poverty by a significant amount.

8) *Lessons from Brazil*

The Brazilian experience suggests that agricultural modernization has the potential to drastically change the farm size – productivity relationship so often observed in the agricultural sectors of developing economies. Technical change has been driving TFP growth in Brazil. This has resulted in a changing relationship between farm size and productivity when measured using TFP, even when the relationship appears inverse and relatively constant over time when measured using land productivity.

The Brazilian evidence paints a picture of the farm size – productivity relationship as an evolving phenomenon, likely dependent upon the relative modernization of the agricultural sector in question. The use of land productivity can fail to capture important changes in this relationship. In the case of Brazil, land productivity was found to be inversely related to farm size in spite of rising crop yields for many products. It was also found to be inversely related to farm size among both family and non-family farms, as well as in every region of the country. When the level of technology in agriculture is relatively low, land productivity may serve as a suitable proxy for TFP. But as agriculture modernizes, the farm size - land productivity relationship can remain inverse even if the farm size – TFP relationship has changed. Evidence was provided of this phenomenon in the more modern and dynamic regions of Brazil.

While size is one of the important determinants of differences in TFP across farms, it is not the only relevant factor. Evidence was presented of substantial differences in TFP growth within and across each farm size. This heterogeneity is related to differences in specialization, levels of education, and the ability of some farmers to more readily adopt new technologies. More research is needed on the relative importance of these alternative determinants of TFP growth.

TFP growth can also play an important role in poverty reduction. Within each region of Brazil, differences between poor and non-poor farmers were due to both lower levels of productivity among the poor and insufficient access to land, with productivity playing a slightly larger role. Both the level of productivity and the amount of land utilized contribute to the total income of farm families. Thus, poverty reduction among poor farmers could be achieved via TFP growth among the poor, land reform, or even land consolidation. While land consolidation could contribute to creating a middle class of farmers, it begs the question of employment opportunities for the farmers who choose to exit the sector.

V. Farm Size and Productivity in Mexico

The relationship between farm size and productivity in Mexico is important for understanding this relationship more broadly in Latin America. Mexico is the second largest producer of agricultural goods in the region (second only to Brazil), producing approximately 15% of the region's agricultural output in 2015. Agricultural output accounted for approximately 3.6% of Mexico's GDP in 2015, employing 13-14% of Mexico's workforce (this share has fallen from approximately 18% at the turn of the century to 13.5% in 2007, where it has remained relatively constant). Approximately 62% of Mexico's rural population is impoverished when measured with the national poverty line.¹¹ An increase in agricultural productivity is one channel through which rural poverty could be alleviated. As shown in the case of Brazil, the farm size – productivity relationship could potentially play a role in achieving this objective.

Previous research using data from the Mexico National Rural Household Survey (ENHRUM), a household survey statistically representative of 80% of rural Mexico, showed evidence of an inverse relationship between farm size and productivity when estimating both an average production function and a stochastic frontier production frontier (Kagin, Taylor, and Yunez-Naude, 2016). Their analysis of the stochastic production frontier revealed that the observed farm size – productivity relationship was driven, in part, by larger farms being further from the frontier (i.e. smaller farms being more efficient than their larger counterparts). The current analysis of the farm size – productivity relationship in Mexico using the MxFLS expands upon these findings by exploring how that relationship may have changed over time, and by using data that is more recent and includes a little bit more of the upper end of the farm size distribution.

The Mexican Family Life Survey

The Mexican Family Life Survey (MxFLS) is a longitudinal survey of Mexican households, representative of the Mexican population at the national, urban, and rural levels.¹² Controlling for unobservable farm and community level characteristics with fixed effects is important for determining the farm size – productivity relationship. The panel nature of the data, together with

¹¹ All data taken from the World Bank: <http://data.worldbank.org/>.

¹² MxFLS was designed, implemented, and is managed by the Iberoamerican University and the Center for Economic Research and Teaching in Mexico, in conjunction with Duke University researchers.

the decade long span of the surveys, allows for a careful analysis of the evolution of the size-productivity relationship over time, making the MxFLS a rich source of data for this analysis.

The first wave of the survey, conducted in 2002, interviewed 8,437 unique households and recorded detailed information on a broad range of characteristics at the household member, household, and community levels. The second wave, conducted in 2005-06, again interviewed 8,437 households.¹³ Of these, 7,572 were first wave households and 865 were newly formed second wave households, implying a re-interview rate of 90%. The third and final wave, conducted between 2009 and 2012, interviewed 10,119 households.¹⁴ Of these, 1,492 households were newly formed, and the survey succeeded in re-interviewing 94% of the first wave households and 83% of the new second wave households. Table 5 summarizes the evolution of the panel by displaying the number of households in each survey year by the survey year in which they first appear.

Table 5: MxFLS Interviews by Year the Household First Appears

Year Household Appears	Survey Year		
	2002	2005	2009
<i>2002 Household</i>	8,437	7,572	7,912
<i>2005 Household</i>	--	865	715
<i>2009 Household</i>	--	--	1,492
<i>Total</i>	8,437	8,437	10,119

While not representative of the Mexican agricultural sector per se, the MxFLS is representative of both rural and non-rural Mexican households. As such, the use of the dataset to study Mexican agriculture has the important caveat that it likely underrepresents the larger, commercial agricultural operations to the degree that they are not family farms. A comparison with the 2007 Agricultural Census reveals that both the census and MxFLS have less than 5% of farms being greater than 50 ha. However, it is important to note that these “large” farms are not necessarily the same as those in the census, as they are family-run farms and do not include corporate-run commercial agricultural operations. In comparison to the 2007 census, the MxFLS over-represents

¹³ The number of households presented here are the number of unique households in the data – several observations were dropped due to missing or duplicate household identifiers. After taking these actions it is unusual that the first and second wave number of households are identical. One check of this unusual result is to calculate the implied annual rates of household formation between each survey wave. From 2002 to 2005 the MxFLS saw an annual rate of household formation of 3.6%, and from 2005 to 2009 a rate of 4.7% - these are similar in order of magnitude to other nationally representative panels, such as the Indonesian Family Life Survey.

¹⁴ The vast majority of third wave interviews, 95%, were conducted in 2009 and 2010.

farms less than 2 ha and under-represents farms between 20 ha and 50 ha. This is true for each survey wave, highlighting that while the MxFLS is not representative of the Mexican agricultural sector in its entirety, it is adequate for studying household farms in Mexico in much the same way that the RIGA data was used for other Latin American countries.

An ideal empirical setting for assessing the relationship between farm size and productivity would consist of a balanced, plot-level panel for multiple farms in multiple villages. Such a dataset enables the analysis of plot specific, household specific, and institutional/market determinants of the observed input-output relationship. Because input data in the MxFLS is recorded at the household level and is therefore not plot specific, the analysis will be conducted at the farm (i.e. household) level for all households engaged in agricultural production. This eliminates the feasibility of studying agricultural productivity at the plot level; however, as our analysis is more concerned with documenting the farm size – productivity relationship in Mexico and how it has changed over time, and less concerned with fully explaining its determinants, a farm level analysis will suffice.

The identification of a panel, however, is desirable to control for unobserved factors at the household level, and to ensure that any changes in the relationship are not driven by changing composition of the farms in each survey year. Agricultural households in the MxFLS move in and out of agricultural production between survey waves – some households engaging in agricultural production in the first wave also do so in the second and third waves as well, some produce in the third wave but not the second, some produce in the second but not the third, and some stop agricultural production altogether after 2002. Similarly, because new households are formed in later survey waves, some agricultural households do not exist in the first wave of the MxFLS. The construction of the panel involves two stages of restricting the data: first, cross-sections of households with complete farm data are identified and cleaned to eliminate outliers, and second, an unbalanced panel is formed out of all households that appear in two or more MxFLS survey waves.¹⁵

¹⁵ Some households have incomplete data on the size and/or output for one or more plots used in agricultural production in a given period, making the inclusion of these households in a farm-level analysis problematic. With inputs recorded at the household level and output and plot size data recorded at the plot level, the inclusion of farms with missing data on any plot used in production will introduce measurement error. After removing such households, we remove households in the top and bottom 1% of farm size, the bottom 1% of land productivity, and the top 5% of land productivity to eliminate outliers. These efforts were in response to concerns over measurement

In Table 6, all households using plots for production in a given period are referred to as agricultural households, whereas all households with plot size and output data for all non-fallow plots are referred to as complete farms. The intermediate group, farms with farm size data, includes all farms with complete farm size data but not necessarily complete production data – this less restricted dataset increases the sample size at the expense of potentially introducing some measurement error, and is an alternative treatment of the data that is pursued below. Lastly, the number of farms in the panel includes the number of households with complete farm data in two or more of the survey years. These restrictions on the data leave us with a sample of 566 farms reappearing in two or more survey years. Table 7 describes these farms according to the combination of survey years in which they appear.¹⁶ Some of the initial descriptive analysis compares results from the cross-sectional and panel datasets on complete farms.

Table 6: Agricultural Households and Complete Farms by Survey Year

	2002	2005	2009
<i>N Households</i>	8,440	8,437	10,119
<i>N Agricultural Households</i>	1,586	1,303	1,410
<i>N with Farm Size Data</i>	1,042	713	696
<i>N Complete Farms</i>	887	626	596
<i>N Farms in Panel</i>	483	412	359

*Note that *N Complete Farms* and *N Farms in Panel* are after respective rounds of cleaning for outliers

Table 7: Panel Sample Size for Complete Farms

	N	N from Interim Report
All Survey Years	122	152
First and Second Surveys Only	207	238
First and Third Surveys Only	154	158
Second and Third Surveys Only	83	88
<i>Total</i>	566	636

error in farm size and/or the data used to calculate output. This trimming is done within farm size bins. After identifying the unbalanced panel, we further trim households in the top and bottom 1% of percentage changes in the output quantity index.

¹⁶ Note that the sample size has been slightly reduced since the interim report. This is partially due to several households only producing crops for which we do not have accurate translations/prices, but primarily due to cleaning the data for outliers.

Description of the Panel

The MxFLS offers two ways to measure farm size. In addition to asking the size of each parcel used by each household, the survey asks for an estimate of what percentage of each parcel is actually utilized for agricultural production.¹⁷ This offers two potential measures of farm size: first, the size of all parcels used in production by a farm, and second, the size of the utilized area of all parcels used in production by a farm. Both are used in the literature, and depending upon the definition we have a different interpretation of any farm size – productivity relationship identified. We use the former definition of farm size in our base estimates, and use the latter as an alternative measure of farm size in a robustness exercise.

We classify farms into one of 7 farm size groups, as shown below in Table 8, showing the number and percentage of farms in each bin in each year for the cleaned cross sections and the panel discussed above. The distribution of farms across these bins is roughly constant over time in both sets of data. More importantly, with the exception of the share of farms between 0.5 and 1 ha in 2002, the distribution does not change in any notable way as we restrict the cross section to form our panel, an indication that our use of the panel has not biased our sample along these dimensions. There is a considerable range in farm sizes in the sample, ranging from less than one hundredth of a hectare to 45,000 hectares. The median farm size in the panel is 2.5, 2.1, and 3.0 hectares in 2002, 2005, and 2009, respectively, with mean farm sizes of 101, 232, and 218 hectares. If we were to trim the top 5% of the farm size distribution we would reduce the maximum farm size to 46 hectares, but for now we maintain the largest farms in our analysis. The majority of farms utilize only one plot for production in any given year, although some farms use up to 7 parcels in 2002, 5 parcels in 2005, and 10 in 2009.¹⁸

¹⁷ Survey respondents choose between the 5 categories of 10%, 25%, 50%, 75%, and 100%. Missing values to this question were imputed with mean values of 85%, 84%, and 86% for parcels being farmed in 2002, 2005, and 2009, respectively.

¹⁸ Note that the percentage of farms having just one parcel is also roughly constant between the cross section and the panel, with 76% of the cross sections and 73% of the panel having 1 parcel – the difference is made up in the percentage of households having 2 or 3 parcels.

Table 8: Sample Size by Farm Size Group for Complete Farms

Farm Size Group	Cross Sections			Panel		
	2002	2005	2009	2002	2005	2009
0 to 0.5 ha	199 (22%)	116 (19%)	110 (18%)	103 (21%)	66 (16%)	55 (15%)
0.5 to 1 ha	108 (12%)	102 (16%)	101 (17%)	45 (9%)	60 (15%)	57 (16%)
1 to 2 ha	141 (16%)	109 (17%)	96 (16%)	83 (17%)	75 (18%)	58 (16%)
2 to 5 ha	182 (21%)	133 (21%)	122 (20%)	108 (22%)	88 (21%)	75 (21%)
5 to 10 ha	143 (16%)	93 (15%)	91 (15%)	79 (16%)	76 (18%)	65 (18%)
10 to 20 ha	65 (7%)	34 (5%)	40 (7%)	39 (8%)	23 (6%)	27 (8%)
> 20 ha	49 (6%)	39 (6%)	36 (6%)	26 (5%)	24 (6%)	22 (6%)
<i>Total</i>	887	626	596	483	412	359

For each parcel, MxFLS records the ownership status of the piece of land, and additionally the type of document that guarantees that status. The vast majority of plots are either privately owned property or are part of an *ejido* – a piece of communally held land where plots are farmed by designated households.¹⁹ At least 91% of privately held plots in the MxFLS have some form of formal documentation in any given year, while just 75-84% of MxFLS *ejido* properties do. Privately held plots primarily have a formal deed or title to the land as documentation, whereas *ejido* plots primarily have a certificate of *ejido* status or agricultural rights.²⁰

Because *ejidos* may function differently than privately owned parcels, we assess the share of panel farms that are *ejido*. For example, certificate of *ejido* status is often not acceptable to private financial institutions for use as collateral, whereas formal deeds are. Table 9 below shows the share of parcels, the share of farms, the share of complete farms, and the share of farms in the final panel that are *ejidos*.²¹ First, it is surprising that more than half of the agricultural parcels in the MxFLS are part of an *ejido* in each period. Second, this share is roughly constant over time, although in 2005 there is a decline in *ejido* prevalence. Lastly, as the data moves through the cleaning process from parcels, to farms, to complete farms forming the cross section, the shares are effectively constant in each survey year. It is only when we restrict the cross section to form the panel that we see any qualitative difference, where the panel now contains a slightly larger share of *ejidos* than

¹⁹ Other ownership categories include rented land, borrowed land, sharecropped land, and other.

²⁰ In any given year, 71-85% of privately held parcels have a formal deed. In contrast, in any given year 69-76% of parcels have an *ejido* certificate.

²¹ Note that farms that are *ejido* are almost always (at least 95% in any survey year) comprised completely of parcels that are part of *ejidos*, so an indicator variable for whether or not a farm is part of an *ejido* captures almost all of the variation along this dimension.

the cross section. This could, in part, be due to increased stability of ejido parcels, leading to less movement into and out of agricultural production and therefore an increased prevalence of ejido farms in our panel. It will be important to control for being a part of an ejido in the analysis below.

Table 9: Percent of Parcels/Farms Belonging to an Ejido

	All Parcels	All Farms	Cross Section Farms Only	Panel Farms Only
2002	59%	56%	58%	64%
2005	50%	49%	47%	51%
2009	56%	55%	54%	58%
<i>Total</i>	<i>56%</i>	<i>54%</i>	<i>54%</i>	<i>58%</i>

Geographical Distribution

There are three levels of geographical identification in the MxFLS: the community of each household, the municipality of each household, and the state of each household. The distribution of panel farms across the states covered by the MxFLS is shown in Table 10, grouped into five regions in Mexico: the North, Center, Pacific, South, and Gulf. Agriculture in the Northern states is characterized by having larger commercial farms with greater importance of the commercial production of maize, whereas agriculture in the Southern and Central states is characterized by more traditional, smallholder maize producers and commercial production of fruits and edible vegetables (Prina, 2013). The distribution of farms in the panel is stable across both states and regions, with the majority of farms coming from the North and the Center regions.²² The panel spans 62 municipalities, averaging 20 farms per municipality. There are 92 communities represented across these 62 municipalities, each of which has an average of 14 farms per community over all survey years.²³

²² An unfortunate element of the community data is that several communities do not appear in the 2005 and 2009 community dataset, leading to 2 observations not having state, municipality, and community identifiers and community data in 2005, and 34 such observations in 2009. We assume geographical stationarity of these households over time in order to recover these geographical identifiers. This assumption reduces the number of observations with missing geographical identifiers to just 2 in each of 2005 and 2009.

²³ Across all survey years, there is an average of 1.4 communities per municipality, ranging from just 1 to 5.

Table 10: Share of Farms in Each State, by Survey Year

Region	2002	2005	2009
North	26%	28%	28%
<i>Baja California Sur</i>	1%	1%	1%
<i>Coahuila</i>	1%	1%	1%
<i>Durango</i>	15%	17%	16%
<i>Nuevo Leon</i>	1%	1%	1%
<i>Sinaloa</i>	8%	8%	9%
<i>Sonora</i>	0%	0%	0%
Center	31%	28%	32%
<i>Distrito Federal</i>	0%	0%	0%
<i>Guanajuato</i>	8%	7%	8%
<i>Estado de Mexico</i>	11%	9%	13%
<i>Morelos</i>	3%	3%	3%
<i>Puebla</i>	9%	9%	8%
Pacific	7%	8%	6%
<i>Jalisco</i>	1%	2%	1%
<i>Michoacan</i>	6%	6%	5%
South	19%	19%	19%
<i>Oaxaca</i>	19%	19%	16%
Gulf	15%	15%	17%
<i>Veracruz</i>	11%	11%	12%
<i>Yucatan</i>	4%	4%	4%
Missing	0%	<1%	1%

Farm Size and Land Productivity

As with much of the literature, we begin our discussion of the farm size – productivity relationship using a measure land productivity, measured as output per hectare. Figures 12 through 14 show the relationship between the log of farm size and the log of output per hectare in each survey year using the cross-sectional data. Figures 15 through 17 show the relationship between the log of farm size and the log of output per hectare in each survey year for the panel. Farm output in the cross sections is measured in 2002 Mexican pesos, whereas output in the panel is measured using a Fisher quantity index. Appendix 2 provides a detailed discussion of the construction process for these measures of output.

