**Re-examining Changes in Farm Size Distributions Worldwide Using a Modified Generalized Method of Moments Approach**

A THESIS

SUBMITTED TO THE FACULTY OF THE UNIVERSITY OF MINNESOTA BY

MASON HURLEY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF SCIENCE

Advisor: Dr. Philip Pardey

Committee Members: Dr. Jeffrey Coulter, Dr. Robert King

October 2016

© October 2016, Mason Hurley

**Dedication**

*For my friend and mentor, Jason Beddow.*

**Abstract**

This thesis is focused on measuring and assessing country-level trends in average farm size and farm size distributions over space and time and across income levels worldwide. We develop and implement a new variant of the Generalized Method of Moments approach to estimate farm size distributions derived from world census data. Notwithstanding a major data collection effort, global generalizations are difficult due to incomplete and inconsistent farm size samples over time. However, we did find that average farm sizes have been increasing overall for both high and low to middle income countries. We also detected complex structural differences between countries stratified by income class. These differences were revealed by estimating farm size distributions (and a range of associated summary statistics) that would otherwise be masked when considering the global dynamics of farm size based only on changes in the average size of farm per country.

**Keywords:** farm size distributions, time trends, generalized method of moments

# Table of Contents

# List of Tables

## List of Figures

## 1. Introduction

Poor countries generally experience significantly lower productivity levels in agriculture than their rich counterparts and a much larger proportion of their work force is involved in farming, often for (semi-) subsistence purposes. Over the past half century, average farm size in high-income countries has been increasing with the aid of new technology, and agricultural labor productivity has also risen. At the same time, farms in poor countries have seen little change or, in some places, a decline in average farm size while adopting few, if any, improved inputs. As a result, labor productivity growth in poor countries has stagnated, contributing to the widening of the productivity gap between rich and poor countries. A key component of agricultural labor productivity is physical farm size. Larger farms have significantly higher labor productivity, which seems largely attributable to size dependent technology (e.g., tractors, cultivators, planters, and other large scale machinery) (Adamopoulos et al. 2014).

Economists (and others) have had a long standing interest in farm size, and there is a large literature on this topic (see, for example, Carter 1984; van Zyl et al. 1995; Heltberg 1998; Eastwood et al. 2010). Sumner (2014) summarizes evidence on the agricultural and farm size structure of the United States and how it has changed over the years. He focuses mostly on commercial farming and considers farm size in terms of value of production (or, more specifically, sales) instead of area. He reports that the number of farms in the United States has barely changed over the past three decades. Moreover, the United States has transitioned to a point where roughly 6 percent of farms produce 75 percent of U.S. farm output in terms of value of production, meaning that drastic changes to the make-up of farm size on the smaller end of the spectrum (by value) would have little impact on the overall output of the sector as a whole. Commercial farms have been growing in size very quickly while small farms have also been growing, albeit slowly, so the lion's share of sales belongs to fewer and fewer farms over time. Sumner points out that large farms are small businesses relative to firms in other sectors of the U.S. economy, and notes that competition is still high in most sectors so there are no discernable market power issues as a result of the farm size consolidation regarding the value share of output. He concludes by arguing that manager capability is one of the more important determinants of farm size. Skilled managers are able to handle much larger farms and will be lured from small farms to larger, higher paying farms. Meanwhile, small (often family run) farms are limited in growth by the head of the household's skill level and ability to adapt to new technologies that improve the productivity of land and labor, and enable an increase in farm size in terms of area and/or value.

In a short discussion on the techniques used to estimate farm size, Hall (1987) raises several concerns; the most notable being that there is a selection bias when looking at farm size changes due to farms entering and exiting over time, the result of this bias depending on the reasons for entry and exit. However, she argues that correcting for this bias would likely have little if any meaningful effect on the estimated trends in average farm size. She points out that this bias is present in other industries and that farm size distributions are very similar in nature to the size distribution of firms in other sectors of the economy.

Lowder et al. (2016a) is the latest in a line of studies seeking to provide an overview of the global dynamics of farm size. They use a similar, but less comprehensive, dataset to that used in this paper to examine differences in average farm size trends across income levels and geopolitical regions. They report that in most low and low-middle income countries, average farm size has decreased. At the same time, among upper-middle income countries, average farm size has increased in more than half of these countries while decreasing overall. High-income countries have seen an increase in average farm size from just over 20 ha per farm to roughly 30 ha per farm during the period of 1960 to 2000, with the 30 highest income countries (81.1% of the high-income countries sampled, 29.1% of the entire sample) all reporting increasing average farm size over this period. Grouping countries by geopolitical region – specifically, East Asia and the Pacific, Latin America and the Caribbean, Middle East and North Africa, South Asia and Sub-Saharan Africa – the preponderance of countries in each of these regions have shrinking average farm sizes, although the authors note that their sample size is unlikely to be sufficiently large to be representative of overall trends in each region.

Adamopoulos and Restuccia (2014) examined the relationship between farm size and productivity. Using a snapshot of 1990 data for 63 countries, they attempted to show that farm size is an important determinant of the low productivity problem in agriculture of poor countries. They point to some of the same differences in farm size at different income levels that were shown in Lowder et al. (2016a), namely that rich countries tend, on average, to have much larger farms than poor countries. They reveal that the positive link between farm size and the level of development at the country level is clear and independent of other seemingly important factors (e.g., land endowment, geographical location, land quality, and type of agriculture). For their empirical work, they use *average* farm size in terms of area and show that farm size is highly correlated with overall country development. They also examine policies in low-income countries that establish explicit and implicit limits to farm size and find that these policies may have hindered productivity growth in the agricultural sector in some cases.

Getting the data right on measures of average farm size or, arguably of more importance, attributes of the distribution of farm size is important. Many studies have taken the available data at face value to analyze one of several dimensions of farm size and its relationship to agricultural productivity (Barrett et al. 2010), input use (Mottaleb and Mohanty 2015; Ju et al. 2016) – including the links between farm size and natural inputs like climate, soil and so forth (Barrett et al. 2010) –, and farm livelihoods and the prevalence of (rural) poverty (Fan and Chan-Kang 2005). This paper is focused on reassessing the nature and robustness of the underlying farm size data and develops and implements new methods to standardize farm size distribution estimates and the associated summary statistics (mean, median, percentiles) across time and space.