There is a clear inverse relationship between farm size and land productivity over the entire range of farm sizes, one that is strikingly consistent over time and across samples. In each year, land productivity falls rapidly up to approximately 1 ha. Around 1 ha the relationship levels, becoming relatively flat up to approximately 20 ha, at which point a dramatic decline in land productivity once again occurs. The consistency of these results across the cross section and panel for each survey year is encouraging, indicating that the rigorous treatment of the data has not obscured or skewed the relationship between farm size and productivity that we wish to assess.

Figure 12: Non-parametric Relationship between Farm Size and Productivity, Cross-Section (2002)

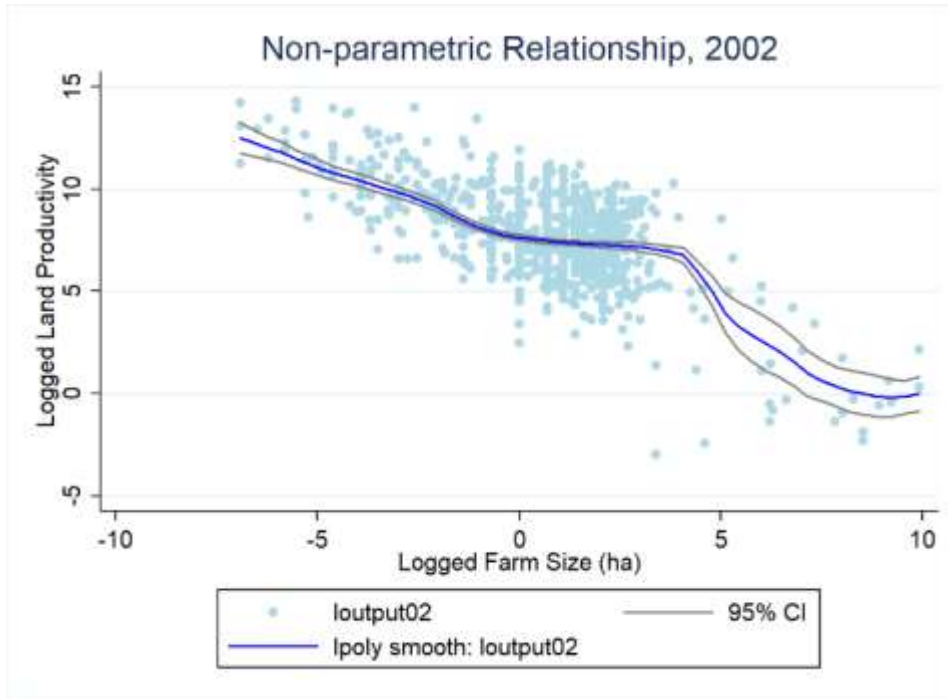


Figure 13: Non-parametric Relationship between Farm Size and Productivity, Cross-Section (2005)

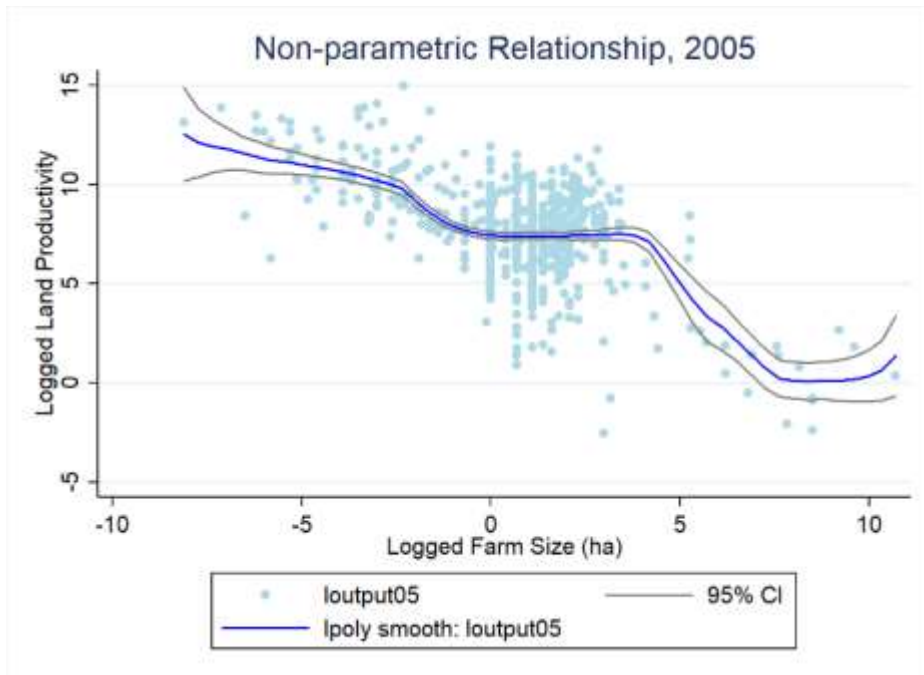


Figure 14: Non-parametric Relationship between Farm Size and Productivity, Cross-Section (2009)

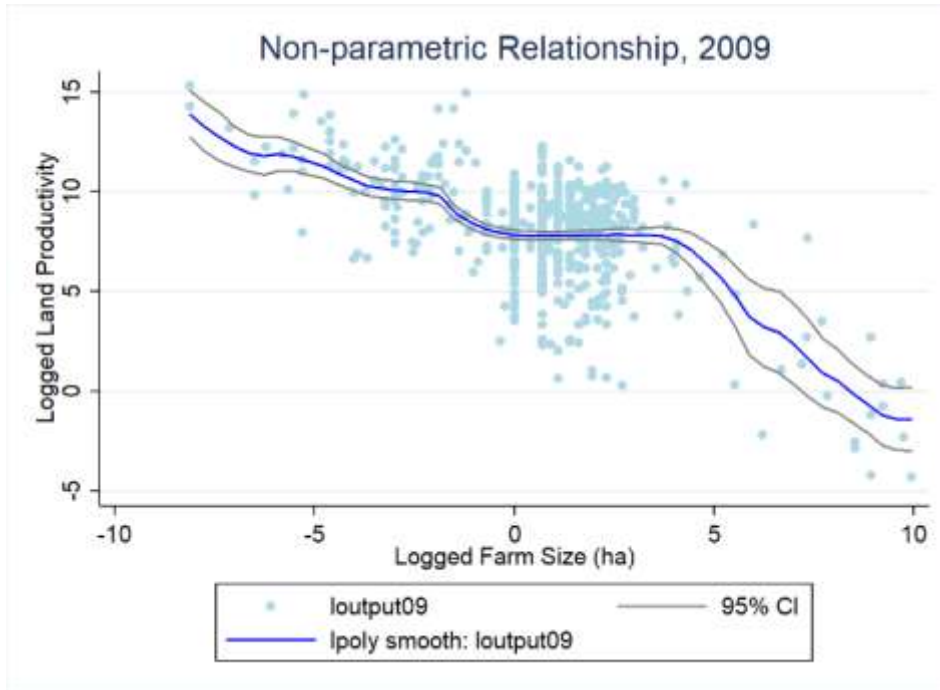


Figure 15: Non-parametric Relationship between Farm Size and Productivity, Panel (2002)

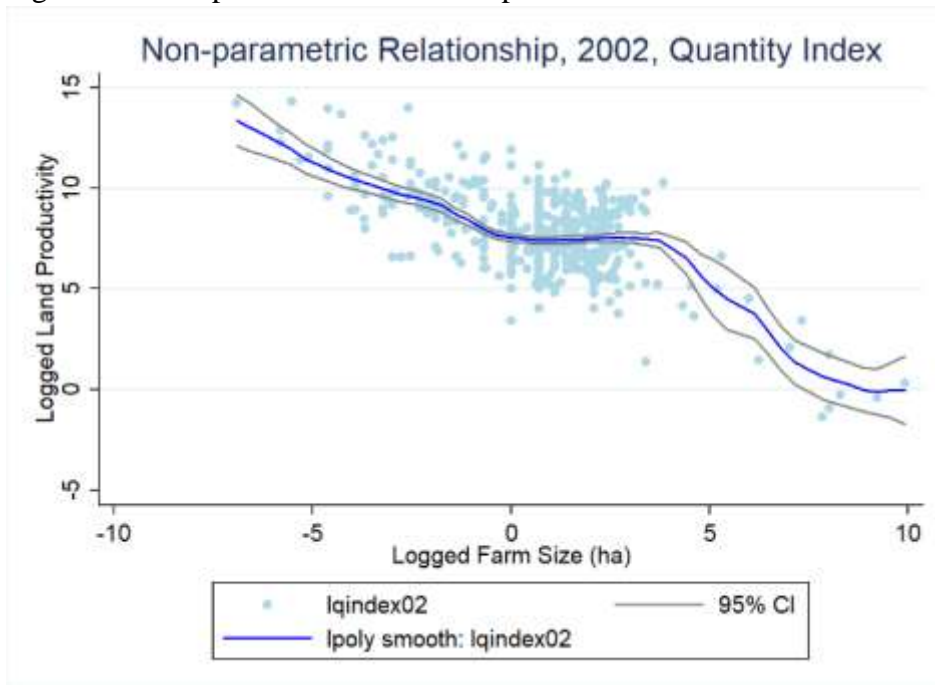


Figure 16: Non-parametric Relationship between Farm Size and Productivity, Panel (2005)

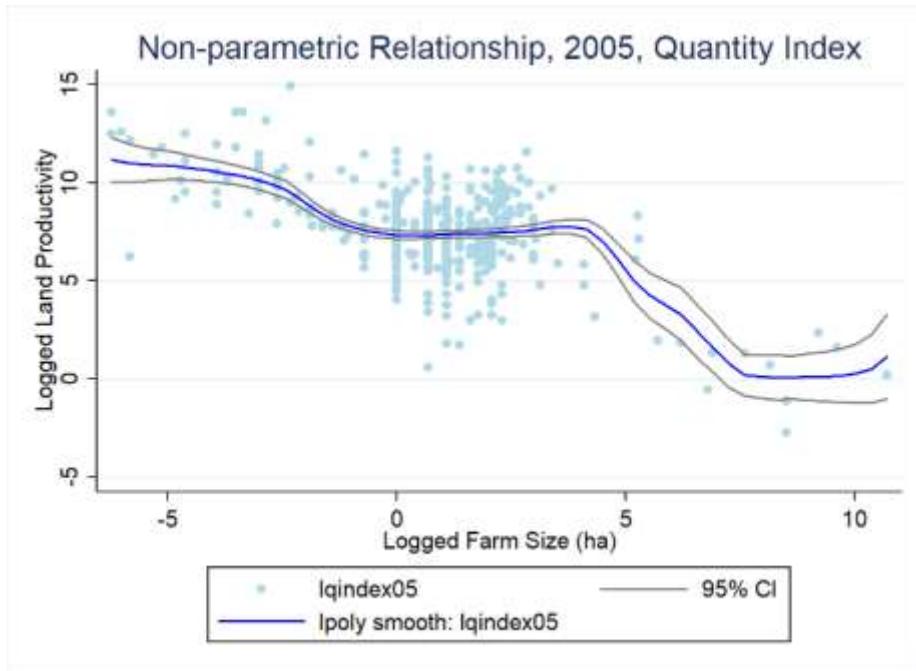
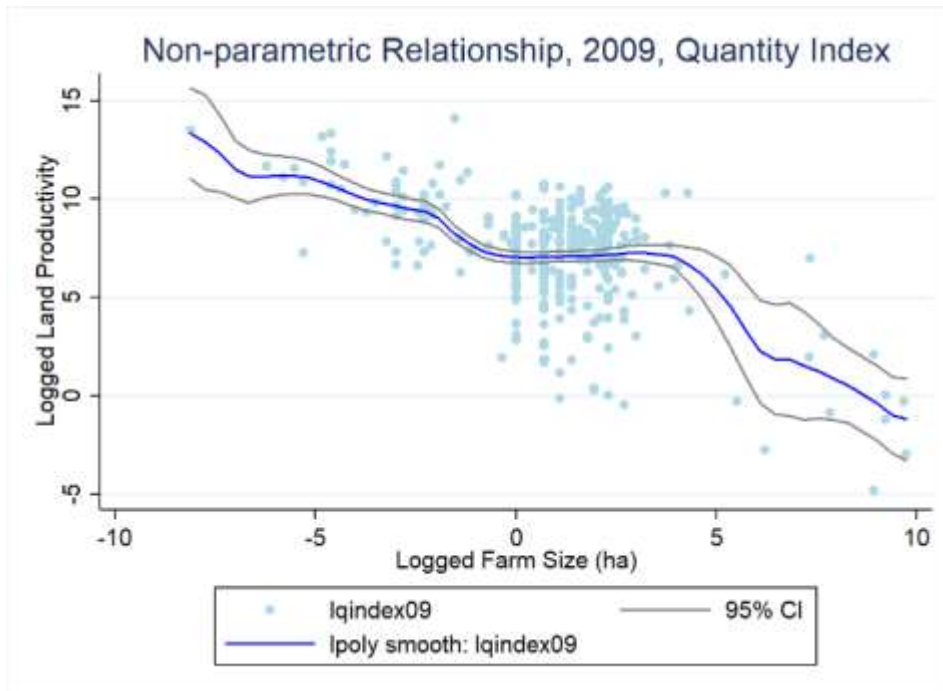


Figure 17: Non-parametric Relationship between Farm Size and Productivity, Panel (2002)



The Use of Inputs

As shown in Helfand and Taylor (2016), land productivity is a partial measure of productivity and does not account for the use of inputs other than land. Where other inputs are used in production, failing to account for the use of those resources potentially introduces bias into estimates of the relationship between farm size and productivity if the intensity of input use (inputs per hectare) varies with farm size. Controlling for all inputs in agricultural production amounts to the estimation of a production function and uncovering TFP, the comprehensive and preferable measure of productivity. The MxFLS offers data on five agricultural inputs other than land: family labor, non-family labor, physical capital, draft animals, and purchased intermediates. Appendix 3 provides a detailed discussion of the source and construction of these input variables.

Table 11 shows the share of households using the different input categories in each year for both the cross sections of farms and the panel. Purchased intermediates are shown collectively, and then also disaggregated into their 9 components. The prevalence of input usage and patterns over time are broadly consistent between the cross sections and panel. Again, this is an encouraging indicator that our treatment of the data has not introduced bias into our sample. For all of the inputs there exist at least some, if not a majority, of households that have zeros for that input category. This is plausible, as farms in the sample are expected to span a range from low technology subsistence agriculture to more modern and input intensive operations. Furthermore, many inputs may be substitutes for each other, and farms can access these inputs by owning them or by purchasing them in factor markets. Tractor services, for example, may be substituted for with draft animals. Households can either own some combination of these capital stocks, or purchase their services on the market. The prevalence of zero inputs necessitates some treatment of the data for a production function to be estimated, and this process is detailed in the empirical methodology below.

Table 11: Percent of Households Using Selected Intermediate Inputs

Input Category	Cross Sections			Panel		
	2002	2005	2009	2002	2005	2009
Family Labor	93%	94%	92%	94%	94%	91%
Non-family Labor	52%	48%	41%	52%	49%	39%
Physical Capital	10%	10%	12%	13%	12%	14%
Draft Animals	32%	28%	22%	35%	30%	27%
Purchased Intermediates	70%	67%	72%	70%	70%	71%
<i>Fertilizer</i>	51%	44%	50%	51%	48%	46%
<i>Manure</i>	17%	15%	18%	17%	17%	18%
<i>Pesticides</i>	31%	24%	31%	33%	26%	27%
<i>Seeds</i>	22%	20%	26%	24%	23%	25%
<i>Tractor Services</i>	29%	23%	34%	32%	25%	36%
<i>Animal Power</i>	3%	11%	11%	3%	10%	11%
<i>Labor</i>	4%	24%	25%	5%	25%	27%
<i>Water</i>	3%	15%	19%	3%	16%	20%
<i>Fuel</i>	2%	9%	14%	2%	10%	15%

Of principle importance is any relationship between inputs per hectare and farm size. We calculate the correlation coefficients between logged input per hectare and logged farm size for those farms with non-zero values of usage of that input. These correlations are shown in Table 12. Conditional on using the input, the intensity (per hectare) of all inputs used declines with farm size, emphasizing the importance of moving from partial measures of productivity to a comprehensive measure such as TFP.

Table 12: Correlation Coefficients between Logged Farm Size and Logged Inputs per ha

Input Category	Farms with Non-zero Values		
	2002	2005	2009
Family Labor	-0.93*	-0.90*	-0.89*
Non-family Labor	-0.90*	-0.91*	-0.89*
Physical Capital	-0.34*	-0.20	-0.28*
Draft Animals	-0.84*	-0.84*	-0.62*
Purchased Intermediates	-0.69*	-0.65*	-0.68*

Note: * indicates statistical significance at the 10% level

Community Variables

The community level questionnaire of the MxFLS surveys several individuals from each community, generating variables with relevance to agricultural productivity.²⁴ First, we examine the existence of farming technological centers and farming technological high schools in the community. While approximately 10% of communities in the survey have one of these types of agricultural institutions, less than 2% of farms in the panel have such institutions in their community.

Second, the prevalence of paved roads in the community proxies for the transaction costs associated with market access. We create a dummy variable equaling 1 if a community's roads are primarily covered in cement or pavement and 0 otherwise. Third, the existence of key institutions in the community, including a public transportation hub, a permanent market, a mobile market, a post office, an elementary school, a secondary school, a high school, a college, a hospital, and a library. Table 1 in Appendix 5 displays the share of panel farms living in communities with paved roads, along with the share of panel farms living in communities with the different institutions within the community. Access to markets at the community level is plausibly exogenous to individual farms in the community, so the permanent and mobile market variables are of particular interest. We combine them to construct an indicator variable for access to a local market. The other important set of institutions are schools, the existence of which reduce the opportunity costs of investing in human capital. Table 2 in Appendix 5 shows the share of farms with paved roads and different institutions in their communities by ejido status. Ejido farms appear less likely to have access to permanent markets, hospitals, libraries, and schools beyond the secondary level.

Household Variables

Additional household level controls are grouped into two broad categories: variables describing agricultural practices that are mostly endogenous, and demographic variables that are largely exogenous.

²⁴ All community level variables are based upon averages across respondents for each community for the relevant year.

We classify a farm that doesn't bring any of its crop to market as a subsistence farm. While endogenous, this measure identifies farms that don't sell their produce to markets and may behave differently than those who do. There is little difference in the prevalence of subsistence farming between ejido and non-ejido farms. As shown in Table 3 in Appendix 5, subsistence farming decreases with farm size, as expected. We calculate the share of each farm's crop that is marketed – on average, those farms in the sample that do participate in the market sell around 75% of their production. This appears relatively constant across farm size bins.

Alongside subsistence farming practices, Table 3 in Appendix 5 shows the share of farms engaging in monocropping. The farms in the sample that monocrop do so on the vast majority of their farm, not just on specific plots. In each survey year over half of the plots being monocropped are growing maize, with approximately 10% each growing beans and coffee. As shown in Table 3 of Appendix 5, there is no discernible difference in monocropping across farm sizes. However, ejido farms appear to consistently employ monocropping more prevalently than do non-ejido farms.²⁵

The MxFLS asks households about their participation in a variety of government programs. The two most important programs for our panel are Procampo and Progres/Oportunidades. Procampo is an income transfer program designed to support agricultural producers of staple crops, stabilizing and increasing farmers' incomes. Progres, later renamed Oportunidades, is a conditional cash transfer program designed to combat poverty and incentivize investments in children. Data limitations do not allow us complete information on participation in Progres²⁶ so we focus exclusively on participation in Procampo. Table 3 in Appendix 5 shows the share of farms participating in Procampo by year and farm size. With the exception of the largest farms, participation increases with farm size. In addition, we consider participation with Alianza, a government-run program designed to aid farmers in diversifying into crops for export. While less than 3% of the sample participated in this program in any survey round, we consider participation in this program for its potentially important impact on farmers.

²⁵ The shares of ejido and non-ejido farms engaging in monocropping are 75% and 65% in 2002, 77% and 68% in 2005, and 76% and 64% in 2009.

²⁶ In the MxFLS, data is available for participation in 2005 only.

The MxFLS also asks households about crop and livestock loss in recent years. To account for the potential that negative productivity shocks may be persistent, we generate dummy variables for whether the household suffered crop loss in either of the previous two years or livestock loss in either of the previous two years. As shown in Table 4 in Appendix 5, livestock loss appears to be much less prevalent than does crop loss in the panel.

The MxFLS provides data on the existence of savings of households, the use of credit over the previous year, and the forms of credit that individuals have access to. Table 5 in Appendix 5 below shows the share of farms having savings, regardless of how those savings are held, alongside whether or not the household used credit in the previous year (again, regardless of the source of the credit), by farm size and year. As expected, the prevalence of having savings increases with farm size, however there is no clear relationship between farm size and having used credit.

These measures capture the behavior of family farms, but because they reflect use, rather than access, they are largely endogenous. Having access to credit, in contrast, is not only plausibly exogenous to the agricultural production function, but the existence of credit constraints and differential access to credit is one theoretical source of a relationship between farm size and productivity. As such, controlling for access to credit rather than use of credit is preferable. Table 6 in Appendix 5 shows “access to credit” by farm size, where a household is considered to have access if the household head knows where they can go to borrow or ask for a loan. Clearly, this is a very crude measure as it does not account for credit rationing and the likelihood that a household could succeed in obtaining a loan. A follow up question regarding the source of that credit allows us to identify if access is through a formal or an informal financial institution.²⁷ There are no clear relationships between farm size and this measure of access to credit. Curiously, nearly all farms had access to some line of credit in 2009. This is likely driven by a notable change in the wording of the question in that survey year. For this reason, we focus on access to formal lines of credit in our analysis which appears more stable.

The household demographic variables are based on the characteristics of the household head. The panel farms predominantly have male, married, and Spanish speaking heads of household, with little differences across farm sizes or ejido status. Table 7 in Appendix 5 shows that farms

²⁷ Formal sources of credit include banks, cooperative savings funds, and government credit programs.

larger than about 5 ha appear to be less likely to have an indigenous household head and more likely to have a literate household head than do smaller farms.

Literacy is just one way to measure educational attainment of the household head, and it captures a rather low bar. We measure the education of the household head by creating indicator variables for the highest level of formal schooling attended, from having no formal education to elementary school, secondary school, high school, and college. With little variation across survey years, Table 8 in Appendix 5 shows educational obtainment by farm size for 2002 only, showing that a majority of farms have household heads with no more than an elementary school education, and almost one quarter of the panel's household heads have no formal education at all.

Empirical Methodology

We take two complementary approaches to explore the relationship between farm size and TFP with the MxFLS data. We use an average production function to estimate average TFP and any relationship with farm size. We also use a stochastic production frontier to estimate TFP along the frontier along with technical inefficiency, identified as deviations from the frontier. The frontier analysis allows us to identify and relationship between farm size and frontier TFP and any relationship between farm size and technical inefficiency. As is standard in the literature (Coelli et al., 2005 and Kumbhakar et al., 2015), we view TFP change as a combination of changes in the technological frontier and changes in the deviations from the frontier.

We estimate a Cobb-Douglas production function with inputs measured per hectare to identify TFP, effectively imposing constant returns to scale on the production technology. In such a setting, the inclusion of a measure of farm size as an explanatory variable identifies any relationship between farm size and TFP (Helfand and Taylor, 2016). We estimate the following production function:

$$y_{it} = \beta_0 + \beta_k \mathbf{x}_{kit} + \theta_t + \gamma_i + f(A_{it}) + \varepsilon_{it} \quad (6)$$

where:

- y_{it} is log of output per ha for farm i in year t

- x_{kit} is log of input per ha for the $k = 1, \dots, 5$ non-land inputs: purchased intermediates, physical capital, draft animals, family labor, and non-family labor
- θ_t are survey year dummies using 2002 as the base
- γ_i are household fixed effects
- $f(A_{it})$ is a measure of farm size, A_{it} , taking one of the following forms: linear, quadratic, cubic, and a flexible dummy variable structure
- ε_{it} is our standard error term

Additional controls, such as land tenure status (ejidos vs. owners), can be added to further explain variation in TFP if they vary over time within farms. We can also include survey year interactions with the measurement of farm size to allow the farm size – productivity relationship to vary over time. Omitting survey year interactions with inputs effectively assumes that the technology is time-invariant, forcing any changes in technology into the TFP term.

The nature of the MxFLS data is such that for all inputs other than land and family labor many households have zero values in any given period. We follow Battese (1997) in our approach to estimating production functions with observations having zero inputs. For each input, k , for each farm in each survey year, we generate a dummy variable, D_{kit} , equaling 1 if there is zero input for that farm in that survey year and zero otherwise. We then define a new measure of the input, x_{kit}^* , equaling 0 if $D_{mit} = 1$ and the log of that input per ha (x_{mit}) otherwise. The inclusion of the dummy variables and newly constructed inputs allows for unbiased estimation of the production function's parameters in the presence of zeros while using the full sample. The final production function to be estimated is:

$$y_{it} = \beta_0 + \mathbf{D}_{kit} + \beta_k \mathbf{x}_{kit}^* + \theta_t + \gamma_i + \delta_{it} \text{Ejido} + f(A_{it}) + \theta_t \times f(A_{it}) + \varepsilon_{it} \quad (7)$$

Our primary approach relies upon the use of household fixed effects to control for time-invariant unobservable determinants of productivity. While the use of household level fixed effects accounts for as much of the time-invariant unobservable variation in productivity as possible, it does not allow for analysis of household level determinants of productivity. Therefore, we use geographical fixed effects as an alternative, allowing for the inclusion of additional explanatory variables. If we utilize community-level or municipality-level fixed effects we can include a set of household-level explanatory variables, \mathbf{Z}_{it} . If we utilize municipality fixed effects (or,

alternatively, state fixed effects) we can additionally include a set of community-level explanatory variables, \mathbf{W}_{ct} . Our second set of production functions introduces household specific controls only using community level fixed effects, γ_c , allowing us to explore household explanations of agricultural productivity differences:

$$y_{ict} = \beta_0 + \mathbf{D}_{kit} + \beta_k \mathbf{x}_{kit}^* + \omega \mathbf{Z}_{it} + \theta_t + \gamma_c + f(A_{it}) + \theta_t \times f(A_{it}) + \varepsilon_{ict} \quad (8)$$

The vector of household specific controls, \mathbf{Z}_{it} , includes dummy variables indicating the practice of monocropping, engaging in subsistence agriculture, crop and livestock loss, Procampo participation, access to formal lines of credit, ejido status, having an indigenous household head, the gender, marital status, and age of the household head, and a series of indicator variables for the last level of formal schooling attended.

Our third set of production functions introduces community level controls, \mathbf{W}_{ct} , while using municipality level fixed effects, γ_m , allowing for the analysis of institutional determinants of agricultural productivity:

$$y_{icmt} = \beta_0 + \mathbf{D}_{kit} + \beta_k \mathbf{x}_{kit}^* + \omega \mathbf{Z}_{it} + \phi \mathbf{W}_{ct} + \theta_t + \gamma_m + f(A_{it}) + \theta_t \times f(A_{it}) + \varepsilon_{icmt} \quad (9)$$

This specification includes variables for whether or not the majority of roads were paved (a proxy for transportation costs), a measure of access to markets using a dummy variable for having a permanent or mobile market in the community, and dummy variables for having a high school, college, or library in the community. In addition, we calculate community-wide participation in Procampo and community-wide prevalence of monocropping and subsistence agriculture, generating plausibly exogenous proxies for the impact of government programs and agricultural practices. These replace the monocropping and Procampo participation variables in equation (8).

The stochastic production frontier approach begins with a measure of technical inefficiency. We take an output oriented approach, measuring technical inefficiency as the difference between what is actually produced by a farm, Y , and the maximum possible production given the inputs used by that farm, $f(\mathbf{X})$:

$$\text{Technical Inefficiency} = \frac{f(\mathbf{X})}{Y} \quad (10)$$

Rearrangement of the log of technical inefficiency generates the following relationship:

$$\ln Y = \ln f(\mathbf{X}) - \ln(\text{Technical Inefficiency}) \quad (11)$$

Econometric estimation of the frontier requires the introduction of an idiosyncratic error term, v , and requires a distributional assumption on the log of technical inefficiency, u :

$$\ln Y = \ln f(\mathbf{X}) + v - u \quad (12)$$

Stochastic production frontier analysis differs from the estimation of an average production function because of the introduction of the two-part error term – a standard idiosyncratic error term coupled with a one-sided error term that measures inefficiency, or deviations from the production frontier.

Such analysis requires the assumption of a functional form of the frontier, a distributional assumption for v , and a distributional assumption for u . Functional forms for the frontier are typically Cobb-Douglas or the more general trans-log functional form, and the standard normal distribution is a regular assumption for v . Common assumptions for the distribution of u include the exponential distribution, the half normal distribution, and a more general truncated normal distribution. Stochastic frontier models allow for the simultaneous estimation of the frontier and heterogeneity in the inefficiency term that is a function of explanatory variables, estimated with maximum likelihood methods.

Greene’s (2005) “true” fixed effects model is probably the most appropriate for our setting, as it allow us to take advantage of panel data by additionally controlling for time invariant unobserved heterogeneity in production with a fixed effects specification. If employed assuming a truncated normal distribution for the inefficiency component of the error term, the model allows both the mean and/or the variance of the inefficiency distribution to depend on a vector of explanatory variables, whereas if a half normal is assumed it allows only for the variance to be a function of such variables (Wang, 2002). Attempts to estimate such a model using the *sfpanel* command in Stata²⁸ were unsuccessful, as the models do not converge regardless of specification.²⁹

²⁸ See Belotti et al. (2012) for a discussion of *sfcross* and *sfpanel*.

²⁹ We suspect that this is because the MxFLS has only three time periods. With an average of just 2.2 observations per household, employing household fixed effects is asking too much of the data in the stochastic frontier setting. We also tried models that continued to use the “true” fixed effects approach with the more restrictive assumption of a half normal distribution for the inefficiency component, and models that maintained the assumption of the truncated normal

We use community fixed effects with a half normal distribution to estimate models with the MxFLS data. While working with household fixed effects would have been preferable, working with community level fixed effects has the advantage of allowing the inclusion of household level explanatory variables. This model allows us to estimate an equation explaining the variance of the half normal inefficiency term in addition to the stochastic frontier. We estimate five alternative specifications, each assuming a Cobb-Douglas functional form for the production frontier with inputs and output measured in logs per unit of land, a normal distribution for the idiosyncratic component of the error term, a half normal distribution for the inefficiency component of the error term, and using dummy variables to implement fixed effects at the community level. Standard errors are clustered at the community level. More formally, the alternative models are given by:

$$y_{ict} = \beta_0 + \boldsymbol{\beta}\mathbf{x}_{it} + \mathbf{D}_{kit} + \boldsymbol{\omega}\mathbf{Z}_{it} + \boldsymbol{\theta}_t + \rho A_{it} + \boldsymbol{\gamma}_c + v_{ict} - u_{it} \quad (13)$$

$$v_{ict} \sim N(0, \sigma_{vc}^2) \quad (14)$$

$$u_{it} \sim N^+(0, \sigma_{uit}^2) \quad (15)$$

$$\sigma_{uit}^2 = \alpha_0 + \alpha_1 A_{it} + \boldsymbol{\theta}_t + \boldsymbol{\varphi}\mathbf{V}_{it} + \epsilon_{it} \quad (16)$$

where \mathbf{x}_{it} are inputs per ha in logs, \mathbf{D}_{kit} are dummy variables for zero values of the inputs, A_{it} is log farm size, \mathbf{Z}_{it} and \mathbf{V}_{it} are vectors of household level controls used in the frontier and inefficiency equations, respectively, $\boldsymbol{\theta}_t$ are time dummies, $\boldsymbol{\gamma}_c$ are community dummies, v_{ict} is the standard normal idiosyncratic component of the error term, u_{it} is the half normal inefficiency component of the error term, and ϵ_{it} is a standard normal error term used in the inefficiency equation. For simplicity, we assume farm size (A_{it}) enters linearly.³⁰

Empirical Results

For reference, we begin our analysis of the farm size – productivity relationship by estimating the linear relationship between land productivity and farm size. Farm size is inversely related to land productivity at the 1% level of significance, as shown in column 1 of Table 13, where we

distribution but employed fixed effects at the community level. None of these approaches would converge with the MxFLS data.

³⁰ The results with farm size dummies are largely equivalent.

estimate the elasticity of land productivity with respect to farm size to be -0.854. We then estimate the average production function identified by equation (7) above, assuming four alternate specifications of the farm size – productivity relationship. These specifications vary in the flexibility of the relationship, ranging from a linear relationship to a flexible relationship captured by farm size dummies for the 7 farm size bins, and are shown in columns 2 through 5 of Table 13. These regressions weight observations by the expansion factors provided by MxFLS (to be discussed further below), use our preferred measure of the family labor index (the first approach in the technical appendix), employ household fixed effects, and cluster standard errors at the household level.

The results indicate an inverse relationship between farm size and TFP, as shown by the negative and statistically significant coefficient on the linear Farm Size variable in model 2. In our sample, a 1% increase in farms size is correlated with a 0.82% decrease in output per hectare, *ceteris paribus*. These coefficients on farm size are slightly less negative than in column 1, but not statistically different, indicating that the relationships between farm size and land productivity and farm size and TFP are almost identical in our sample. This partially reflects the fact that, with the exception of purchased intermediates in model 5, none of the production inputs per ha have a statistically significant relationship with output per ha. The inclusion of fixed effects means that the model is identifying off of variation in input use within, but not across, households. Overall, the coefficients are relatively stable across model specifications.

Models 3 and 4 allow for a quadratic and cubic relationship between farm size and TFP, but the coefficients on the higher ordered terms are not significant (statistically or substantively) and do not have a noticeable impact on the linear model. Model 5 captures some non-linearity in the farm size – TFP relationship by using dummy variables for farm size bins. The smallest of farms, those less than one half of a hectare, are significantly more productive than all other farms, while the largest, those greater than 20 hectares, are significantly less productive than all smaller farms.³¹ Productivity between these two extremes, however, appears relatively stable. This closely mirrors the non-parametric relationships between farm size and land productivity shown in Figures 15 to

³¹ There is an emerging literature on the possibility that measurement in production could be driving the estimated inverse relationship found in so many studies. When independent “crop-cut” measures of production are used, several studies have reported that the inverse relationship disappears. This is an intriguing result that needs additional research to test the external validity of these studies, and whether they hold in settings with much larger farms such as Brazil. See Desiere and Jolliffe (2017) and Gourlay et al. (2017).

17, highlighting the need to assume a flexible functional form to fully understand the farm size – productivity relationship. The linear relationships identified in the parametric specifications 2 through 4 do not capture these subtleties.

Across all models, we see little change in the inverse relationship over time. None of the farm size and survey year interaction terms are statistically significant. There is, however, some evidence for a decline in productivity over time in this sample. The coefficient on our 2005 dummy variable is consistently negative, albeit not statistically significant, whereas the 2009 dummy variable is negative and statistically significant in four of the five specifications. The exception here is model 5, where the decline in TFP over time disappears. Overall, the results show a time invariant inverse relationship between farm size and productivity, with the largest productivity differences observed at the extremes of the size distribution. Curiously, we find some evidence of declining TFP over time in all but model 5 having the dummy variable specification.