The primary objective of this paper is to reexamine how farm size has changed over time around the world. To accomplish this objective, we used farm size distribution data from national census data collected and compiled by the Food and Agricultural Organization of the United Nations (FAO), as well as distribution data for some additional countries that were obtained by the International Science and Technology Practice and Policy Center (InSTePP) from a variety of sources (see Appendix 1, Table 1). We apply a variation of the Generalized Method of Moments approach to fit gamma and beta distributions to the data and determine which distribution is best when it comes to estimating country-level farm size distributions. We then use the estimated distributions to obtain estimates for the 5th, 25th, 50th, 75th and 95th percentiles of farm area for each country and examine how those measures changed over time in high and low-middle income countries. We also estimate the Gini coefficient for each distribution to examine the extent of inequality in farm size distributions.

Presaging our results, we find that in the majority of cases the gamma distribution is the best choice for estimating farm size distributions, but introducing a constraint that requires the parameters to estimate a country-level distribution with an average farm size set equal to the country's actual observed average farm size is preferred. We find that high-income countries experienced farm size patterns that are distinctly different from the distribution of low- and middle-income countries for all size cohorts. We also find that farm size is distributed more unequally in high-income countries and that the extent of that inequality is growing. In contrast, we find that farm size inequality is declining in low-income countries.

## 2. Data Description

We use data collected by the FAO (1981, 1997 and 2013). The data were taken from national censuses and represent the overall distribution of farm size for each country measured in area per farm terms. Figure 1 illustrates the diversity in farm size distributions using data for Jamaica and the United States for the year 2000. The distribution statistics are also included and help show that the majority of Jamaican farms are at the lowest end of the spectrum with very few farms over 20 ha. In contrast, the United States distribution is skewed less heavily to the right, which is common in countries with large farms. To harmonize the distribution data for over time and across country consistency, we used an approach decidedly different than the FAO's. The FAO standardized their data by first defining a set of consistent farm size cohorts (see Table 1) and then reallocating reported national size distributions into their consistent cohorts (FAO 1981; FAO 1997; FAO 2013). [1]

[Figure 1: United States versus Jamaica Distribution of Farms, 2000]

[Table 1: Example of Disparity in Reported Farm Size Cohorts]

The FAO draws together farm size (and other) data in a series of world census reports which they publish every decade. The reports we draw on for this study pertain to data published in FAO's 1970, 1990 and 2000 reports.[2][3] We use World Bank (2016) income level and region descriptors for each country from 1970, which we take to be the base year for comparisons of interest. The FAO distributes guidelines to national statistical departments and sources its data from each of the available national agricultural censuses. However, for a particular world census report the available country-level data for the year that falls closest to the census year. For

[1] Cohorts for all three censuses are (in hectare): 0-1, 1-2, 2-5, 5-10, 10-20, 20-50, 50-100, 100-200, 200-500, 500-1000, 1000 and over

[2] The *World Programme for the Census of Agriculture 2020* report states that: "It is the tenth round in the decennial programme of agricultural censuses, which started in 1930. The 1930 and 1940 rounds were sponsored by the International Institute of Agriculture (IIA). The six subsequent rounds – in 1950, 1960, 1970, 1980, 1990, 2000 and 2010 – were promoted by the Food and Agriculture Organization of the United Nations (FAO), which assumed the responsibilities of IIA following its dissolution in 1946 (FAO 2015, p. 3)." We rely on (among other sources) the FAO reports from 1970, 1990 and 2000.

[3] We exclude the 1980 world census because of its small sample size. The *1980 World Census of Agriculture Methodological Review* states that: "It is known that 103 countries carried out a census of agriculture during 1976-1985, but FAO had received the census results from only 86 countries as of April 1990 (FAO 1992, p. 12)."

example, for the 2000 world census, Albania's data are actually from 1998 while Botswana's data are from the year 2004. The distributions include the number of farms in a farm-size cohort (e.g., between 0 and less than 1 hectare, and so on) and in most cases they include the area in each cohort as well. The distributions are discrete and bounded on the low end at zero while unbounded at the high end (i.e., the largest size cohort will refer to farms that are **X** hectares or larger). Notably, the reported size range for each farm size cohort is not standardized across countries, and are often loosely dependent on the characteristics of farming in the country. Table 1 illustrates the considerable inconsistency in the reported farm size cohorts for a number of countries. For example, the Republic of Korea with more than two million farms in total, reports only three cohorts, with the lower bound of the largest farm-size cohort being 2 hectares per farm. In contrast, Guatemala with less than one million farms, groups their farms into 11 farm-size cohorts with the lower bound of their largest farm-size cohort being 9,032 hectares per farm.

Rather than drawing our primary data from the re-aggregated, standardized farm size cohorts reported by FAO (FAO 1981; FAO 1997; FAO 2013) we opted for an alternative, replicable, approach to developing standardized farm size distributions. To do this, we compiled our primary data from the underlying country-specific tabulations in each of these FAO census reports. The advantage of this approach is that we did not introduce an undocumented (and thus non-replicable) re-aggregation of the reported farm size cohorts into our measurement schema. To generate inter-temporal and cross-country consistency in the farm size *distributions* we then used a Generalized Method of Moments approach to obtain fitted distributions, which we then compared across time and among countries looking at the estimated means, medians, percentiles and skewness. We also estimated the Gini coefficient for each country-year observation to assess changes in the inequality of farm size.

While this approach is a transparent way of resolving measured inconsistencies in the distribution data, some unavoidable inconsistencies remain in the definition of a farm among countries and over time. The World Census of Agriculture lays down guidelines as to what constitutes a farm but many countries use alternative definitions, some of which may have substantive empirical consequences (for example, assigning an area or total value of sales threshold to what constitutes a census farm, assigning a minimum number of animals threshold, only including farms producing certain crops, excluding farms located in urban areas, and others).[4] Ideally, every country would include all agricultural holdings regardless of type, size, location and other factors. Moreover, FAO's definition of what constitutes a farm changes over time. In 1970, they proposed that all farming activity be included, but in 1990 they excluded urban areas. In 2000, they returned to a definition similar to 1970. Finally, within many countries the definition of a farm changes from year to year as well. For example, in 1990 Réunion defined a holding to be a farm if its physical footprint was greater than one half of a hectare and in 2000 they defined a holding to be a farm if its footprint was greater than one hectare. This could result in the data reflecting *measured* changes in the distribution that are attributable to definitional changes rather than underlying changes in the structure of farm size. These definitional changes are typically small and likely do not have a major impact on the estimated distributions, but it is important to note when comparing farm size trends in a country over time.