Table 13: Estimates of the Production Function

	(1) Linear w/o Inputs	(2) Linear	(3) Quadratic	(4) Cubic	(5) Dummies
Farm Size	-0.854*** (0.076)	-0.821*** (0.103)	-0.808*** (0.099)	-0.773*** (0.143)	
Bin 2: 0.5 to 1 ha					-1.779*** (0.636)
Bin 3: 1 to 2 ha					-2.382*** (0.576)
Bin 4: 2 to 5 ha					-2.072*** (0.587)
Bin 5: 5 to 10 ha					-2.848*** (0.741)
Bin 6: 10 to 20 ha					-2.247** (1.055)
Bin 7: 20+ ha					-6.151*** (1.355)
Family Labor		-0.038 (0.077)	-0.035 (0.082)	-0.032 (0.083)	0.051 (0.103)
Physical Capital		-0.134 (0.170)	-0.139 (0.152)	-0.136 (0.149)	-0.159 (0.213)
Draft Animals		-0.013 (0.102)	-0.015 (0.104)	-0.009 (0.106)	0.026 (0.104)
Purchased Intermediates		0.093 (0.077)	0.092 (0.075)	0.095 (0.075)	0.173* (0.103)
Non-family Labor		0.030 (0.073)	0.036 (0.079)	0.036 (0.078)	0.116 (0.074)
2005 Dummy	-0.315 (0.200)	-0.243 (0.176)	-0.281 (0.187)	-0.292 (0.218)	-0.160 (0.450)
2009 Dummy	-0.486** (0.226)	-0.381* (0.211)	-0.465** (0.221)	-0.445* (0.267)	0.133 (0.632)
Ejido	-0.051 (0.290)	-0.124 (0.252)	-0.138 (0.252)	-0.148 (0.258)	-0.251 (0.342)
2005* Farm Size	0.066 (0.087)	0.081 (0.091)	0.068 (0.095)	0.084 (0.138)	
2009* Farm Size	-0.108 (0.126)	-0.114 (0.116)	-0.129 (0.110)	-0.146 (0.172)	
Farm Size ²			-0.012 (0.016)	-0.007 (0.020)	
2005*Farm Size ²			0.009 (0.015)	0.009 (0.020)	
2009*Farm Size ²			0.017 (0.020)	0.013 (0.031)	
Farm Size ³				-0.001	

2005*Farm Size ³				(0.004)	
				-0.000	
2009*Farm Size ³				(0.004)	
				0.001	
				(0.005)	
2005*Bin 2					-0.208
					(0.705)
2005*Bin 3					0.615
					(0.668)
2005*Bin 4					-0.513
					(0.607)
2005*Bin 5					0.365
					(0.585)
2005*Bin 6					-0.574
					(0.860)
2005*Bin 7					0.053
					(1.057)
2009*Bin 2					-0.672
					(0.848)
2009*Bin 3					-0.562
					(0.793)
2009*Bin 4					-0.759
					(0.888)
2009*Bin 5					-0.844
					(0.849)
2009*Bin 6					-0.867
					(1.060)
2009*Bin 7					-1.828
					(1.686)
Constant	9.626***	11.561***	11.672***	11.601***	11.292***
	(0.307)	(2.034)	(1.924)	(1.958)	(2.681)
Household FE	Yes	Yes	Yes	Yes	Yes
R ²	0.85	0.86	0.86	0.86	0.85
N	1235	1235	1235	1235	1235

Standard errors in parentheses

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

Robustness Tests

In the tables below, we evaluate the robustness of the farm-size – TFP relationship observed in the MxFLS to our treatment of the data. We estimate these models assuming the farm size – TFP relationship is best captured by the linear and dummy variable models above, as the quadratic and cubic models provide little additional information.

In the sensitivity analysis presented here, we explore three decisions made regarding the treatment of the MxFLS data. First, our decisions regarding the construction of the family labor index. Models 1 and 2 in Table 14 show the linear and dummy variable models using our alternative index of family labor (this is our third measure of family labor, as described in Appendix 3). Second, our decision regarding the weighting of observations in our sample. Whereas the core results apply the MxFLS weights designed to make the sample statistically representative of Mexican households in each survey year, models 3 and 4 in Table 14 show the linear and dummy variable specifications when we apply no weighting at all. We explore sensitivity to our use of weights because we are interested in Mexican agriculture, not rural Mexican households, and our treatment of the data reduces the sample size; therefore, it is not clear that these weights remain appropriate. Third, our measurement of the dependent variable – farm output. Our core results use our preferred approach of calculating a quantity index for each household (see Appendix 2 for more detail). Alternatively, we can deflate the nominal value of production in each year for each household and use the real value of output (in 2002 Mexican pesos). Model 5 displays the linear specification of the production function using the real value of output.³²

Overall, these alternative treatments of the data generate qualitatively similar results to those of our core regressions in Table 13 above. This is largely true in terms of the coefficient signs and orders of magnitude. In a number of cases the statistical significance of our estimated coefficients changes, such as for the size dummies in model 2. Employing the alternative family labor index in models 1 and 2 decrease the size and significance of the inverse farm size – TFP relationship, but not enough to change it qualitatively. Models 3 and 4 show that, with the exception of increasing the significance of purchased intermediates, removing sample weights appears to have had little impact on our results. While model 5 shows that our choice of how to measure output does not

³² The results for the dummy variable specification are not shown in this table for the sake of space, however the model was estimated and the results are consistent with those using the quantity index.

affect the farm size – TFP relationship we have identified, it does affect the estimates of average TFP changes over time. When using real output as in model 5, we see a decline in TFP over the period 2002 to 2005, but no difference in average TFP between 2002 and 2009.

In general, the consistency across models reassures us that our treatment of the data is not driving the results. As a similar check, we estimate the regressions from Table 13 and Table 14 using crop production only as our measure of output. These results are shown in Appendix 4, showing that our conclusions regarding the farm size – productivity relationship are robust to this dimension as well.

The finding of declining average TFP over time is curious, as it runs counter to both the long-run country-level analyses in the literature and the micro-level analysis of Kagin et al. (2016) over a similar time period. One important caveat of the MxFLS sample is that it is a household rather than a farm survey, and thus does not include non-family owned commercial farms as does the Brazilian data and other national-level studies. To the extent that such farms have more effectively harnessed the gains from technological change, the potentially heightened productivity of such large farms is not included in this evaluation of the farm size – TFP relationship in Mexican agriculture. This is likely responsible, at least in part, for the finding of both a time-invariant farm size – TFP relationship and declining average TFP over time.

The lack of commercial farms does not, however, reconcile this finding with that of Kagin et al. (2016), who find rising average TFP over a similar period in a different sample of rural households. On the one hand, the MxFLS used here includes more of the larger family farms than does the ENHRUM dataset. If the larger family farms behave differently than smaller family farms then this may help to explain our different findings. Alternatively, the different findings could be in part due to our treatment of the data. The 2002-2009 period saw rapidly rising and then falling agricultural prices, and these fluctuations could play some part in finding declining TFP over time. These prices are central to generating the quantity indices, our preferred approach to measuring output. To check the sensitivity of our results to our measurement of output, we estimated the core regressions using the real value of output in 2002 US dollars. The results are shown in Table 3 of Appendix 4, showing that while the dynamics of the farm size – productivity relationship do not change, the finding of declining average TFP over the 2002-2009 period disappears. Similarly, when using just crop production (Tables 1 and 2 of Appendix 4), there is little evidence of a decline

in TFP over time. Thus, while we remain confident in our findings regarding the farm size – productivity relationship and its dynamics, we should not draw any strong conclusions regarding average TFP over time.

Table 14: Robustness Checks for the Production Function

	(1) Linear, Alt. Labor	(2) Dummies, Alt. Labor	(3) Linear, No Weights	(4) Dummies, No Weights	(5) Linear, Alt. Output
Farm Size	-0.668*** (0.148)		-0.818*** (0.076)		-0.825*** (0.103)
Bin 2: 0.5 to 1 ha		-1.200** (0.594)		-1.705*** (0.543)	
Bin 3: 1 to 2 ha		-1.283** (0.609)		-2.110*** (0.472)	
Bin 4: 2 to 5 ha		-0.938 (0.586)		-2.094*** (0.493)	
Bin 5: 5 to 10 ha		-1.136 (0.692)		-2.902*** (0.600)	
Bin 6: 10 to 20 ha		-0.271 (1.077)		-2.580*** (0.733)	
Bin 7: 20+ ha		-2.834*** (1.037)		-5.366*** (1.039)	
Alt. Family Labor	0.150 (0.148)	0.569*** (0.104)			
Family Labor			-0.009 (0.060)	0.098 (0.079)	-0.044 (0.076)
Physical Capital	-0.144 (0.171)	-0.208 (0.179)	-0.062 (0.098)	-0.070 (0.134)	-0.136 (0.171)
Draft Animals	-0.013 (0.101)	-0.013 (0.094)	0.016 (0.069)	0.082 (0.084)	-0.011 (0.100)
Purchased Intermediates	0.085 (0.082)	0.091 (0.084)	0.148*** (0.057)	0.214*** (0.069)	0.092 (0.077)
Non-family Labor	0.031 (0.074)	0.086 (0.074)	0.012 (0.051)	0.082 (0.058)	0.036 (0.072)
2005 Dummy	-0.262 (0.181)	-0.493 (0.404)	-0.196 (0.141)	-0.150 (0.387)	-0.335* (0.175)
2009 Dummy	-0.418* (0.217)	-0.287 (0.484)	-0.336** (0.169)	0.312 (0.473)	-0.176 (0.210)
Ejido	-0.171 (0.258)	-0.328 (0.264)	-0.152 (0.195)	-0.198 (0.208)	-0.127 (0.254)
2005* Farm Size	0.072 (0.087)		0.064 (0.063)		0.089 (0.090)
2009* Farm Size	-0.108 (0.121)		-0.013 (0.080)		-0.109 (0.117)
2005*Bin 2		0.269 (0.655)		0.070 (0.622)	
2005*Bin 3		0.829 (0.654)		0.114 (0.562)	
2005*Bin 4		-0.015		-0.235	

		(0.575)		(0.499)	
2005*Bin 5		0.635		0.356	
		(0.548)		(0.495)	
2005*Bin 6		-0.234		-0.069	
		(0.812)		(0.607)	
2005*Bin 7		0.724		-0.328	
		(0.869)		(0.934)	
2009*Bin 2		-0.032		-0.950	
		(0.762)		(0.740)	
2009*Bin 3		-0.196		-0.757	
		(0.725)		(0.625)	
2009*Bin 4		-0.215		-0.688	
		(0.742)		(0.632)	
2009*Bin 5		-0.483		-0.694	
		(0.717)		(0.626)	
2009*Bin 6		-0.354		-0.486	
		(0.977)		(0.622)	
2009*Bin 7		-1.069		-1.130	
		(1.506)		(1.260)	
Constant	11.509***	12.409***	10.032***	9.602***	11.593***
	(1.923)	(2.028)	(1.307)	(1.885)	(2.035)
Sample Weights	Yes	Yes	No	No	Yes
Household FE	Yes	Yes	Yes	Yes	Yes
R ²	0.86	0.86	0.81	0.78	0.86
N	1235	1235	1235	1235	1235

Standard errors in parentheses

* p<0.10, ** p<0.05, *** p<0.01

The models estimated above only include the ejido dummy as a shift of the intercept. We further explored models that have the ejido dummy variable a) interacting with the factors of production, b) changing over time, and c) interacting with the measures of farm size. The results have not been shown here as none of these interaction terms were statistically significant, indicating that ejidos do not use inputs in a statistically different way than do non-ejido farms. Similarly, the farms size – TFP relationship does not appear to vary between ejido and non-ejido farms.

Additional Controls

Tables 15 and 16 below show regression results from the models employing community level and municipality level fixed effects, respectively. The estimated production function and the farm size – productivity relationship in these models remains quite similar to those identified in Table 13 using household level fixed effects. The linear, quadratic, cubic, and dummy variable specifications all show an inverse relationship between farm size and TFP. Unlike in previous models, the quadratic and cubic specifications now capture some of the non-linearity in the data when using community and municipality fixed effects in Tables 15 and 16 (although only at the 10% level of significance). The difference with the previous tables is that the models in Tables 15 and 16 identify the farm size – productivity relationship using across-farm within-community and across-farm within-municipality variation in Tables 15 and 16, respectively.

The dummy variable estimates continue to indicate a non-linear inverse relationship between farm size and TFP. As in the household fixed effect models, the smallest farms are statistically and substantively more productive than all the other farms. The next most productive farms are those between 0.5 and 1 ha, followed by an approximately constant TFP across farms between 1 and 20 ha. The productivity of the largest farms, those 20 ha or greater, falls sharply. These models support the conclusion that the farm size –TFP relationship is inverse, non-linear, and time invariant, and that it is the extremely large and extremely small farms in the sample that are primarily driving the relationship.

The inclusion of household explanatory variables in the community fixed effects specification of Table 15 shows that both monocropping (most frequently of maize) and operating as a subsistence farm are consistently negatively correlated with TFP. In contrast, participating in Procampo is positively correlated with productivity (as is participation with Alianza, although the relationship is not statistically significant). It is important to reiterate that these are potentially endogenous explanatory variables, and we should not interpret the coefficients as identifying causal relationships. Having more education is positively related to TFP, however with the exception of a college education these impacts are not consistently statistically significant at standard levels.

The coefficients on the explanatory variables in the municipality fixed effects models of Table 16 are broadly consistent with these findings. In this model, Procampo participation, Alianza

participation, monocropping, subsistence agriculture, and access to formal lines of credit are all measured at the community level, generating plausibly exogenous controls for the household specific variables used in Table 15.³³ Procampo remains positively correlated with TFP in Table 16. Monocropping is no longer statistically significant when measured at the community level, however the prevalence of subsistence agriculture remains negative and statistically significant. In addition to college education being positively correlated with TFP, having some secondary education is also now positively and significantly correlated with TFP. The other community level variables – having paved roads, more access to credit in the community, and a market present in town – all have insignificant correlations with TFP. One exception is the presence of agricultural teaching institution in the community, which is positively and significantly correlated with agricultural TFP.

To summarize, we find that education, participation in Procampo, operating as a subsistence farm, the existence of agricultural teaching institutions, and monocropping contribute to explaining differentials in agricultural TFP. To explore their impact on the farm size – productivity relationship we interact them with farm size. We assume the farm size – TFP relationship to be time invariant, and simply interact these household indicator variables, along with ejido status and access to formal credit, with farm size. For simplicity, we assume linearity in the farm size – TFP relationship. The results are displayed in Table 17. Overall, the farm size – TFP relationship identified in Tables 15 and 16 remains. Furthermore, none of these additional controls contribute to explaining the farm size – TFP relationship that we have identified. When the interactions were specified with the size dummies (not shown here), the overwhelming majority were statistically insignificant and no discernible pattern was observed for the few coefficients that were statistically significant.

In addition, we introduce controls for various forms of property rights in Table 18. Models 1 and 2 both include dummy variables for farms having property rights secured by a deed/title and by a certificate of agricultural rights, each of which are sequentially interacted with farm size in the respective models. As these variables and their interactions are not statistically significant, model 3 includes a dummy variable for having property rights being secured by any form of

³³ The community level measures are the share of MxFLS farms in the community that participate in the given practice.

documentation and includes an interaction with farm size, neither of which are statistically significant. Model 4 in Table 18 tests for regional differences in the inverse relationship, interacting with farm size an indicator variable for farms located in the more commercially oriented agricultural region in the North of Mexico. The statistical insignificance of this interaction term indicates that the inverse relationship between farm size and TFP persists regardless of regional differences in the agricultural sectors.³⁴

³⁴ The use of household fixed effects and the time invariant nature of farm location do not allow for the inclusion of regional indicator variables. Models with regional indicator variables interacting with controls, however, are permissible. A series of such regional models were run, including a model with household fixed-effects and a model allowing for differential dynamic farm size – TFP relationships, all of which resulted in similar conclusions of a homogenous farm size – TFP relationship across regions.