Using the distribution data, we formed estimates of average farm sizes for 80 1970 census countries and 71 2000 census countries.[5] We have data for both census years for only 30 of these countries. To expand the sample size we draw in the reported averages in Lowder et al. (2016b) such that a pooled sample of 81 countries in 1970 and 77 countries in 2000, of which 48 countries report averages for both census years.

[Table 2: Summary of InSTePP Distributions and Lowder et al. Means]

**3. Methods**

The following is a heuristic summary of the methods developed in this paper. A more rigorous and complete description of the derivations we developed are reported in Appendix 2. The important difference between the methods we developed to estimate farm size distributions and the prior efforts reported by Chotikapanich et al. (2007) and Gholamreza et al. (2012) to estimate the distribution of per capita income is that we introduced the option of a constrained estimation technique as an extension to the unconstrained method they reported. In our case we estimated both unconstrained farm size distributions using the Chotikapanich et al. (2007) and Gholamreza et al. (2012) approach, as well as constrained farm size distributions where the estimated average farm size for a country-year observation was constrained to equal the reported

[4] For example, in 1990, Paraguay defined census holdings based on a set of conditions including: "have at least .1 ha of land under crops (FAO 1997, p. 152)." and "have at least one head of adult cattle (FAO 1997, p. 152)."
[5] We also found 62 averages for 1990 census countries but these data have been set aside for further consideration in this study due to sample size and sample inconsistency issues.

(or directly implied) average farm size.[6] This data required for this approach is number of farms per each size cohort *and* the area for which each size cohort is responsible.

### 3.1 Unconstrained Generalized Method of Moments Approach

We begin by defining a set of theoretical distributional moments designed to utilize as much of the observed data as possible. Traditionally, GMM focuses on matching the estimated mean and variance to the observed mean and variance. We instead create moments for both the proportion of farms in each cohort and the average size of a farm in each cohort. This means that for every cohort there are two moment conditions in the system, one that minimizes the difference between the actual proportion of farms in the cohort and the estimated proportion of farms in the cohort and another that minimizes the difference between the actual average size of a farm in the cohort and the (conditional) average size of a farm in the cohort implied by the distribution. This results in over-identification, where we have a set of moment equations larger in size than the parameter vector used to solve the system (that vector ranges in size between one and three parameters depending on the statistical distribution used and the chosen GMM approach). For this reason, there is no unique solution to the system of equations. Instead, we minimize the weighted square error between the theoretical and sample moments.

Assume the cumulative distribution of farm size is $F(\cdot|\boldsymbol{\beta})$ with density $f(\cdot|\boldsymbol{\beta})$ where $\boldsymbol{\beta}$ is an unknown vector of parameters. To understand how the moment conditions for GMM can be constructed to estimate $\boldsymbol{\beta}$ given cohort information on the number of farms and total farm area in a single farm size cohort (e.g., $I_s$ and $A_s$ for cohorts $s$ =1, 2,…, $S$), define $a_i$ as the area of a single farm from a sample of $I$ farms ($i$ = 1, 2, …, $I$) where $I = \sum_{s=1}^{S} I_s$. Let $\varphi_s \geq 0$ for $s$ = 0, 1, …, $S$ be the size cohort bounds such that any given farm area $a_i$ will fall into some farm size cohort $s$ (i.e., $\varphi_{s-1} < a_i \leq \varphi_s$). These bounds can then be used to define an indicator function

(1) $$g^s(a_i) = \begin{cases} 1, & \text{for } \varphi_s \geq a_i > \varphi_{s-1} \\ 0, & \text{otherwise} \end{cases}$$

for each cohort $s$. The expectation of this indicator function is

(2) $$E(g^s(a_i)) = E(1|\varphi_s \geq a_i > \varphi_{s-1}) = F(\varphi_s|\boldsymbol{\beta}) - F(\varphi_{s-1}|\boldsymbol{\beta}).$$

Similarly, it is useful to recognize

(3) $$E(a_i g^s(a_i)) = E(a_i|\varphi_s \geq a_i > \varphi_{s-1}) = (F(\varphi_s|\boldsymbol{\beta}) - F(\varphi_{s-1}|\boldsymbol{\beta}))\mu^s(\boldsymbol{\beta})$$

[6] The imputed farm size average was obtained by dividing the total reported area in farms by the total reported number of farms in each country-year.

where $\mu^s(\boldsymbol{\beta})$ is the conditional mean when $\varphi_{s-1} < a_i \leq \varphi_s$. Rearranging these two expressions to equal zero, summing over all $i$, and dividing by $I$ then gives

(4) $$\frac{\sum_{i=1}^{I}\left(E\left(g^s(a_i)\right)-\left(F(\varphi_s|\boldsymbol{\beta})-F(\varphi_{s-1}|\boldsymbol{\beta})\right)\right)}{I} = E\left(\frac{I_s}{I} - \left(F(\varphi_s|\boldsymbol{\beta}) - F(\varphi_{s-1}|\boldsymbol{\beta})\right)\right) = 0 \text{ and}$$

(5) $$\frac{\sum_{i=1}^{I}\left(E\left(a_i g^s(a_i)\right)-\left(F(\varphi_s|\boldsymbol{\beta})-F(\varphi_{s-1}|\boldsymbol{\beta})\right)\mu^s(\boldsymbol{\beta})\right)}{I} = E\left(\frac{I_s}{I}\frac{A_s}{I_s} - \left(F(\varphi_s|\boldsymbol{\beta}) - F(\varphi_{s-1}|\boldsymbol{\beta})\right)\mu^s(\boldsymbol{\beta})\right) = 0$$

for $s$ = 1, 2, …, $S$.

Equations (4) and (5) define 2$S$ moment conditions that use all of our observable information and depend only on the unknown parameters of the assumed theoretical distribution. These equations also show us how these moment conditions can be written in terms of some sample of individual farm areas, which allows us to cast them in a more typical GMM context:

(6) $$\mathbf{H}(\boldsymbol{\beta}) = \frac{\sum_{i=1}^{I}\mathbf{h}(a_i|\boldsymbol{\beta})}{I} \text{ where } \mathbf{h}(a_i|\boldsymbol{\beta}) = \begin{bmatrix} g^1(a_i) - \left(F(\varphi_1|\boldsymbol{\beta}) - F(0|\boldsymbol{\beta})\right) \\ \vdots \\ g^{S-1}(a_i) - \left(F(\varphi_{S-1}|\boldsymbol{\beta}) - F(\varphi_{S-2}|\boldsymbol{\beta})\right) \\ a_i g^1(a_i) - \left(F(\varphi_1|\boldsymbol{\beta}) - F(0|\boldsymbol{\beta})\right)\mu^1(\boldsymbol{\beta}) \\ \vdots \\ a_i g^S(a_i) - \left(F(\varphi_S|\boldsymbol{\beta}) - F(\varphi_{S-1}|\boldsymbol{\beta})\right)\mu^S(\boldsymbol{\beta}) \end{bmatrix}.$$

Note that not all of these moment conditions are independent because $\sum_{i=1}^{I}\frac{I_s}{I} = 1$, so we drop one of the redundant moments (i.e., the conditions for equation (4) where $s = S$) yielding a 2$S$ - 1× 1 matrix. Within this context, the GMM estimator can be obtained by solving

(7) $$\min_{\boldsymbol{\beta}} \mathbf{H}(\boldsymbol{\beta})'\mathbf{W}\,\mathbf{H}(\boldsymbol{\beta})$$

where $\mathbf{W}$ is some 2$S$ - 1×2$S$ – 1 positive definite weighting matrix. It is assumed that this optimal weighting matrix exists and is positive definite. By minimizing equation (7), we obtain a consistent and optimal estimator for $\boldsymbol{\beta}$ (Cameron and Trivedi 2005).

The optimal weighting matrix $\mathbf{W}(\boldsymbol{\beta})$ for efficient GMM estimation can be obtained from

(8) $$\mathbf{W}(\boldsymbol{\beta}) = \left(plim\sum_{i=1}^{I}\frac{h(a_i|\boldsymbol{\beta})h(a_i|\boldsymbol{\beta})'}{I}\right)^{-1} = \left(E[h(a|\boldsymbol{\beta})h(a|\boldsymbol{\beta})']\right)^{-1},$$

which is a 2$S$ – 1 × 2$S$ – 1 matrix with six distinct types of terms:

(9) $$E\left[\left(g^s(a) - \left(F(\varphi_s|\boldsymbol{\beta}) - F(\varphi_{s-1}|\boldsymbol{\beta})\right)\right)^2\right] =$$
$$\left(F(\varphi_s|\boldsymbol{\beta}) - F(\varphi_{s-1}|\boldsymbol{\beta})\right)\left(1 - \left(F(\varphi_s|\boldsymbol{\beta}) - F(\varphi_{s-1}|\boldsymbol{\beta})\right)\right),$$

(10) $$E\left[\left(g^s(a) - \left(F(\varphi_s|\boldsymbol{\beta}) - F(\varphi_{s-1}|\boldsymbol{\beta})\right)\right)\left(g^t(a) - \left(F(\varphi_t|\boldsymbol{\beta}) - F(\varphi_{t-1}|\boldsymbol{\beta})\right)\right)\right] =$$
$$-\left(F(\varphi_s|\boldsymbol{\beta}) - F(\varphi_{s-1}|\boldsymbol{\beta})\right)\left(F(\varphi_t|\boldsymbol{\beta}) - F(\varphi_{t-1}|\boldsymbol{\beta})\right),$$

(11) $E\left[\left(g^s(a) - (F(\varphi_s|\boldsymbol{\beta}) - F(\varphi_{s-1}|\boldsymbol{\beta})\right)\left(g^s(a)a - \left(F(\varphi_s|\boldsymbol{\beta}) - F(\varphi_{s-1}|\boldsymbol{\beta})\right)\mu^s(\boldsymbol{\beta})\right)\right] =$

$(F(\varphi_s|\boldsymbol{\beta}) - F(\varphi_{s-1}|\boldsymbol{\beta}))\left(1 - (F(\varphi_s|\boldsymbol{\beta}) - F(\varphi_{s-1}|\boldsymbol{\beta}))\right)\mu^s(\boldsymbol{\beta})$,

(12) $E\left[\left(g^s(a) - (F(\varphi_s|\boldsymbol{\beta}) - F(\varphi_{s-1}|\boldsymbol{\beta})\right)\left(g^t(a)a - \left(F(\varphi_t|\boldsymbol{\beta}) - F(\varphi_{t-1}|\boldsymbol{\beta})\right)\mu^t(\boldsymbol{\beta})\right)\right] =$

$-\left(F(\varphi_s|\boldsymbol{\beta}) - F(\varphi_{s-1}|\boldsymbol{\beta})\right)\left(F(\varphi_t|\boldsymbol{\beta}) - F(\varphi_{t-1}|\boldsymbol{\beta})\right)\mu^t(\boldsymbol{\beta})$,

(13) $E\left[\left(g^s(a)a - \left(F(\varphi_s|\boldsymbol{\beta}) - F(\varphi_{s-1}|\boldsymbol{\beta})\right)\mu^s(\boldsymbol{\beta})\right)^2\right] =$

$\left(F(\varphi_s|\boldsymbol{\beta}) - F(\varphi_{s-1}|\boldsymbol{\beta})\right)\left({\sigma^s}^2(\boldsymbol{\beta}) - \mu^s(\boldsymbol{\beta})^2 - \left(F(\varphi_s|\boldsymbol{\beta}) - F(\varphi_{s-1}|\boldsymbol{\beta})\right)\mu^s(\boldsymbol{\beta})^2\right)$, and

(14) $E\left[\left(g^s(a)a - \left(F(\varphi_s|\boldsymbol{\beta}) - F(\varphi_{s-1}|\boldsymbol{\beta})\right)\mu^s(\boldsymbol{\beta})\right)\right.$

$\left.\left(g^t(a)a - \left(F(\varphi_t|\boldsymbol{\beta}) - F(\varphi_{t-1}|\boldsymbol{\beta})\right)\mu^t(\boldsymbol{\beta})\right)\right] =$

$-\left(F(\varphi_s|\boldsymbol{\beta}) - F(\varphi_{s-1}|\boldsymbol{\beta})\right)\left(F(\varphi_t|\boldsymbol{\beta}) - F(\varphi_{t-1}|\boldsymbol{\beta})\right)\mu^s(\boldsymbol{\beta})\mu^t(\boldsymbol{\beta})$

We can then write $\mathbf{W}(\boldsymbol{\beta}) = \begin{bmatrix} \mathbf{W}^U(\boldsymbol{\beta}) & \mathbf{W}^{UL}(\boldsymbol{\beta})' \\ \mathbf{W}^{UL}(\boldsymbol{\beta}) & \mathbf{W}^L(\boldsymbol{\beta}) \end{bmatrix}$ where $\mathbf{W}^U(\boldsymbol{\beta})$ is an $S - 1 \times S - 1$ matrix with terms from equation (9) on the diagonal and terms from equation (10) off the diagonal; $\mathbf{W}^{UL}(\boldsymbol{\beta})$ is an $S \times S - 1$ matrix with terms from equation (11) on the diagonal and terms from equation (12) off the diagonal; and $\mathbf{W}^L(\boldsymbol{\beta})$ is an $S \times S$ matrix with terms from equation (13) on the diagonal and terms from equation (14) off the diagonal.