Table 15: Community Fixed Effects with Household Controls

	(1) Linear w/o Inputs	(2) Linear	(3) Quadratic	(4) Cubic	(5) Dummies
Farm Size	-0.822*** (0.037)	-0.814*** (0.067)	-0.796*** (0.059)	-0.774*** (0.070)	
Bin 2: 0.5 to 1 ha					-1.564*** (0.260)
Bin 3: 1 to 2 ha					-2.211*** (0.247)
Bin 4: 2 to 5 ha					-2.150*** (0.272)
Bin 5: 5 to 10 ha					-2.555*** (0.382)
Bin 6: 10 to 20 ha					-2.389*** (0.548)
Bin 7: 20+ ha					-5.236*** (0.967)
Family Labor		-0.026 (0.045)	-0.039 (0.047)	-0.038 (0.047)	0.114** (0.051)
Physical Capital		-0.044 (0.082)	-0.046 (0.083)	-0.043 (0.082)	0.078 (0.070)
Draft Animals		0.036 (0.041)	0.027 (0.044)	0.024 (0.048)	0.067 (0.045)
Purchased Intermediates		0.070 (0.051)	0.071 (0.051)	0.070 (0.051)	0.194*** (0.074)
Non-family Labor		0.009 (0.045)	0.015 (0.054)	0.011 (0.052)	0.120** (0.046)
2005 Dummy	-0.293* (0.172)	-0.206 (0.162)	-0.261 (0.193)	-0.289 (0.227)	-0.199 (0.211)
2009 Dummy	-0.666*** (0.124)	-0.544*** (0.118)	-0.631*** (0.152)	-0.665*** (0.187)	0.205 (0.517)
Ejido	0.037 (0.149)	0.033 (0.149)	-0.001 (0.141)	0.003 (0.142)	-0.112 (0.152)
2005* Farm Size	0.020 (0.044)	0.032 (0.050)	0.016 (0.046)	0.055 (0.089)	
2009* Farm Size	-0.072 (0.071)	-0.068 (0.074)	-0.092 (0.072)	-0.064 (0.109)	
Farm Size ²			-0.018* (0.009)	-0.014 (0.012)	
2005*Farm Size ²			0.013 (0.012)	0.016 (0.017)	
2009*Farm Size ²			0.020 (0.013)	0.024 (0.020)	
Farm Size ³				-0.001	

				(0.002)	
2005*Farm Size ³				-0.001	
				(0.003)	
2009*Farm Size ³				-0.001	
				(0.003)	
2005*Bin 2					-0.227
					(0.556)
2005*Bin 3					0.597
					(0.509)
2005*Bin 4					-0.398
					(0.306)
2005*Bin 5					0.154
					(0.263)
2005*Bin 6					0.101
					(0.580)
2005*Bin 7					-0.419
					(0.846)
2009*Bin 2					-0.898
					(0.677)
2009*Bin 3					-0.901
					(0.740)
2009*Bin 4					-0.859
					(0.629)
2009*Bin 5					-1.041
					(0.769)
2009*Bin 6					-1.287
					(0.833)
2009*Bin 7					-1.569
					(1.359)
Monocrop	-0.379**	-0.374**	-0.370**	-0.354**	-0.440***
	(0.150)	(0.154)	(0.151)	(0.155)	(0.143)
Crop Loss	-0.040	-0.019	-0.019	-0.029	-0.050
	(0.272)	(0.236)	(0.230)	(0.235)	(0.224)
Livestock Loss	0.521	0.424	0.382	0.418	0.289
	(0.334)	(0.341)	(0.327)	(0.325)	(0.357)
Procampo	0.481***	0.401**	0.388**	0.383**	0.286*
	(0.146)	(0.156)	(0.156)	(0.157)	(0.163)
Alianza	0.477	0.272	0.324	0.325	0.256
	(0.363)	(0.343)	(0.349)	(0.347)	(0.342)
Formal Credit	0.094	0.052	0.044	0.037	0.123
	(0.275)	(0.261)	(0.268)	(0.266)	(0.246)
Age	0.001	-0.002	-0.002	-0.002	-0.001
	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)
Male	0.043	0.018	0.028	0.030	-0.004
	(0.245)	(0.209)	(0.213)	(0.214)	(0.204)
Married	0.227	0.170	0.168	0.153	0.163

	(0.217)	(0.190)	(0.191)	(0.192)	(0.192)
Indigenous	-0.065	-0.084	-0.092	-0.111	-0.159
	(0.192)	(0.204)	(0.203)	(0.206)	(0.225)
Elementary School	0.103	0.023	0.004	0.011	0.041
	(0.167)	(0.161)	(0.161)	(0.161)	(0.175)
Secondary School	0.604*	0.380	0.341	0.359	0.168
	(0.354)	(0.413)	(0.414)	(0.409)	(0.478)
High School	0.390	0.317	0.333	0.371	0.076
	(0.332)	(0.330)	(0.342)	(0.336)	(0.403)
College	1.330**	1.312**	1.296**	1.341**	0.809
	(0.560)	(0.565)	(0.575)	(0.572)	(0.533)
Subsistence	-0.566***	-0.442***	-0.439***	-0.436***	-0.279
	(0.180)	(0.155)	(0.156)	(0.159)	(0.182)
Constant	7.908***	9.230***	9.913***	10.427***	6.183***
	(0.455)	(1.102)	(1.299)	(1.505)	(1.111)
Community FE	Yes	Yes	Yes	Yes	Yes
R ²	0.70	0.72	0.72	0.72	0.68
N	1235	1235	1235	1235	1235

Standard errors in parentheses

* p<0.10, ** p<0.05, *** p<0.01

Table 16: Municipality Fixed Effects with Household and Community Controls

	(1) Linear w/o Inputs	(2) Linear	(3) Quadratic	(4) Cubic	(5) Dummies
Farm Size	-0.798*** (0.038)	-0.706*** (0.039)	-0.683*** (0.034)	-0.620*** (0.058)	
Bin 2: 0.5 to 1 ha					-1.815*** (0.356)
Bin 3: 1 to 2 ha					-2.426*** (0.228)
Bin 4: 2 to 5 ha					-2.370*** (0.241)
Bin 5: 5 to 10 ha					-2.529*** (0.289)
Bin 6: 10 to 20 ha					-2.750*** (0.426)
Bin 7: 20+ ha					-5.740*** (0.713)
Family Labor		0.062*** (0.020)	0.060*** (0.021)	0.060*** (0.021)	0.076*** (0.024)
Physical Capital		0.104*** (0.031)	0.108*** (0.030)	0.099*** (0.028)	0.123*** (0.027)
Draft Animals		0.012 (0.015)	0.012 (0.015)	0.012 (0.016)	0.003 (0.021)
Purchased Intermediates		0.069*** (0.020)	0.069*** (0.020)	0.066*** (0.021)	0.100*** (0.025)
Non-family Labor		0.064 (0.047)	0.070 (0.057)	0.057 (0.059)	0.193*** (0.051)
2005 Dummy	-0.400** (0.184)	-0.305* (0.170)	-0.340* (0.194)	-0.353 (0.242)	-0.371 (0.257)
2009 Dummy	-0.613*** (0.167)	-0.498*** (0.147)	-0.570*** (0.182)	-0.633*** (0.203)	0.249 (0.402)
Ejido	0.027 (0.196)	0.066 (0.190)	0.037 (0.185)	0.054 (0.182)	-0.045 (0.193)
2005* Farm Size	0.040 (0.041)	0.019 (0.046)	0.008 (0.047)	0.044 (0.085)	
2009* Farm Size	-0.062 (0.064)	-0.068 (0.066)	-0.093 (0.065)	-0.043 (0.082)	
Farm Size ²			-0.015** (0.007)	-0.001 (0.013)	
2005*Farm Size ²			0.008 (0.011)	0.007 (0.018)	
2009*Farm Size ²			0.017 (0.011)	0.022 (0.018)	
Farm Size ³				-0.003	

				(0.002)	
2005*Farm Size ³				-0.000	
				(0.003)	
2009*Farm Size ³				-0.001	
				(0.003)	
2005*Bin 2					-0.204
					(0.680)
2005*Bin 3					0.745
					(0.546)
2005*Bin 4					-0.264
					(0.348)
2005*Bin 5					-0.025
					(0.361)
2005*Bin 6					0.406
					(0.556)
2005*Bin 7					0.015
					(0.766)
2009*Bin 2					-1.056
					(0.664)
2009*Bin 3					-0.898
					(0.744)
2009*Bin 4					-0.732
					(0.598)
2009*Bin 5					-0.797
					(0.739)
2009*Bin 6					-0.990
					(0.690)
2009*Bin 7					-1.005
					(1.133)
Monocrop	0.598	0.434	0.401	0.356	-0.192
	(0.487)	(0.418)	(0.414)	(0.390)	(0.531)
Subsistence	-1.764***	-1.596***	-1.602***	-1.570***	-1.752***
	(0.477)	(0.494)	(0.494)	(0.529)	(0.444)
Crop Loss	-0.062	-0.076	-0.078	-0.093	-0.173
	(0.267)	(0.246)	(0.242)	(0.245)	(0.240)
Livestock Loss	0.618**	0.509*	0.490*	0.552*	0.388
	(0.287)	(0.289)	(0.284)	(0.279)	(0.367)
Procampo	1.230**	1.207**	1.198**	1.106*	1.005
	(0.549)	(0.556)	(0.554)	(0.571)	(0.614)
Alianza	0.313	0.218	0.165	0.375	-0.038
	(0.939)	(0.867)	(0.918)	(0.907)	(1.413)
Formal Credit	-0.180	-0.342	-0.362	-0.402	-0.454
	(0.643)	(0.601)	(0.602)	(0.586)	(0.675)
Age	0.011*	0.007	0.007	0.006	0.005
	(0.006)	(0.006)	(0.006)	(0.006)	(0.007)
Male	0.055	-0.026	-0.017	0.002	-0.098

	(0.244)	(0.220)	(0.223)	(0.224)	(0.208)
Married	0.204	0.128	0.126	0.083	0.124
	(0.209)	(0.204)	(0.204)	(0.199)	(0.200)
Indigenous	-0.215	-0.229	-0.238	-0.259	-0.345
	(0.251)	(0.245)	(0.245)	(0.238)	(0.260)
Elementary School	0.276	0.191	0.177	0.184	0.111
	(0.167)	(0.156)	(0.154)	(0.156)	(0.157)
Secondary School	0.993***	0.853**	0.818**	0.831**	0.605
	(0.348)	(0.343)	(0.343)	(0.336)	(0.463)
High School	0.512	0.438	0.441	0.509	0.235
	(0.357)	(0.359)	(0.373)	(0.354)	(0.448)
College	1.686***	1.801***	1.762***	1.832***	1.373***
	(0.446)	(0.417)	(0.425)	(0.421)	(0.441)
Paved Roads	-0.178	-0.115	-0.112	-0.086	0.074
	(0.200)	(0.193)	(0.196)	(0.193)	(0.192)
Market Present	-0.067	-0.133	-0.129	-0.123	-0.186
	(0.202)	(0.187)	(0.188)	(0.189)	(0.212)
Farm School	0.851*	0.915*	0.906*	0.924*	0.663
	(0.441)	(0.463)	(0.459)	(0.467)	(0.717)
Constant	8.561***	7.663***	7.769***	7.630***	10.147***
	(0.598)	(0.620)	(0.616)	(0.644)	(0.540)
Municipality FE	Yes	Yes	Yes	Yes	Yes
R ²	0.65	0.68	0.68	0.68	0.64
N	1235	1235	1235	1235	1235

Standard errors in parentheses

* p<0.10, ** p<0.05, *** p<0.01

Table 17: Community Fixed Effects with Household Control Interactions

	(1) Monocrop	(2) Ejido	(3) Procampo	(4) Access to Formal Credit	(5) Higher Education	(6) Subsistence
Farm Size	-0.828*** (0.090)	-0.826*** (0.066)	-0.840*** (0.064)	-0.821*** (0.060)	-0.818*** (0.059)	-0.821*** (0.070)
Farm Size*Monocrop	0.010 (0.071)					
Farm Size*Ejido		0.018 (0.075)				
Farm Size*Procampo			0.075 (0.069)			
Farm Size*Access				0.013 (0.083)		
Farm Size*Education					-0.072 (0.088)	
Farm Size *Subsistence						-0.000 (0.059)
2005 Dummy	-0.213 (0.142)	-0.192 (0.150)	-0.228 (0.138)	-0.199 (0.145)	-0.203 (0.148)	-0.155 (0.142)
2009 Dummy	-0.596*** (0.097)	-0.592*** (0.098)	-0.644*** (0.102)	-0.591*** (0.097)	-0.538*** (0.110)	-0.582*** (0.096)
Community FE	Yes	Yes	Yes	Yes	Yes	
R ²	0.72	0.72	0.72	0.72	0.72	0.72
N	1235	1235	1235	1235	1235	1235

Standard errors in parentheses

* p<0.10, ** p<0.05, *** p<0.01

Table 18: Community FE with Property Rights and Regional Interactions

	(1) Title/Deed	(2) Certificate of Rights	(3) Any Documentation	(4) North
Farm Size	-0.813*** (0.070)	-0.817*** (0.066)	-0.809*** (0.106)	-0.819*** (0.054)
Farm Size*Deed	-0.008 (0.048)			
Farm Size*Certificate		-0.000 (0.060)		
Farm Size*Documentation			-0.013 (0.078)	
Farm Size*North				-0.020 (0.209)
2005 Dummy	-0.193 (0.154)	-0.195 (0.159)	-0.192 (0.155)	-0.192 (0.157)
2009 Dummy	-0.586*** (0.102)	-0.587*** (0.107)	-0.589*** (0.106)	-0.586*** (0.100)
Community FE	Yes	Yes	Yes	Yes
R ²	0.72	0.72	0.72	0.72
N	1235	1235	1235	1235

Standard errors in parentheses

* p<0.10, ** p<0.05, *** p<0.01

Frontier Analysis

Estimating a stochastic frontier complements our analysis of the average production function by identifying productivity at the frontier and production inefficiencies. Together, these components determine average TFP identified with the average production function. In similar fashion, whereas the estimation of the average production function allows us to assess the relationship between farm size and average productivity, stochastic frontier analysis allows us to assess any relationships between farm size and productivity at the technical frontier and between farm size and technical inefficiency.

The results of five specifications of the stochastic production frontier are shown in Table 19, with the top and bottom panels displaying the results from the frontier and variance of inefficiency equations, respectively. Model 1 is our baseline model with no additional controls in either the

frontier (Z) or the inefficiency equations (W). Model 2 includes dummy variables for the household head's level of education in the frontier equation and includes a dummy variable for the household head being of indigenous ethnicity in the inefficiency equation. Model 3 alternatively assumes that education of the household head should be included as a control in the inefficiency equation but not the frontier equation. Model 4 assumes that education belongs in both equations. Model 5 includes education in the frontier equation only, but adds interaction terms between farm size and the survey year dummies in both the frontier and the inefficiency equations.

The estimated coefficients from models 1 – 5 in Table 19 are largely consistent. They indicate a strong inverse relationship between farm size and frontier TFP, and that the frontier is increasing over time, reflecting positive technical change. The coefficients on inputs are positive and stable across specifications, with family labor and purchased intermediates being significant. Thus, differential use of inputs, primarily of family labor and intermediates, is an important explanation for different levels of output for the most productive farms. The variance of the inefficiency term σ_u^2 is roughly double the size of the variance of the noise σ_v^2 in all models, and lambda – which is the ratio of the two variances – indicates that estimation of a stochastic frontier is appropriate with the MxFLS data. An excessively small (or large) lambda is indicative of the model struggling to differentiate between the two components of the error term, and this is not the case in Table 19.³⁵

The models indicate a stable inverse relationship between farm size and productivity at the frontier of the technology, with negative coefficients around -0.64 in the time invariant models 1 through 4. This inverse relationship is of the same magnitude as what was estimated with the average production function in the tables above. The coefficients on the survey year dummies in Table 19 are all positive and significant, indicating that the frontier is increasing over time. Thus, in contrast to the results of our average production function analysis above, here we find evidence of positive technical change at the frontier even though we did not find rising TFP on average. The interactions between farm size and the survey year dummies in model 5 identify a positive and significant relationship between farm size and technical change. This suggests that technical

³⁵ In models estimated with a constant variance of the inefficiency distribution (σ_u^2), and thus no explanatory variables in equation (7), Stata provides a p-value for the test of lambda equal to zero. This hypothesis is rejected at greater than the 1% level of significance, providing evidence in support of the stochastic frontier model.

change has been biased towards larger farms, and that the inverse relationship along the frontier became less steep over time.

When education of the household head is included in the frontier specifications but not in the explanation of inefficiency (models 2 and 5), having secondary education or a college education is positively associated with higher levels of productivity among the frontier producers. Similarly, when education is included only in the inefficiency equation (model 3), secondary and college education reduce the variance of the one-sided inefficiency term, a result quite similar to the result in models 2 and 5. However, when education is included in both the frontier and inefficiency equations (model 4), none of the education dummies are significant as the model appears to struggle to identify the separate relationships with education.

Models 1 through 4 find that, while the variance of the inefficiency distribution increased over time, there is no relationship between farm size and inefficiency. The inclusion of interactions between farm size and survey year dummy variables in model 5 reveals a dynamic relationship between farm size and technical inefficiency that is more nuanced. The estimates indicate that larger farms were more efficient than smaller farms in 2002 (i.e. they operated closer to the frontier). While inefficiency is increasing with time across the entire farm size distribution (the coefficients on the 2005 and 2009 dummies are 1.20 and 1.87 in model 5), it is increasing even faster for larger farms (the coefficients on the year*size interactions are 0.37 and 0.42). These differential changes in inefficiency across the farm size distribution have caused the farm size - inefficiency relationship to disappear in the latter waves of the MxFLS.³⁶

Taken with the results from the frontier equation, model 5 reveals that rising technical inefficiency has accompanied technological change, suggesting that the majority of farms have been unable to keep with the TFP growth of the most productive farms, and in particular of the larger farms operating along the frontier. This increased inefficiency is driving the decline of TFP over time identified in the average production function estimates. In models not shown here, we estimate a stochastic frontier including the household controls from Table 15 as explanatory variables of the inefficiency term. We find that, in addition to educational achievements of the

³⁶ This can be seen by adding the farm size coefficient (-0.32) in model 5 with the year*size interaction from 2005 (0.37) or 2009 (0.42). In either case, the sum of the two coefficients is not statistically significantly different from zero.

household head, technical inefficiency is lower among Procampo participants and higher among farms practicing monocropping. When interacted with farm size, none of the interaction terms are statistically significant, suggesting that they do not fundamentally change the relationships observed in Table 19.

These findings are also consistent with the time invariant inverse relationship between farm size and TFP. Along the frontier, the inverse relationship between farm size and productivity is becoming less pronounced as technological advance has favored larger farms. At the same time, this has been offset by growing inefficiencies among larger farms. Inefficiency was initially smaller for larger farms, but this was no longer the case in the later waves of the MxFLS. The combination of these two forces has resulted in a farm size – TFP relationship that is relatively stable over time. These findings suggest that not all large farms are participating equally in adopting and effectively harnessing technological change.