*Implementation Issues*

Operationalizing equation (7) requires some choice of $F(\cdot|\boldsymbol{\beta})$. Exploratory plots of our country level distributions revealed that all were skewed right, often to an extreme. Additionally, the upper area was often unbounded, meaning that there was no given maximum farm size (e.g., ***X** hectare and larger).* The farm size distributions are also naturally bound by zero on the low end. These facts made the gamma distribution an obvious candidate:

(15) $f(a|k,\theta) = \frac{1}{\Gamma(k)\theta^k} a^{k-1} e^{-\frac{a}{\theta}}$ for $a > 0$

with $\Gamma(t) = \int_0^\infty x^{t-1}e^{-x}\,dx$, mean $\mu = k\theta$ and variance $\sigma^2 = k\theta^2$. Alternatively, the conditional mean and variance can be written as

(16) $\mu(x^l, x^u) = \int_{x^l}^{x^u} \frac{\frac{x}{\Gamma(k)\theta^k} x^{k-1} e^{-\frac{x}{\theta}}}{F(x^u|k,\theta) - F(x^l|k,\theta)}\,dx = \left(\frac{F(x^u|k+1,\theta) - F(x^l|k+1,\theta)}{F(x^u|k,\theta) - F(x^l|k,\theta)}\right) k\theta$

and

$$(17) \quad \sigma^2(x^l, x^u) = \int_{x^l}^{x^u} \frac{\frac{x^2}{\Gamma(k)\theta^k} x^{k-1} e^{-\frac{x}{\theta}}}{F(x^u|k,\theta) - F(x^l|k,\theta)} dx = \frac{F(x^u|k+2,\theta) - F(x^l|k+2,\theta)}{F(x^u|k,\theta) - F(x^l|k,\theta)} k(k+1)\theta^2.$$

While the largest farm size cohort is typically unbounded in the observed data, there is some upper limit to the size of farms in any given country. This constitutes an inconsistency with the theoretical range of the gamma distribution (which does not have an upper limit) that may be of practical significance. Therefore, we also considered the beta distribution due to its flexibility and theoretical range:

$$(18) \quad f(a|\alpha, \beta, \bar{a}) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \left(\frac{a}{\bar{a}}\right)^{\alpha-1} \left(1 - \frac{a}{\bar{a}}\right)^{\beta-1} \text{ for } \bar{a} > a > 0$$

with mean $\mu = \frac{\alpha}{\alpha+\beta} \bar{a}$ and variance $\sigma^2 = \frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2} \bar{a}^2$. The conditional mean and variance are then

$$(19) \quad \mu(x^l, x^u) = \int_{x^l}^{x^u} x \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} dx = \frac{\alpha}{\alpha+\beta} \left( F(x^u|\alpha+1, \beta) - F(x^l|\alpha+1, \beta) \right)$$

and

$$(20) \quad \sigma^2(x^l, x^u) = \int_{x^l}^{x^u} x^2 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} dx = \frac{(\alpha+1)\alpha}{(\alpha+1+\beta)(\alpha+\beta)} \left( F(x^u|\alpha+2, \beta) - F(x^l|\alpha+2, \beta) \right)$$

With these two alternative distributions, we used Matlab's unconstrained minimization function (fminunc) to solve equation (7) numerically. Matlab's fminunc function requires starting values. If the chosen starting values are poor, the objective function in equation (7) may not be defined numerically. That is, fminunc may stop before finding an estimate, or fminunc may find an answer that is a local rather than global minimum. To address these potential problems, we increased the number of iterations to 4,000 the maximum function evaluations per iteration to 2,000. We used the observed information to choose better starting values by incorporating the observed average farm size and an estimated variance. To assess the robustness of our results, we ran fminunc repeatedly for a wide range of starting values determined by our mean and variance estimates and a set of arbitrary scalars picking the results with the lowest value for the objective function.

The observed data made it possible to directly calculate the average farm size: $\hat{\mu} = \frac{\sum_{s=1}^{S} A_s}{I}$. We could also obtain an estimate of the variance: $\hat{\sigma}^2 = \frac{\sum_{s=1}^{S} I_s (A_s - \hat{\mu})^2}{I}$. To pick better starting values $k^0$ and $\theta^0$ for the gamma distribution, we used the overall average farm size from the observed data and set $\theta^0 = \frac{\hat{\mu}}{k^0}$. We then pick a value of $k^0 = c$ and solved for $\theta^0$. Note that

$\hat{\mu} = k^0\theta^0$ for all values of $c$, so a set of scalar values were used for $c$ and the minimization process was repeated for each $c$. Again, the final reported results correspond to the starting values that yielded the lowest value for equation (7).

To pick better starting values $\alpha^0$, $\beta^0$ and $\bar{a}^0$ for the beta distribution, we used the observed mean as well as the approximated variance. Substituting $\hat{\mu}$ for $\mu$ into the mean equation for beta and rearranging gives an estimate for $\bar{a}^0$

(21) $$\bar{a}^0 = \hat{\mu}\,\frac{\alpha^0+\beta^0}{\alpha^0}.$$

Next, substituting into the variance equation,

(22) $$\hat{\sigma}^2 = \frac{\alpha^0\beta^0\left(\hat{\mu}\frac{\alpha^0+\beta^0}{\alpha}\right)^2}{(\alpha^0+\beta^0+1)(\alpha^0+\beta^0)^2} = \frac{\beta^0\hat{\mu}^2}{\alpha^0(\alpha^0+\beta^0+1)}$$

and solving for $\beta^0$ gives a starting value for one parameter as a function of the other, the observed mean and the estimated variance:

(23) $$\beta^0 = \frac{\frac{\hat{\sigma}^2}{\hat{\mu}^2}\left({\alpha^0}^2+\alpha^0\right)}{\left(1-\alpha^0\frac{\hat{\sigma}^2}{\hat{\mu}^2}\right)}$$

The starting value for $\alpha$ was then assigned in the same manner as $k^0$ for the gamma distribution. Once the value for $\alpha^0$ was assigned, $\beta^0$ and $\bar{a}^0$ are calculated using equations (23) and (21).

**3.2 Constrained Generalized Method of Moments Approach**

An initial inspection of our GMM results revealed that the mean implied by our parameter estimates often deviated substantially from the observed mean. To the extent that our data were derived from a sample of farms, some deviation is expected. However, for a sample that is meant to represent census data, estimated means that are rarely within 10 percent of the observed mean are disconcerting. Therefore, we also estimated equation (7) while imposing the constraint that the estimated mean must equal the observed mean. An adjustment must be made to the objective function when introducing this constraint. Specifically, equation 6 becomes:

(24) $$\mathbf{H}(\boldsymbol{\beta}) = \frac{\sum_{i=1}^{I}\mathbf{h}(a_i|\boldsymbol{\beta})}{I} \text{ with } \mathbf{h}(a_i|\boldsymbol{\beta}) = \begin{bmatrix} g^1(a_i) - (F(\varphi_1|\boldsymbol{\beta}) - F(0|\boldsymbol{\beta})) \\ \vdots \\ g^{S-1}(a_i) - (F(\varphi_{S-1}|\boldsymbol{\beta}) - F(\varphi_{S-2}|\boldsymbol{\beta}) \\ a_i g^1(a_i) - \big(F(\varphi_1|\boldsymbol{\beta}) - F(0|\boldsymbol{\beta})\big)\mu^1(\boldsymbol{\beta}) \\ \vdots \\ a_i g^{S-1}(a_i) - \big(F(\varphi_{S-1}|\boldsymbol{\beta}) - F(\varphi_{S-2}|\boldsymbol{\beta})\big)\mu^{S-1}(\boldsymbol{\beta}) \end{bmatrix},$$

because not all of these moment conditions are independent. This is due to the fact that when the 'estimated' mean is fixed at $\hat{\mu}$, then $\sum_{i=1}^{I} \frac{a_i}{I} = \hat{\mu}$. So, similar to equation (6) above, we dropped one of the redundant moment conditions yielding a $2S - 2 \times 1$ matrix. In this case, the GMM estimator can be obtained by estimating equation (7) with the redefined $\mathbf{H}(\boldsymbol{\beta})$ and $\mathbf{W}(\boldsymbol{\beta})$.

Applying this constraint to the gamma distribution relates to $\theta$. Using the observed mean, we set $\theta = \frac{\hat{\mu}}{k}$ and minimize equation (7) with respect to $k$. Applying this constraint to the beta distribution relates to $\beta$. Using the observed mean and variance estimate, we set $\beta = \frac{\frac{\hat{\sigma}^2}{\hat{\mu}^2}(\alpha^2+\alpha)}{\left(1-\alpha\frac{\hat{\sigma}^2}{\hat{\mu}^2}\right)}$ and minimize equation (7) with respect to $\alpha$ and $\bar{\alpha}$.

## 4. Results

The results from our analysis are reported in two sections. First, the estimation results are discussed. This focuses on observations and distribution selection from the estimation process. Second, an analysis of farm size trends were performed and are discussed in section 4.2.

### 4.1 Estimation Results

We estimated distributions for farm size using the four methods discussed above, specifically, unconstrained and constrained estimates of both the gamma and beta distributions. Figure 2 plots the observed data along with each of the four estimated distributions for the 2000 United States farm size estimates. To compare the distributional estimates and determine which best represents the observed data, we used a set of criteria including successful fit rate, minimized value of the objective function, shape and difference between the estimated averages and the observed averages.

[Figure 2: Fitted Distributions for the United States, 2000]

We first examined the rate of successful fit because there are occasions when numerical minimization fails to produce a minimum, thus suggesting that a minimum does not exist. For about 9% of the observed distributions a fit was not possible for either the gamma or the beta distribution, regardless of whether or not we constrained the mean. In all but one case, Guatemala's 2000 census report, this happened because of evident errors in the data or missing area data. Most of the failed estimations were because countries reported area totals for a given farm size cohort that were mathematically impossible given the number of farms in that cohort. In most of these instances, the calculated average farm size in a cohort actually fell well outside that cohort (upper) bound.[7] For example, Sweden reports 1,140 farms and 26,530 ha in the 0 to 2 ha cohort for the year 2000. This implies that the average size of a farm within that cohort is 23.3 ha, well in excess of the 2 hectare upper limit of this farm size cohort. This problem often occurs in multiple cohorts for the affected country-year distribution. Table 2 describes the specific issues that occurred in each case that an estimate was not produced. In the case of Guatemala for the year 2000, the distribution is unusually and questionably flat from roughly 20 hectare to 9,032 hectare with most of the farms falling below 20 hectare (see Figure 3), this is likely because of unusual cohort classifications that distorted the shape of the distribution, resulting in incompatibility with either statistical distribution.[8]

[Table 3: Fitting Problems and Causes]

[Figure 3: Guatemala Farm Size Distribution, 2000]

Theoretically, the unconstrained distributions should fit better and more frequently than the constrained distributions. After setting aside the 9% of country-year distributions for which a (constrained or unconstrained) fitted distribution was unobtainable, the unconstrained gamma and beta each fit all but one distribution which represent 99.4% of the attempted distributions. The constrained gamma fitted all but two distributions for a 98.9% fit rate, while the constrained beta resolved for 96.9% of the distributions attempted.

A comparison of the minimized objective function values, which is essentially a pseudo weighted sum of square errors, for the gamma and beta distributions with and without constraining the means is more discerning. When the distributions were constrained to match the observed mean of the country-year observation, the gamma distribution returned a lower objective function value than the beta distribution for 92.2% of the observations. The difference was even more evident when the constraint was removed, wherein the gamma distribution yielded a lower objective function value for 96.9% of the observations. For the unconstrained estimates, we then compared the averages, 25$^{th}$, 50$^{th}$ and 75$^{th}$ percentiles across the fitted and observed distributions to gain a better understanding of why the gamma consistently outperformed the beta. The beta returned a higher average farm size than the unconstrained

[7] For the distributions that could not be fitted numerically this was a frequent problem spanning multiple census years. It was especially problematic in high-income, Western European countries such as Denmark, Finland, France, Norway, Sweden and others.