Table 19: Stochastic Frontier Production Function Results

	(1)	(2)	(3)	(4)	(5)
Frontier Equation					
Farm Size	-0.642*** (0.051)	-0.653*** (0.050)	-0.646*** (0.053)	-0.638*** (0.052)	-0.805*** (0.062)
2005 Dummy	0.477** (0.186)	0.475** (0.177)	0.446** (0.174)	0.458** (0.186)	0.400** (0.201)
2009 Dummy	0.790*** (0.212)	0.799*** (0.207)	0.769*** (0.201)	0.765*** (0.210)	0.711*** (0.223)
2005*Farm Size					0.192*** (0.064)
2009*Farm Size					0.204* (0.108)
Family Labor	0.077** (0.032)	0.077** (0.033)	0.072** (0.034)	0.074** (0.034)	0.068** (0.033)
Physical Capital	0.008 (0.047)	0.012 (0.042)	0.016 (0.044)	0.019 (0.045)	0.037 (0.046)
Draft Animals	0.028 (0.034)	0.026 (0.032)	0.022 (0.032)	0.023 (0.033)	0.006 (0.030)
Purchased Intermediates	0.148*** (0.038)	0.139*** (0.038)	0.146*** (0.040)	0.148*** (0.039)	0.145*** (0.041)
Non-family Labor	0.045 (0.034)	0.041 (0.034)	0.051 (0.034)	0.053 (0.034)	0.024 (0.033)
Elementary School		0.044 (0.090)	-0.048 (0.142)		0.057 (0.094)
Secondary School		0.517** (0.205)	0.293 (0.332)		0.531** (0.209)
High School		0.008 (0.204)	-0.069 (0.344)		0.083 (0.207)
College		0.703** (0.307)	-0.334 (0.494)		0.699** (0.302)
Inefficiency Equation					
Farm Size	0.037 (0.062)	0.031 (0.060)	0.035 (0.060)	0.040 (0.061)	-0.317*** (0.119)
2005 Dummy	1.152*** (0.377)	1.163*** (0.361)	1.112*** (0.359)	1.146*** (0.403)	1.198*** (0.430)
2009 Dummy	1.838*** (0.407)	1.878*** (0.387)	1.840*** (0.379)	1.871*** (0.431)	1.870*** (0.401)
2005*Farm Size					0.368** (0.149)
2009*Farm Size					0.417** (0.167)
Indigenous		0.001 (0.230)	-0.038 (0.238)	-0.046 (0.245)	0.001 (0.233)

Elementary School			-0.193 (0.271)	-0.142 (0.193)	
Secondary School			-0.499 (0.549)	-0.882** (0.411)	
High School			-0.190 (0.819)	-0.118 (0.579)	
College			-1.680* (0.451)	-1.603*** (0.480)	
$E(\sigma_u^2)$	1.679	1.666	1.661	1.641	1.620
σ_u^2	0.846	0.840	0.838	0.852	0.853
λ	1.985	1.983	1.982	1.926	1.899
N	1235	1235	1235	1235	1235

Standard errors in parentheses

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

Discussion

The analysis of Mexican data reveals an inverse and time-invariant relationship between farm size and TFP. This confirms earlier empirical work by Kagin et al. (2016), who estimate both an average production function and a stochastic production frontier using a different panel of Mexican family farms (the ENHRUM sample used has two survey years, 2003 and 2008). As with our analysis, their estimates with fixed effects clearly show inverse relationships between farm size and TFP and between farm size and frontier productivity. Similarly, they also find that both technical change and technical inefficiency increased over time.

Unlike previous research, we show that the relationship between farm size and TFP is non-linear, with relatively high productivity of the smallest farms (those less than 0.5 ha) and relatively low productivity of the largest farms (those greater than 20 ha) driving the results. This emphasizes the limitations of imposing a linear function form on the farm size – productivity relationship. Further, our estimate of declining average TFP over time—in some though not all models--suggests that increased technical inefficiency may have dominated technical change, whereas Kagin et al. (2016) find that average TFP increased over a similar time period. In their stochastic production frontier model with fixed effects, Kagin et al. (2016) find that farm size is directly related to technical inefficiency; we find the opposite in 2002, and that this relationship disappeared in later years due to more rapidly growing inefficiency at the upper end of the farm

size distribution. More rapid technical change at the upper end of the farm size distribution indicates an advantage for some larger farms in harnessing more modern agricultural practices, yet this advantage has not been widespread enough to translate into increased TFP due to the inability of the non-frontier households to keep up. Thus, we provide evidence of a more nuanced and dynamic relationship between farm size and technical inefficiency and between farm size and productivity at the frontier.

Our results are complemented by previous work on the farm size – productivity relationship in Brazil. Whereas the Brazilian experience suggests a dynamic farm size – TFP relationship, with an inverse relationship in traditional agriculture becoming flat and potentially positive with modernization, we observe no such dynamics in the Mexican sample. The relationship observed in the MxFLS is time invariant and persistently inverted, contrasting with the emerging U-shaped relationship observed in the modernizing regions of Brazil. It is quite similar, however, to the more traditional agricultural regions in Brazil that display a persistent inverse relationship between farm size and TFP. The frontier analysis using MxFLS data finds that technical change has been biased towards larger farms, weakening the farm size – productivity relationship at the frontier. This indicates that if inefficiency had not increased, the average inverse relationship between farm size and productivity would have weakened, if not reversed, with modernization of the agricultural sector.

Participation in Procampo, increased education, and the presence of agricultural educational institutions are found to be positively correlated with the agricultural productivity of Mexican family farms. The practices of monocropping and operating as a subsistence farm are found to be negatively correlated with TFP. We are tentative in drawing stronger conclusions about the causal impact of these variables, as these controls are likely endogenous. Inclusion of these variables either additively or interacted with farm size, however, does not alter the farm size – TFP relationship that we have identified.

Both the Brazilian experience and the Mexican analysis suggest roles for extension services. In Brazil, we've seen that different policies for large and small farms were effective in contributing to productivity growth, with extension services more important for the large farms and education important for all size farms. In Mexico, we estimate a positive correlation between TFP and the presence of agricultural schools, providing some direct evidence for the role of such institutions.

Less directly, the growing technical inefficiency that we observe in Mexico indicates the potential for policies designed to promote and support the adoption and efficient use of best practices to achieve gains in agricultural productivity. Even as technical change has occurred, TFP has been relatively stagnant due to the increasing inefficiencies, emphasizing that policies helping to move farms towards the frontier would be an effective approach for increasing agricultural productivity. Our finding that participation in Procampo is correlated with lower technical inefficiency in Mexico is evidence that policies designed to increase the competitiveness and productivity of agriculture can be effective.

We consistently estimate an inverse relationship between farm size and average TFP in all periods of the Mexican sample, yet the analysis casts doubt on the ability to exploit the existing inverse relationship between farm size and TFP to generate productivity gains. The production frontier is expanding more rapidly for larger farms, suggesting that successfully reducing technical inefficiency could mediate, if not reverse, the farm size - productivity relationship. Policies designed to reduce inefficiencies, such as extension services, should be a priority. The Mexico analysis shows that a significant share of farms do not have formal documentation of property rights. Policies to ensure that farms have the necessary documentation would help provide farms with the opportunity to keep abreast of technical change, as documented property rights are an important condition for accessing credit and thus facilitating adoption.

While the Mexico analysis indicates the potential for larger farms to be the key drivers of future productivity growth in Mexico, the Brazilian experience shows a U-shaped relationship between agricultural productivity growth and farm size, where small farms experienced relatively rapid productivity growth compared to medium sized farms. This is in part due to increased specialization and education amongst the farms with the most rapid gains in productivity. Policies geared towards smaller family farms may not have large returns in terms of increasing overall agricultural productivity, but they are likely very important for poverty reduction. Even if small farms generate an increasingly smaller share of agricultural output, they are likely here to stay in the near future because of their roles in generating livelihoods for rural households. Increasing their productivity remains an important component of facilitating poverty reduction in rural areas.

VI. Conclusions

We have documented an inverse relationship between farm size and land productivity in Latin America, one that is persistent across space and time. Theoretically, however, a focus on land productivity is insufficient to assess fully the farm size – productivity relationship, and comprehensive measures such as total factor productivity (TFP) are more appropriate. Throughout the region, agricultural TFP has increased steadily between 1960 and 2010 at an average annual rate of 1-2%. Within the region there has been considerable heterogeneity, with Brazil often identified as one of the fastest performers, and Mexico with positive yet more average performance. Heterogeneity in TFP growth is salient to the farm size – productivity relationship in the region, as modernization of the agricultural sector is expected to render the farm size – land productivity relationship an inadequate proxy for the more important farm size – TFP relationship.

The Brazilian experience highlights this, where we observe a persistent inverse relationship between farm size and land productivity. In contrast, analysis of the farm size – productivity relationship using TFP reveals a relatively flat relationship up to 500ha, with farms above this size being increasingly more productive than the rest of the farm size distribution. A regionally disaggregated analysis in Brazil reveals that the farm size – land productivity relationship can be a reasonable proxy for the farm size – TFP relationship in settings where agriculture is less developed. However, in more technologically advanced locations, land productivity is no longer a suitable measure of performance. Technological change and modernization are forces altering the farm size – productivity relationship. In dynamic regions, such as the Southeast and Center-West of Brazil, it is imperative to use broader measures of productivity such as TFP.

We extend the discussion of Brazil to an exploration of the farm size – productivity relationship in Mexico based on a panel of household farms from the MxFLS. After documenting a persistent inverse relationship between farm size and land productivity over the sample period, 2002 to 2009, we use an average production function to estimate the mean farm size – TFP relationship. This is complemented by the estimation of a stochastic frontier production function, allowing for analysis of the relationship between farm size and frontier productivity and between farm size and technical inefficiency.

We find a time-invariant inverse relationship between farm size and TFP, driven by the relatively high productivity of the smallest farms and relatively low productivity of the largest

farms. Analysis of the production frontier reveals a dynamic inverse relationship between farm size and frontier productivity, where technical change has increased the frontier for larger farms at a faster rate than for smaller farms, weakening the inverse relationship at the frontier. Despite these changes at the frontier, the farm size – average TFP relationship has remained constant due to technical inefficiencies growing faster for larger rather than smaller farms. In essence, many of the larger farms were not able to keep up with technical change at the frontier. This analysis complements the Brazilian experience, and is likely to be applicable to many other countries in Latin America, where the farm size distributions and the dynamism of the agricultural sectors are more similar to Mexico than to Brazil.

To the extent that the inverse relationship between farm size and TFP has flattened in more modernizing regions of Brazil, and along the frontier for household farms in Mexico, suggests that size may fade as one of the key determinants of productivity differences as agricultural sectors modernize. Other factors are likely to grow in importance. In the Mexican case, in addition to farm size, we find that education, participation in government programs such as Procampo, operating as a subsistence farm, the existence of agricultural teaching institutions, and monocropping contribute to explaining differentials in agricultural TFP. More rigorous study of these correlations is necessary to identify causal impacts of these factors. They could potentially provide important opportunities for government policies aimed at increasing agricultural productivity. Similarly, policies that help family farms keep abreast of improvements in agricultural technology could contribute to reducing technical inefficiency over time. The frontier analysis suggests that education was important in this regard in Mexico. In Brazil, both education and extensions services played a significant role in boosting TFP growth.

Robust agricultural TFP growth is also important for poverty reduction. For the agricultural sectors as a whole, it facilitates food supply growing more rapidly than demand so that prices can fall, benefiting poor consumers wherever they may live. And for the small farms that continue to exist, either because they are competitive or because they have few other opportunities, TFP growth helps to boost income. Where farms are too small, as in many parts of Mexico or in the Northeast of Brazil, increased productivity may still be insufficient to lift households out of poverty. Households in regions with access to non-agricultural employment may persist, and some will escape poverty, but migration is likely to continue.

An important caveat to our empirical analysis is the absence of non-family commercial farms in the Mexican sample. Future research should extend our analysis to a nationally representative sample of farms, such as the 2007 Mexican Agricultural Census, which would include family and non-family agricultural operations. Extending the analysis to the entire range of farm sizes and farm types would allow for a more complete analysis of the farm size – productivity relationship. Together with a theoretical analysis of a dynamic farm size – TFP relationship, such extensions would inform policy efforts to increase agricultural productivity.

References

- Assuncao, J. J., & Braido, L. H. (2007). Testing household-specific explanations for the inverse productivity relationship. *American Journal of Agricultural Economics*, 89(4), 980-990.
- Bardhan, P.K. (1973). Size, productivity, and returns to scale: An analysis of farm-level data in Indian agriculture. *The Journal of Political Economy*, 1370-1386.
- Barrett, C. B. (1996). On price risk and the inverse farm size-productivity relationship. *Journal of Development Economics*, 51(2), 193-215.
- Barrett, C. B., Bellemare, M. F., & Hou, J. Y. (2010). Reconsidering conventional explanations of the inverse productivity-size relationship. *World Development*, 38(1), 88-97.
- Battese, G. E. (1997). A note on the estimation of Cobb-Douglas production functions when some explanatory variables have zero values. *Journal of Agricultural Economics*, 48, 250-252.
- Belotti, F., & Ilardi, G. (2012). Consistent Estimation of the “True” Fixed-effects Stochastic Frontier Model (No. 231). Tor Vergata University, CEIS.
- Belotti, F., Daidone, S., Ilardi, G., & Atella, V. (2012). Stochastic frontier analysis using Stata. *Stata Journal*, 13(4), 718-758.
- 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. (1979). *Agrarian structure and productivity in developing countries: a study prepared for the International Labour Office within the framework of the World Employment Programme*. Baltimore: Johns Hopkins Univ. Press.
- Bevis, L., & Barrett, C. B. (2016). Close to the Edge: Do Behavioural Explanations Account for the Inverse Productivity Relationship? Working Paper.
- Binswanger, H. P., Deininger, K., & Feder, G. (1995). Power, distortions, revolt and reform in agricultural land relations. *Handbook of Development Economics*, 3, 2659-2772.
- Bravo Ortega, C., & Lederman, D. (2004). Agricultural productivity and its determinants: revisiting international experiences. *Estudios de economía*, 31(2).
- Carletto, C., Savastano, S., & Zezza, A. (2013). Fact or artifact: the impact of measurement errors on the farm size-productivity relationship. *Journal of Development Economics*, 103, 243-261.
- Coelli, T. J., Rao, D. S. P., O'Donnell, C. J., & Battese, G. E. (2005). *An introduction to efficiency and productivity analysis*, 2nd edition. Springer Science & Business Media.
- Desiere, S., & Jilliffe, D. M. (2017). Land productivity and plot size: Is measurement error driving the inverse relationship? World Bank Policy Research Working Paper No. 8134.
- Eastwood, R., Lipton, M. & Newell, A. (2010). Farm size. *Handbook of agricultural economics*, 4, 3323-3397.
- Eswaran, M. & Kotwal, A. (1986). Access to capital and agrarian production organization. *The Economic Journal*, 482-498.

- 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), 297-313.
- Ferreira Filho, J. B., & Freitas Vian, C. (2015). The Evolving Role of Large and Middle Size Farms in Brazilian Agriculture (No. 212026). International Association of Agricultural Economists.
- Foster, A. D., & Rosenzweig, M. R. (2017). Are There Too Many Farms in the World? Labor-Market Transaction Costs, Machine Capacities and Optimal Farm Size (No. w23909). National Bureau of Economic Research.
- Fuglie, K. O. (2008). Is a slowdown in agricultural productivity growth contributing to the rise in commodity prices?. *Agricultural Economics*, 39(s1), 431-441.
- Gourlay, S., Kilik, T., & Lobell, D. (2017). Could the debate be over? Errors in farmer-reported production and their implications for the inverse scale-productivity relationship in Uganda. World Bank Policy Research Working Paper No. 8192.
- Greene, W. (2005). Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. *Journal of econometrics*, 126(2), 269-303.
- Headey, D., Alauddin, M., & Rao, D. S. (2010). Explaining agricultural productivity growth: an international perspective. *Agricultural Economics*, 41(1), 1-14.
- Helfand, S. M., & Levine, E. S. (2004). Farm size and the determinants of productive efficiency in the Brazilian Center- West. *Agricultural Economics*, 31(2- 3), 241-249.
- Helfand, S. M., Magalhães, M. M., & Rada, N. E. (2015). Brazil's Agricultural Total Factor Productivity Growth by Farm Size. Inter-American Development Bank.
- Helfand, S. M., Rada, N. E., & Magalhães, M. M. (2017). Brazilian Agriculture: Is it all about the Large Farms? *EuroChoices* 16(1): 17-22.
- Helfand, S. M., Rada, N. E., & Magalhães, M. M. (2017). Agricultural productivity growth in Brazil: Large and small farms excel. Mimeo, 47p.
- Helfand, S. M., Moreira, A. R. B., & Figueiredo, A. M. R. (2011). Explaining the Differences in Poverty among Agricultural Producers in Brazil: Counterfactual Simulations with the 1995-96 Agricultural Census. *Revista de Economia e Sociologia Rural*, 49(2), 391-418.
- Helfand S. M., and Taylor M. P., (2016). "The Inverse Relationship between Farm Size and Productivity: Refocusing the Debate." Paper presented at the Western Economic Association International Conference, Portland, OR, June 29-July 3.
- Kagin, J., Taylor, J. E., & Yúnez-Naude, A. (2016). Inverse Productivity or Inverse Efficiency? Evidence from Mexico. *The Journal of Development Studies*, 52(3), 396-411.
- Kumbhakar, S. C., Wang, H. J., & Horncastle, A. P. (2015). *A practitioner's guide to stochastic frontier analysis using Stata*. Cambridge University Press.
- Mundlak, Y. (2001). Production and Supply. *Handbook of Agricultural Economics*, 3-85.
- Prina, S. (2013). Who Benefited More from the North American Free Trade Agreement: Small or Large Farmers? Evidence from Mexico. *Review of Development Economics*, 17(3), 594-608.