[8] Cohort classifications were (in hectare): 0.4-0.7, 0.7-1.4, 1.4-3.5, 3.5-7.1, 7.1-22.6, 22.6-45.2, 45.2-452, 452-903.2, 903.2-2258, 2258-4516, 4516-9032, 9032 and over

gamma more than 84% of the time. The gamma distribution yielded an average value closer to the observed mean than the beta distribution 88.6% of the time. However, many of the unconstrained mean estimates were questionable: only 13.9% of the estimates were within ten percent of the observed mean for gamma and not a single estimated mean was within ten percent of the observed mean for the beta. For the 25$^{th}$ percentile the gamma distribution returned a higher value 97.3% of the time when constrained and 89.1% of the time when unconstrained. The estimated median was higher for the gamma 92.3% of the time when using the constraint, while the beta estimated higher medians three-quarters of the time when the distributions were unconstrained. The 75$^{th}$ percentile was higher in 53.8% of the distributions for the constrained gamma estimates when compared to the constrained beta estimates. Without the constraint, the beta distribution yielded a higher 75$^{th}$ percentile over 85% of the time. Overall, the gamma distribution returned a better fit at the lower end of the distribution, where the bulk of the observations were clustered, and tended to underestimate the higher end because of its longer tail. Table 4 summarizes the criteria used in distribution selection.

[Table 4: Summary of Distribution Selection Criteria]

To sum up, we selected the gamma distribution for further analysis because it was at least as effective as the beta distribution for all of our selection criteria. For the subsequent analysis we also opted to only report the mean constrained estimates for a number of reasons. First, the unconstrained estimated means were rarely within 10% of the observed mean. While the estimated means for the gamma distribution were consistently closer to the observed means than the beta distribution, nonetheless, more than 85% of these estimated means deviated from the observed means by more than 10%. Second, although the primary data are less than ideal (e.g., truncation, inconsistent reporting standards) they are the data and the observed mean is one of the few “facts” we have about these distributions. For this reason we opted to anchor our distributional estimates around these reported sample means. We then used the gamma distribution estimates with the mean constraint to assess farm size trends and inequality over time, differentiating countries by income levels and geopolitical regions.

**4.2 Farm Size Findings**

We begin by assessing the measured change from, 1970 to 2000, in the distribution of average farm sizes worldwide using our imputed country-wide averages supplemented with data from Lowder et al. (2016).[9] The 2000 distribution is further subdivided into income classes. We then use our distribution estimates to assess the estimated change in different summary descriptive statistics to characterize changes in the distribution (i.e. beyond averages) of farm size.

*Average Farm Size, 1970 versus 2000*

We used the data for the observed distributions to calculate the average size of a farm for each country-year observation. These averages were pooled with any data Lowder et al. had that we did not, adding another 27 country-year observations for a total of 158. In the case of redundant observations between our set and Lowder et al.'s we retained our observation and discarded theirs. This is discussed in more detail in section (2). Income level and regional classifications were assigned based on World Bank's definitions for the year 1970 (see Appendix 1, Table 1 for classifications).

There were cases where a country had an observation for 1970 and not 2000, or vice versa, thus resulting in an unbalanced sample and potentially confounding over-time comparisons. We compared the distribution of average farm size for the full sample[10] with the frequency distribution[11] of averages for the consistent sample, the sample of 48 countries[12] that collected data in both 1970 and 2000.

[Figure 4: Frequency Distribution of Averages, 1970 versus 2000]

Figure 4 shows the frequency distributions of average farm size for 1970 and 2000 and includes both the full sample and the consistent sample of countries across census years. Visual inspection suggests that the full sample is fairly representative of the consistent sample. Temporal changes in the distribution for the consistent sample of countries were evident when using the full sample. In both census years a good portion of the countries had average farm sizes of less than 2 ha per farm; 21% of countries in 1970; 22% of countries in 2000 for the full

[9] Calculated by summing the area in and number of farms in each cohort to obtain the country's total area in farms and total number of farm and dividing the total area by the total number of farms.
[10] 81 countries for 1970; 77 countries for 2000.
[11] Distributions were compared over 0-100 ha, for both 1970 and 2000 only about 6% of countries reported average farm size over 100 ha.
[12] Our sample is missing two countries crucial to global food production; Australia and China. While we do have distribution data for the number of farms we do not have distributions of area, which is necessary for distribution estimation and calculation of average farm size. China alone accounts for over 200 million farms, with 99.7% being below 6.66 ha. Australia accounts for nearly 140 thousand farms and over 22% of them are greater than one thousand ha in size.

sample. We then looked at the distribution of average farm size across income levels for the full sample. Table 5 summarizes the key differences.

[Table 5: Summary of the Distribution of Average Farm Size]

Overall, average farm size in both high and low-middle income countries increased from 1970 and 2000 and in both years the average farm size in high-income countries was more than double the average size in low-middle income countries. In low-middle income countries over a quarter of farms were less than 2 ha in area for both 1970 and 2000. In contrast, in both 1970 and 2000, less than one-tenth of farms in high-income countries fell below that threshold. In 1970, 11.5% of high-income countries reported an average farm size of over 100 ha, increasing to 13.6% in 2000. Only 3.6% of low-middle income countries reported an average of more than 100 ha in both 1970 and 2000. A look at the frequency distributions of averages for each group showed significant differences in both 1970 and 2000 (Figure 5).

[Figure 5: Frequency Distribution of Averages by Income Level, 1970 and 2000]

There were major differences in the distributions of averages for high and low-middle income countries. The distribution was extremely right skewed for low-middle income countries in 1970 and became increasingly skewed to the right by 2000. High-income countries were trending the other way, their mean distribution was still somewhat right skewed in both years, though not nearly to the extent of the distribution for low-middle income countries, and it was less skewed in 2000 than in 1970. The distributions of average farm size in high and low-middle income countries exhibited fundamental differences in both 1970 and 2000. To determine if they are becoming more or less similar over time we examined changes in average farm size for a consistent sample of 48 countries, 21 of which were classified as high-income and 27 as low-middle income. Table 6 summarizes the results.

[Table 6: Differing Trends in Average Farm Size at Different Income Levels, 1970 to 2000]

For the group of high-income countries, 81% experienced an increase in average farm size between 1970 and 2000. Low-middle income countries trended in the opposite direction, with 70% of the group experiencing decreasing average farm size between 1970 and 2000.