- Sadoulet, E., & De Janvry, A. (1995). *Quantitative development policy analysis* (Vol. 1). Baltimore: Johns Hopkins University Press.
- Sen, A.K. (1966). Peasants and Dualism with or without Surplus Labor. *The Journal of Political Economy*, 425-450.
- Trindade, F. J., & Fulginiti, L. E. (2015). Is there a slowdown in agricultural productivity growth in South America?. *Agricultural Economics*, 46(S1), 69-81.
- United Nations Conference on Trade and Development. (2013). *Mexico's Agricultural Development: Perspectives and Outlook*.
- Vollrath, D. (2007). Land distribution and international agricultural productivity. *American Journal of Agricultural Economics*, 89(1), 202-216.
- Wang, H. (2002). Heteroscedasticity and Non-Monotonic Efficiency Effects of a Stochastic Frontier Model. *Journal of Productivity Analysis*, 18, 241–253.

Appendix 1: The Distribution of Households and Land Productivity by Farm Size in Selected Latin American Countries

Table 1: Number of households and land productivity by farm size in Bolivia, 2005

Farm Size	Number of households	Share	Land Productivity (\$/ha)	Land Productivity Index
0-0.5 ha	462	33.33	4,967	1.00
0.5-1 ha	245	17.67	411	0.08
1-2 ha	271	19.57	317	0.06
2-5 ha	296	21.31	243	0.05
5-10 ha	71	5.12	149	0.03
10 ha -	42	3	78	0.02
Total	1,387	100	-	-

*There were also 2,693 households in the survey without land in Bolivia in 2005. These were excluded from the analysis. Gross income is computed by adding gross crop and livestock income and converted into dollars using the exchange rate in 2005 from the World Bank's World Development Indicators (8.07 Bs = 1.0 USD). Farm size is the size of land owned and operated by the household in hectares.

Table 2: Number of households and land productivity by farm size in Guatemala, 2006

Farm Size	Number of households	Share	Land Productivity (\$/ha)	Land Productivity Index
0-0.5 ha	1,926	41.37	2,294	1.00
0.5-1 ha	1,092	23.45	906	0.39
1-2 ha	1,024	22	657	0.29
2-5 ha	558	11.98	523	0.23
5-10 ha	53	1.15	371	0.16
10 ha -	2	0.05	184	0.08
Total	4,655	100	-	-

*There were also 9,027 households in the survey without land in Guatemala in 2006. These were excluded from the analysis. Gross income is computed by adding gross crop and livestock income and converted into dollars using the exchange rate in 2006 from the World Bank's World Development Indicators (7.60 Qz = 1.0 USD). Farm size is the size of land owned and operated by the household in hectares.

Table 3: Number of households and land productivity by farm size in Nicaragua, 2005

Farm Size	Number of households	Share	Land Productivity (\$/ha)	Land Productivity Index
0-0.5 ha	62	3.35	1,453	1.00
0.5-1 ha	161	8.71	1,438	0.99
1-2 ha	207	11.15	458	0.32
2-5 ha	429	23.17	354	0.24
5-10 ha	257	13.89	446	0.31
10 ha -	736	39.73	94	0.06
Total	1,852	100	-	-

*There were also 4,996 households in the survey without land in Nicaragua in 2005. These were excluded from the analysis. Gross income is computed by adding gross crop and livestock income and converted into dollars using the exchange rate in 2005 from the World Bank's World Development Indicators (16.733 NIO = 1.0 USD). Farm size is the size of land owned by the household in hectares.

Table 4: Number of households and land productivity by farm size in Panama, 2003

Farm Size	Number of households	Share	Land Productivity (\$/ha)	Land Productivity Index
0-0.5 ha	576	34.21	4,688	1.00
0.5-1 ha	35	2.08	287	0.06
1-2 ha	202	12	296	0.06
2-5 ha	296	17.6	170	0.04
5-10 ha	185	11.01	120	0.03
10 ha -	389	23.11	53	0.01
Total	1,683	100	-	-

*There were also 4,665 households in the survey without land in Panama in 2003. These were excluded from the analysis. Gross income is computed by adding gross crop and livestock income and converted into dollars using the exchange rate in 2003 from the World Bank's World Development Indicators (1 Balboas = 1.0 USD). Farm size is the size of land owned by the household in hectares.

Table 5: Number of households and land productivity by farm size in Ecuador, 1995

Farm Size	Number of households	Share	Land Productivity (\$/ha)	Land Productivity Index
0-0.5 ha	418	24.6	5,025	1.00
0.5-1 ha	152	8.95	1,199	0.24
1-2 ha	225	13.24	1,382	0.27
2-5 ha	375	22.08	799	0.16
5-10 ha	151	8.86	372	0.07
10 ha -	379	22.27	224	0.04
Total	1700	100	-	-

*There were also 4,021 households in the survey without land in Ecuador in 1995. These were excluded from the analysis. Gross income is computed by adding gross crop and livestock income and converted into dollars using the exchange rate in 1995 from the World Bank's World Development Indicators (2,564.49 Sucres = 1.0 USD). Farm size is the size of land owned by the household in hectares. 11 observation on land productivity in the 0-0.5ha size class were dropped due to the presence of outliers.

Table 6: Number of households and land productivity by farm size in Paraguay, 2000-01

Farm Size	Number of households	Share	Land Productivity (\$/ha)	Land Productivity Index
0-1 ha	189,693	0.43	743	1.00
1-2 ha	19,607	0.04	336	0.45
2-5 ha	42,211	0.10	264	0.36
5-10 ha	69,454	0.16	135	0.18
10-20 ha	69,546	0.16	166	0.22
20-50 ha	31,699	0.07	120	0.16
50-100 ha	7,766	0.02	152	0.20
100 ha -	10,194	0.02	54	0.07
Total	440,170	100	-	-

*There were also 197,837 households in the survey without land in Paraguay in 2000-01. These were excluded from the analysis. Gross income is computed by adding livestock, crop, dairy, and processed products income. The local currency is converted into dollars using the exchange rate in 2000 from the World Bank's World Development Indicators (3486.35 Guarani = 1.0 USD). Farm size is the size of land owned by the household in hectares.

Source: Calculated from data in Toledo (2010), p. 92.

Appendix 2: Construction of the Output Index

There are three components to the construction of the output index. First, the valuation of crop production for each farm. Second, the valuation of livestock production. Third, the construction of the quantity index.

Valuing Crop Production

For each parcel used in production, MxFLS records the name of the three most important crops grown on the parcel, along with the quantities of production, the quantities of any sales, the monies received from any sales, and the units of measurement for the quantities for each of those three crops (the majority of plots produce just one or two crops). These entries are in Spanish, so are first translated and grouped into 90 crop groupings, including an “Other” category containing a wide range of entries that have yet to be clearly translated so are not currently valued.

The preferred approach to generating a useful vector of prices would have been to use all crop observations that recorded both quantity sold and the values received from sales to generate per unit prices. Average prices, for Mexico together and/or for individual states, would be calculated by summing quantities sold and values received for each crop and then calculating the average price per unit (metric tons is the standardized unit chosen). If done in each of the three survey waves, a price vector would be created that could then be applied to all crop observations with recorded quantities of production. The construction of the quantity index requires a price from each of the three periods for each product produced, regardless of whether or not it was produced in all periods. Whereas there was quite good coverage of the crops produced within any given year, many crops did not have prices in all three years. More importantly, marked fluctuations in the crop prices generated across years raised concerns about this approach for generating consistent crop valuations.

An alternative approach to valuing the production of each plot is to abandon the price vectors generated using MxFLS data and to bring in external data on crop prices over time. A trusted and defensible data series of average Mexican crop prices over time covering each of the major crops produced by MxFLS households meets the above requirements that we have complete coverage of the most important crops in each year, and that the prices have been consistently vetted as reasonable. We use price data from the Food and Agriculture Organization of the United Nations

(FAO)³⁷ which provides Mexican producer prices over the relevant time period for approximately 110 crops, resulting in a near one to one mapping to the crop grouping generated during the initial crop translations of the MxFLS parcel data.³⁸

With a complete set of prices in each period, we calculate the value of production of each crop on each parcel for each farm using each of the three periods' prices. We then aggregate across crops for each parcel for each year's prices, and finally across parcels for each farm for each set of prices. The result is three valuations for the crop production of each farm in each year, one using the price from each of the three survey years. This provides the basis for valuing crop production and the construction of our quantity indices.

Valuing Livestock Production

The MxFLS records the existence and value of the stock of many household assets. These asset categories include Horses, Cows and Bulls, Pigs and Goats, and Chickens. Whereas Horses are most likely an input in the agricultural production process, the latter three categories constitute the production of livestock and their related goods. For the 20% - 25% of households owning Cows and/or Bulls in any period, the 23% - 28% owning Pigs and Goats, and the 37% - 46% of households owning Chickens, the MxFLS provides a valuation of those asset stocks and some measures of the product of those asset stocks.

The product of livestock consumed and sold is recorded in two separate sections of the MxFLS survey. For the product sold, the MxFLS asks for the previous year's sales of dairy products, meat products, and fattened animals. On the consumption side, the MxFLS asks for the value of food products "given as a gift, as a payment, or obtained from its crops, animals, or business" over the previous week. The relevant foods for livestock production are "Beef", "Pork", "Cheese", "Other milky products", "Other animal origin products", "Chicken", "Hen Eggs", and "Pasteurized Milk". In estimating the value of livestock production consumed we sum across these categories for each

³⁷ See <http://www.fao.org/faostat/en/#data/PP>.

³⁸ For the "Blackberries" and "Nuts" crop groupings the FAO price data is supplemented using prices generated by the Mexican government, found at <http://www.economia-sniim.gob.mx/2010prueba/Agricolas.asp>. For the "Herbs" crop grouping we generate average prices from the MxFLS data itself. The "Pasture" grouping is not valued, and the "Other" and "Flowers" groupings are not valued either.

household in each year and then multiply by 52. Treating this summation as the annual value of livestock production consumed implicitly assumes that (i) all, or nearly all, of these consumption values come from home production and not gifts or as payment, and (ii) the previous week's consumption patterns are representative of consumption patterns over the course of the year.

A final value of nominal livestock production is then given by summing the value of livestock sales with the value of livestock consumption. Nominal values are then deflated to 2002 values, forming the measure of the value of livestock production to be used in the calculation of our quantity index.

Construction of the Quantity Index

In each period, we begin by aggregating the total value of production for each farm in each survey year. For those households that have complete farm data in two or more years (i.e. complete farm size data on all parcels and valuation data of all crops on all parcels) we then construct a Fisher quantity index. Having identified “complete farms” in the panel, we then generate the following Panel IDs:

- 1 if the farm is in the panel in 2002 and 2005 only
- 2 if the farm is in the panel in 2002, 2005, and 2009
- 3 if the farm is in the panel in 2005 and 2009 only
- 4 if the farm is in the panel in 2002 and 2009 only
- 0 otherwise

We generate the quantity index for farm, f , producing crop, i , by first calculating changes in the Fisher quantity index over all relevant pairs of periods. These changes are then applied to base year values. The changes in the Fisher quantity index are calculated as follows:

- $Q_{02,05}^f = \sqrt{\left(\frac{\sum_i p_{i,2002} q_{fi,2005}}{\sum_i p_{i,2002} q_{fi,2002}}\right) \left(\frac{\sum_i p_{i,2005} q_{fi,2005}}{\sum_i p_{i,2005} q_{fi,2002}}\right)}$... if Panel = 1 or 2
- $Q_{02,09}^f = \sqrt{\left(\frac{\sum_i p_{i,2002} q_{fi,2009}}{\sum_i p_{i,2002} q_{fi,2002}}\right) \left(\frac{\sum_i p_{i,2009} q_{fi,2009}}{\sum_i p_{i,2009} q_{fi,2002}}\right)}$... if Panel =2 or 4

- $Q_{05,09}^f = \sqrt{\left(\frac{\sum_i p_{i,2005} q_{i,2009}}{\sum_i p_{i,2005} q_{i,2005}}\right) \left(\frac{\sum_i p_{i,2009} q_{i,2009}}{\sum_i p_{i,2009} q_{i,2005}}\right)}$... if Panel=2 or 3

With changes in the quantity index in hand we then generate the level of the quantity index for each year as follows. Here $Value_{year}^f$ is the value of output in a given year using nominal prices for farm f :

- $QI_{2002}^f = Value_{2002}^f$... if Panel = 1, 2, or 4
- $QI_{2005}^f = \begin{cases} Value_{2002}^f * Q_{02,05}^f & \text{if Panel = 1 or 2} \\ Value_{2005}^f / Deflator & \text{if Panel = 3} \end{cases}$
- $QI_{2009}^f = \begin{cases} Value_{2002}^f * Q_{02,09}^f & \text{if Panel = 4} \\ Value_{2005}^f * Q_{05,09}^f & \text{if Panel = 3} \\ QI_{2005}^f * Q_{05,09}^f & \text{if Panel = 2} \end{cases}$

Appendix 3: Construction of Inputs

We consider five factors of production in addition to land: physical capital, draft animals, purchased intermediate inputs, non-family labor, and family labor. The portion of the MxFLS questionnaire recording the value of different types of assets asks the value of “tractor, or other machinery or equipment” and “horses, mules, donkeys” owned by each household. These valuations serve as our proxies for physical capital and draft animals, respectively. Each household is similarly asked the value of their expenditure on 9 agricultural inputs over the course of the previous year.³⁹ The summation of these expenditures forms our measure of purchased intermediate inputs. All of the above values are deflated to 2002 values. The construction of the labor indices are outlined below.

Family Labor Index

There are two approaches that we can use to generate a measure of family labor as an input to the production process. The first uses categorical variables from the parcel level questionnaire, which asks a binary yes/no for whether or not the survey respondent, their spouse, their children, their parents, their parents in law, their siblings, their siblings in law, and “other” household members helped farm the plot (each in a separate variable, binary yes or no response) These measures are plot specific, but they do not include any intensive measures of labor, i.e. they only capture labor on the extensive margin (who worked or not) and not on the intensive margin (how much did each person work on the plot that year). The second approach uses time-use data from the individual level. This is a relatively rich dataset including data on type of work for first and second primary jobs, and average hours worked per week on those activities over the prior year. While this approach has advantages on the intensive margin, it is less comprehensive and less complete on the extensive margin.

We develop a set of three indicators. The first uses time-use data and is our preferred approach, whereas the second and third use the binary yes/no data regarding family members’ participation on each plot.

³⁹ These inputs are: chemical fertilizers, organic manure, pesticides, improved seeds, tractor rental, yoke, laborers, water, and fuel/light.

The construction of Family Labor Index 1 is as follows. First, for any core household member⁴⁰ we estimate annual hours worked on the farm. If a core household member indicates that agricultural work on the family farm was either their primary or secondary job then average hours worked per week is the basis for that individual’s annual agricultural labor.⁴¹ If not, then hours spent on household agricultural activities in the previous week provides the basis for the individual’s annual agricultural labor.⁴² The questionnaire asks a similar set of questions about time use for children in the household, which provides the basis for the annual agricultural labor of these household members. For non-core family members the most comprehensive data comes at the plot level which asks which family members worked each particular plot. We calculate a measure of annual hours worked for these family members using group averages of time spent on household agricultural activities and the number of family members in each group, by type of family member.⁴³

Summing hours worked by the core family members and the non-core family members generates our Family Labor Index 4. This approach prioritizes employment data over the time use data, avoids double-counting of those two measures, and uses as much of the data as possible. Equation (1) summarizes this preferred Family Labor Index:

$$Family\ Labor\ 1_i = H_i^h * 52 + H_i^s * 52 + \sum_k H_{ik}^c * 52 + \sum_{j=1}^{10} N_{ij} * \bar{H}_j * 52$$

where H_i^h , H_i^s , and H_{ik}^c are the weekly hours worked of household i ’s household head, household head’s spouse, and household head’s k^{th} children as described above, \bar{H}_j is the average weekly hours worked of non-core family member type j (unique for each of the 10 possible categories)⁴⁴, and N_{ij} is the number of non-core family members in group j of household i .

⁴⁰ Defined as the household head, the household head’s spouse, and the household head’s adult children.

⁴¹ This includes not only those who claim that their job is as a “peasant on your own plot”, but also those who work in agriculture as a “family worker in a household owned business, without remuneration” or a “boss, employer, or business proprietor.” Note that this treatment of the data assumes that work in agriculture as a boss or employer assumes that this is on the family farm and not off of it.

⁴² Individuals are asked about the use of their time on different activities over the previous week, one of which is “make any agricultural activity like weeding hoe[ing], cleaning, sowing, [etc.]”

⁴³ The average number of hours spent engaged in agriculture in the past week is 18.99 hours for non-core family members.

⁴⁴ These family member types are: parents, parents in law, siblings, siblings in law, grandchildren, grandparents, aunts and uncles, nephews and nieces, cousins, and ex-spouses.

The approach for constructing indices two and three has two stages, involving the creation of an indicator for the involvement of family members in household production followed by aggregation across family member types for each household in each year. For each measure we calculate an indicator for each of the $j = 1 \dots 14$ types of family members (identified in relation to the household head).⁴⁵ In recognition that for multi-parcel farms a given family member type may not help on all plots, we weight each family member type's indicator by the share of the farm they participated on (measured as the size of the parcels that they participated on divided by the size of the total farm). For family member type j of farm i , the indicator I_{ij} is given by:

$$I_{ij} = \frac{\text{size of farm } i\text{'s parcels on which group } j \text{ helped}}{\text{total size of farm } i}$$

When aggregating, we then have the option of summing the indicator functions for each family member type or summing with weights that reflect the number of individuals in each family member type in each household in each survey year. The second index uses the former aggregation procedure, with no weights for the number of members of each family member type. The third index uses the latter aggregation procedure, applying weights that reflect the number of individuals in each family member type, N_{ij} :

$$\text{Family Labor } 2_i = \sum_{j=1}^{14} I_{ij}^2$$

$$\text{Family Labor } 3_i = \sum_{j=1}^{14} I_{ij}^3 * N_{ij}$$

Family Labor Indices 1, 2, and 3 are positively correlated with each other. The correlation coefficients between index 2 and 3 range from 0.68 in 2002 to 0.71 in 2005. Family Labor Index 1 is less highly correlated with 2 and 3 than they are with each other, but this is reasonable given that it is based upon time use and is fundamentally different than the other two. Family Labor Index 1 is a measure of annual hours of agricultural labor from family members, capturing the

⁴⁵ These family member types are: spouse, children, step children, children in law, parents, parents in law, siblings, siblings in law, grandchildren, grandparents, aunts and uncles, nephews and nieces, cousins, and ex-spouses.

intensity of agricultural labor of those family members included in the individual Adult and Child surveys, whereas indices 2 and 3 measured the extent of family participation in the agricultural process. Family Labor Index 1 is our preferred measure because it is less crude and it takes advantage of as much of the data as possible, so it is used in our core regression. Family Labor Indices 2 and 3 provide alternative measures and are used for sensitivity analysis.

Non-Family Labor

The MxFLS records the number of non-household members that worked on each parcel used in agricultural production. This forms the basis of our index of non-family labor. For each parcel, we weight this number of individuals by that parcel's share of the farm. These parcel level indicators are then aggregated across parcels for each household in each survey year to form our final measure.

A second measure of non-family labor is recorded in the household's expenditure on agricultural inputs, one of which is expenditure on laborers. These two measures are potentially very different, the former being unpaid non-family labor and the latter being paid labor. This might be especially true for ejido farms. Prior to including both measures we check for correlation between having both expenditures on laborers and non-family laborers helping out on a farm. For those farms with non-family labor, 92%, 54%, and 39% of farms with such workers recorded no expenditure on labor in 2002, 2005, and 2009, respectively. There is a negative correlation coefficient of -0.39 between having non-household members help with agricultural production and having paid for laborers, suggesting that these are distinct measures of labor and are not redundant. There appears to be no substantive difference between the use of these types of labor across ejido and non-ejido farms.

Appendix 4: Regression Results Using Crop Production Only

Table 1: Estimates of the Production Function, Using Crop Production Only

	(1) None	(2) Linear	(3) Quadratic	(4) Cubic	(5) Dummies
Farm Size	-0.808*** (0.068)	-0.764*** (0.098)	-0.749*** (0.095)	-0.707*** (0.148)	
Bin 2: 0.5 to 1 ha					-1.184* (0.672)
Bin 3: 1 to 2 ha					-2.020*** (0.608)
Bin 4: 2 to 5 ha					-1.785*** (0.608)
Bin 5: 5 to 10 ha					-2.612*** (0.777)
Bin 6: 10 to 20 ha					-1.892* (1.095)
Bin 7: 20+ ha					-5.583*** (1.290)
Family Labor		-0.023 (0.076)	-0.021 (0.081)	-0.018 (0.082)	0.068 (0.101)
Physical Capital		-0.097 (0.146)	-0.111 (0.136)	-0.119 (0.135)	-0.108 (0.182)
Draft Animals		0.005 (0.094)	0.002 (0.095)	0.008 (0.099)	0.047 (0.102)
Purchased Intermediates		0.112 (0.081)	0.111 (0.080)	0.111 (0.080)	0.190* (0.104)
Non-family Labor		-0.000 (0.067)	0.000 (0.074)	-0.000 (0.073)	0.084 (0.067)
2005 Dummy	-0.072 (0.195)	-0.025 (0.174)	-0.045 (0.193)	-0.009 (0.228)	0.362 (0.414)
2009 Dummy	-0.317 (0.201)	-0.255 (0.202)	-0.326 (0.221)	-0.305 (0.265)	0.582 (0.560)
Ejido	0.013 (0.255)	-0.057 (0.238)	-0.069 (0.239)	-0.068 (0.248)	-0.206 (0.298)
2005* Farm Size	0.003 (0.076)	0.016 (0.086)	0.005 (0.090)	-0.026 (0.136)	
2009* Farm Size	-0.176 (0.108)	-0.181* (0.109)	-0.191* (0.101)	-0.210 (0.167)	
Farm Size ²			-0.013 (0.013)	-0.005 (0.019)	
2005*Farm Size ²			0.006 (0.013)	-0.002 (0.020)	
2009*Farm Size ²			0.014 (0.019)	0.009 (0.030)	

Farm Size ³				-0.002	
				(0.004)	
2005*Farm Size ³				0.002	
				(0.004)	
2009*Farm Size ³				0.001	
				(0.004)	
2005*Bin 2					-0.784
					(0.737)
2005*Bin 3					0.198
					(0.665)
2005*Bin 4					-0.954
					(0.592)
2005*Bin 5					-0.001
					(0.574)
2005*Bin 6					-0.922
					(0.855)
2005*Bin 7					-0.687
					(1.024)
2009*Bin 2					-1.328
					(0.808)
2009*Bin 3					-0.995
					(0.767)
2009*Bin 4					-1.293
					(0.863)
2009*Bin 5					-1.207
					(0.802)
2009*Bin 6					-1.197
					(1.058)
2009*Bin 7					-2.482
					(1.596)
Constant	9.486***	10.379***	10.575***	10.582***	9.843***
	(0.285)	(2.031)	(2.001)	(2.034)	(2.660)
Household FE	Yes	Yes	Yes	Yes	Yes
R ²	0.84	0.85	0.85	0.85	0.84
N	1235	1235	1235	1235	1235

Standard errors in parentheses

* p<0.10, ** p<0.05, *** p<0.01

Table 2: Robustness Checks for the Production Function, Using Crop Production Only

	(1) Linear, Alt. Index	(2) Dummies, Alt. Index	(3) Linear, No Weights	(4) Dummies, No Weights	(5) Linear, Alt. Output
Farm Size	-0.625*** (0.149)		-0.796*** (0.076)		-0.825*** (0.103)
Bin 2: 0.5 to 1 ha		-0.633 (0.638)		-1.335** (0.574)	
Bin 3: 1 to 2 ha		-0.989 (0.634)		-1.905*** (0.465)	
Bin 4: 2 to 5 ha		-0.721 (0.606)		-1.985*** (0.496)	
Bin 5: 5 to 10 ha		-1.009 (0.734)		-2.736*** (0.597)	
Bin 6: 10 to 20 ha		-0.025 (1.117)		-2.475*** (0.753)	
Bin 7: 20+ ha		-2.466** (1.047)		-5.169*** (1.043)	
Alt. Family Labor	0.144 (0.153)	0.551*** (0.106)			
Family Labor			0.020 (0.063)	0.124 (0.080)	-0.044 (0.076)
Physical Capital	-0.107 (0.148)	-0.155 (0.147)	-0.061 (0.096)	-0.068 (0.124)	-0.136 (0.171)
Draft Animals	0.005 (0.092)	0.011 (0.090)	0.013 (0.071)	0.078 (0.086)	-0.011 (0.100)
Purchased Intermediates	0.104 (0.085)	0.110 (0.086)	0.150*** (0.057)	0.214*** (0.069)	0.092 (0.077)
Non-family Labor	0.002 (0.068)	0.055 (0.069)	-0.000 (0.051)	0.068 (0.059)	0.036 (0.072)
2005 Dummy	-0.041 (0.178)	0.047 (0.377)	-0.090 (0.139)	0.176 (0.379)	-0.335* (0.175)
2009 Dummy	-0.288 (0.207)	0.178 (0.418)	-0.327* (0.174)	0.556 (0.462)	-0.176 (0.210)
Ejido	-0.099 (0.246)	-0.281 (0.239)	-0.106 (0.201)	-0.151 (0.212)	-0.127 (0.254)
2005* Farm Size	0.009 (0.082)		0.042 (0.061)		0.089 (0.090)
2009* Farm Size	-0.175 (0.113)		-0.038 (0.081)		-0.109 (0.117)
2005*Bin 2		-0.256 (0.669)		-0.087 (0.599)	
2005*Bin 3		0.461 (0.642)		0.017 (0.502)	
2005*Bin 4		-0.475		-0.174	

		(0.550)		(0.450)	
2005*Bin 5		0.381		0.600	
		(0.511)		(0.442)	
2005*Bin 6		-0.436		0.028	
		(0.782)		(0.563)	
2005*Bin 7		-0.155		0.012	
		(0.911)		(0.736)	
2009*Bin 2		-0.801		-0.676	
		(0.729)		(0.686)	
2009*Bin 3		-0.652		-0.614	
		(0.728)		(0.612)	
2009*Bin 4		-0.779		-0.374	
		(0.729)		(0.569)	
2009*Bin 5		-0.824		-0.177	
		(0.701)		(0.572)	
2009*Bin 6		-0.802		-0.431	
		(0.967)		(0.636)	
2009*Bin 7		-1.886		-0.628	
		(1.462)		(1.039)	
Constant	10.426***	11.075***	9.492***	9.088***	11.593***
	(1.953)	(2.041)	(1.350)	(1.845)	(2.035)
Sample Weights	Yes	Yes	No	No	Yes
Household FE	Yes	Yes	Yes	Yes	Yes
R ²	0.85	0.85	0.80	0.78	0.86
N	1235	1235	1235	1235	1235

Standard errors in parentheses

* p<0.10, ** p<0.05, *** p<0.01

Table 3: Estimates of the Production Function – Alternative Measure of Output

	(1) Linear w/o Inputs	(2) Linear	(3) Quadratic	(4) Cubic	(5) Dummies
Farm Size	-0.856*** (0.077)	-0.825*** (0.103)	-0.811*** (0.099)	-0.773*** (0.144)	
Bin 2: 0.5 to 1 ha					-1.768*** (0.623)
Bin 3: 1 to 2 ha					-2.361*** (0.568)
Bin 4: 2 to 5 ha					-2.033*** (0.574)
Bin 5: 5 to 10 ha					-2.851*** (0.739)
Bin 6: 10 to 20 ha					-2.216** (1.065)
Bin 7: 20+ ha					-6.182*** (1.354)
Family Labor		-0.044 (0.076)	-0.042 (0.080)	-0.038 (0.082)	0.044 (0.102)
Physical Capital		-0.136 (0.171)	-0.144 (0.152)	-0.142 (0.149)	-0.161 (0.213)
Draft Animals		-0.011 (0.100)	-0.014 (0.101)	-0.007 (0.104)	0.032 (0.102)
Purchased Intermediates		0.092 (0.077)	0.090 (0.075)	0.093 (0.075)	0.170* (0.103)
Non-family Labor		0.036 (0.072)	0.040 (0.079)	0.039 (0.078)	0.122* (0.073)
2005 Dummy	-0.406** (0.200)	-0.335* (0.175)	-0.374** (0.186)	-0.380* (0.216)	-0.266 (0.442)
2009 Dummy	-0.282 (0.226)	-0.176 (0.210)	-0.262 (0.220)	-0.254 (0.264)	0.337 (0.630)
Ejido	-0.054 (0.294)	-0.127 (0.254)	-0.141 (0.255)	-0.147 (0.261)	-0.261 (0.342)
2005* Farm Size	0.074 (0.087)	0.089 (0.090)	0.075 (0.095)	0.087 (0.137)	
2009* Farm Size	-0.104 (0.127)	-0.109 (0.117)	-0.123 (0.111)	-0.131 (0.173)	
Farm Size ²			-0.014 (0.016)	-0.008 (0.020)	
2005*Farm Size ²			0.009 (0.015)	0.009 (0.020)	
2009*Farm Size ²			0.018 (0.020)	0.016 (0.031)	
Farm Size ³				-0.001	

2005*Farm Size ³				(0.004)	
				-0.000	
2009*Farm Size ³				(0.004)	
				0.001	
				(0.005)	
2005*Bin 2					-0.184
					(0.703)
2005*Bin 3					0.653
					(0.659)
2005*Bin 4					-0.540
					(0.602)
2005*Bin 5					0.435
					(0.580)
2005*Bin 6					-0.530
					(0.864)
2005*Bin 7					0.094
					(1.050)
2009*Bin 2					-0.731
					(0.840)
2009*Bin 3					-0.532
					(0.796)
2009*Bin 4					-0.777
					(0.880)
2009*Bin 5					-0.803
					(0.846)
2009*Bin 6					-0.801
					(1.069)
2009*Bin 7					-1.832
					(1.683)
Constant	9.590***	11.593***	11.745***	11.663***	11.310***
	(0.311)	(2.035)	(1.920)	(1.954)	(2.683)
Household FE	Yes	Yes	Yes	Yes	Yes
R ²	0.85	0.86	0.86	0.86	0.85
N	1235	1235	1235	1235	1235

Standard errors in parentheses

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

Appendix 5: Descriptive Statistics for Household and Community Controls

Table 1: Share of Farms with Paved Roads and Institutions in the Community, by Survey Year

Variable	2002	2005	2009
Paved Roads	33%	37%	51%
Public Transport	56%	65%	72%
Permanent Market	18%	18%	18%
Mobile Market	32%	25%	26%
Post Office	41%	29%	22%
Hospital	7%	7%	15%
Elementary School	97%	99%	99%
Secondary School	79%	76%	60%
High School	28%	34%	40%
College	2%	16%	10%
Library	32%	39%	39%

Table 2: Share of Farms with Paved Roads and Institutions in the Community, by Survey Year

Variable	2002		2005		2009	
	Non-Ejido	Ejido	Non-Ejido	Ejido	Non-Ejido	Ejido
Paved	35%	31%	43%	32%	52%	49%
Public Transport	52%	58%	63%	66%	70%	72%
Permanent Market	26%	13%	21%	15%	25%	14%
Mobile Market	35%	30%	27%	24%	32%	21%
Post Office	35%	44%	42%	15%	29%	17%
Hospital	10%	5%	10%	4%	17%	14%
Primary School	97%	97%	99%	99%	97%	100%
Secondary School	81%	77%	74%	79%	54%	63%
High School	40%	21%	39%	28%	43%	38%
College	3%	1%	26%	7%	19%	4%
Library	44%	25%	51%	27%	42%	36%

Table 3: Prevalence of Subsistence Farming, Monocropping, and Procampo Participation, by Farm Size and Survey Year

Farm Size	Subsistence Farming			Monocropping			Participation in Procampo		
	2002	2005	2009	2002	2005	2009	2002	2005	2009
0 to 0.5 ha	70%	53%	73%	76%	89%	75%	39%	38%	25%
0.5 to 1 ha	58%	50%	54%	82%	77%	68%	42%	40%	37%
1 to 2 ha	54%	36%	38%	73%	72%	83%	67%	47%	67%
2 to 5 ha	27%	37%	30%	71%	63%	62%	56%	51%	46%
5 to 10 ha	28%	21%	23%	58%	71%	69%	82%	74%	60%
10 to 20 ha	21%	8%	44%	72%	67%	67%	85%	79%	78%
> 20 ha	24%	25%	18%	72%	67%	82%	44%	42%	68%
Total	43%	36%	41%	71%	73%	71%	45%	42%	39%

Table 4: Share of farms suffering crop and livestock loss

	2002	2005	2009
Crop Loss	9%	7%	15%
Livestock Loss	5%	2%	3%

Table 5: Savings and Credit of Panel Households, by Farm Size

Farm Size	Has Savings			Used Credit		
	2002	2005	2009	2002	2005	2009
0 to 0.5 ha	5%	6%	2%	21%	26%	22%
0.5 to 1 ha	4%	8%	5%	13%	22%	14%
1 to 2 ha	7%	7%	7%	17%	26%	26%
2 to 5 ha	3%	5%	12%	27%	23%	26%
5 to 10 ha	15%	9%	11%	39%	22%	23%
10 to 20 ha	18%	13%	22%	23%	21%	30%
> 20 ha	20%	13%	14%	40%	17%	18%
Total	8%	8%	9%	25%	23%	23%

Table 6: Share of Farms with Access to Credit, by Farm Size

Farm Size	2002		2005		2009	
	Credit	Formal	Credit	Formal	Credit	Formal
0 to 0.5 ha	31%	8%	24%	5%	95%	13%
0.5 to 1 ha	36%	7%	33%	12%	93%	7%
1 to 2 ha	27%	6%	39%	20%	90%	14%
2 to 5 ha	47%	22%	34%	19%	92%	23%
5 to 10 ha	47%	20%	26%	11%	91%	23%
10 to 20 ha	46%	21%	33%	21%	89%	15%
> 20 ha	40%	24%	29%	17%	95%	14%
Total	39%	14%	32%	14%	92%	16%

Table 7: Share of Farms with Indigenous and Literate Household Head, by Farm Size

Farm Size	Indigenous Ethnicity			Literate		
	2002	2005	2009	2002	2005	2009
0 to 0.5 ha	28%	29%	29%	75%	76%	76%
0.5 to 1 ha	38%	32%	28%	71%	77%	72%
1 to 2 ha	38%	39%	36%	77%	76%	78%
2 to 5 ha	27%	34%	26%	77%	78%	78%
5 to 10 ha	20%	12%	11%	90%	93%	83%
10 to 20 ha	13%	8%	19%	79%	92%	89%
> 20 ha	24%	13%	32%	80%	79%	91%
Total	28%	27%	25%	79%	81%	79%

Table 8: Share of last level of education attended, by farm size, 2002

Farm Size	None	Elementary or Less	Secondary School	High School	College
0 to 0.5 ha	24%	60%	8%	7%	1%
0.5 to 1 ha	18%	71%	9%	0%	0%
1 to 2 ha	26%	60%	12%	1%	1%
2 to 5 ha	24%	63%	7%	1%	5%
5 to 10 ha	18%	67%	9%	5%	1%
10 to 20 ha	18%	64%	13%	2%	3%
> 20 ha	28%	52%	12%	4%	4%
Total	23%	62%	10%	3%	2%