*Farm Size Distributions, 1970 versus 2000*

We used our estimated distributions for the consistent sample of countries (i.e. countries with observations for both 1970 and 2000) to examine differences across income levels in measures of central tendency (e.g., mean and median), dispersion (e.g., standard deviation and the Gini coefficient) and symmetry (e.g., skewness[13] and percentiles). Our consistent sample consisted of 30 countries; 20 in the low-middle income group and 10 in the high-income group.[14] Table 7 summarizes the temporal changes in the statistics for these groups.

[Table 7: Changes in Select Distribution Statistics, 1970 versus 2000]

We found that average farm size trends for this group of 30 countries were similar to those observed in the full sample and the consistent average farm size sample (i.e., the 48 country sample discussed above). In the high-income group, 7 out of 10 countries experienced an increase in average farm size while 80% of the low-middle income countries exhibited a decrease in average farm size. Median farm size increased in 6 of the 10 high-income countries but only 6 of the 20 low-middle income countries. The standard deviation of the country level farm size distribution decreased in 65% of the low-middle income countries, compared with only 30% of the high-income countries. We calculated the Gini coefficient for each country-year observation and found that it increased in 60% of the high-income countries and 50% of the low-middle income countries. For those countries with increasing Gini coefficients, inequality in the farm size distributions of these countries was increasing, but the change in the structure of farm size distribution is markedly different in high versus low-medium income countries. In the high-income countries, cropland is consolidating within the largest farms, while in the low-middle income countries the smallest farms are becoming even smaller and growing in number. The skewness statistic indicated that 60% of high-income farm size distributions became less right skewed while half of the low-middle income countries did the same.

The percentile estimates provided more evidence for the distribution trend differences across income levels. At the 5th percentile[15], 7 out of 10 high-income countries exhibited an increase compared with low-middle income countries where 65% experienced a decrease. This means that the smallest farms (relative to the country's distribution) in most high-income countries are growing in size while those same small farms are shrinking in most low-middle income countries. At the 95th percentile, the largest farms in the country, farm size was

---

[13] Skewness is a statistic derived from the gamma distribution's $k$ parameter, it is always positive for the gamma distribution because the gamma distribution is right skewed by definition, as it increases the gamma distribution approaches the normal distribution in shape (i.e. an increase in the skewness statistic means that the distribution became less right skewed). The formula for skewness is $\frac{2}{\sqrt{k}}$.

[14] See Appendix Table 1 notes

[15] The Xth percentile in these distributions describes the area value that X% of farms fall beneath.

increasing in 70% of the high-income countries and decreasing in 75% of the low-middle income countries. This pattern was consistent in the small to medium and medium to large farms of these countries as well (see Table 7); that is, most high-income countries experienced increasing farm size at all size levels while most low-middle income countries experienced decreasing farm size at all size levels.

Our farm size distribution estimates reveal that the structure of farm sizes in high-income countries has evolved in significantly different ways from that in low-middle income countries. While our results regarding average farm size were consistent with those reported by Lowder et al. Thus, using more detailed distributional data on farm sizes has yielded more nuanced insights into the changing nature of farm sizes within high versus low-medium income countries.

**5. Conclusion**

We estimated gamma and beta distributions to fit observed farm size distributions for a group of countries with data reporting the distribution of farms and farm area in 1970 and 2000 using the Generalized Method of Moments. We found that unconstrained estimations of the distributions frequently resulted in a differences of over 20% between the estimated average farm size and the observed average farm size. To account for this we imposed a constraint on the optimization procedure that required the final parameter estimates to estimate a distribution with an overall mean equal to that observed in the data. In comparing the beta and gamma distributions' objective function values and shapes, we determined that on balance the gamma distribution was more effective at representing the reported distribution of farm size. The fitted distributions provided estimates for median farm size as well as the $5^{th}$, $25^{th}$, $75^{th}$ and $95^{th}$ percentiles, Gini coefficients and skewness. We used these data and drew on the outline provided in Lowder et al. (2016) to compare temporal trends in farm size across income and region groups. When pooling our mean farm size data with Lowder et al.'s (2016), we found similar results. Specifically, there is a distinctly different structure to the change in farm size for low-medium income countries versus high-income countries. The high-income country group experienced increases more often than decreases at all size levels. Low-middle income countries mostly experienced decreasing farm size across the board. Our estimates indicated that the distribution of farm size was trending in opposite directions for high and low-middle income countries.

There are many possibilities for future research. The first would be to increase the sample size. We worked with what census data were available and there were many observations available but the sample size was cut drastically when we began examining trends because many of the countries in the sample only had a distribution for one year. An increased effort by the FAO and organizations within countries to collect distribution data as well as average area data would be ideal, with explicit reporting standards and size definitions. In some cases the fitted distributions do not appear to fit the data remarkably well, often due to disproportionate cohort bounds where cohorts on the low end of farm size are small in range but cohorts are large and could be more descriptive at the high end. Developing a way to determine which category bounds are ideal for fitting and analysis would be a valuable extension of this work. Methodologically, there is the issue of truncation brought on by countries placing an arbitrary upper or lower bound on the area (or value) per farm they include in their sampling frames, which skews the observed data and may have an impact on the results. This can be handled with similar methodology used in this paper being applied with truncated statistical distributions. It would be worthwhile to use a wider range of distributions in addition to beta and gamma to determine if the gamma distribution really is the best for representing the observed farm size data or, relatedly, determine if a flexible approach in which the distribution is allowed to vary across countries and years in order to obtain the best possible estimates.

## References


Adamopoulos, T., and D. Restuccia "The Size Distribution of Farms and International Productivity Differences." *American Economic Review* 104 (2014):1667-97.

Barrett, C.B., M.F. Bellemare, and J.Y. Hou "Reconsidering Conventional Explanations of the Inverse Productivity–Size Relationship." *World Development* 38 (2010):88-97.

Cameron, A.C. and P.K. Trivedi. *Microeconometrics: Methods and Applications*. Cambridge; New York: Cambridge University Press, 2005.

Carter, M.R. "Identification of the Inverse Relationship between Farm Size and Productivity: An Empirical Analysis of Peasant Agricultural Production." *Oxford Economic Papers* 36 (1984):131-45.

Chotikapanich, D., W.E. Griffiths, and D.S.P. Rao "Estimating and Combining National Income Distributions Using Limited Data." *Journal of Business & Economic Statistics* 25 (2007):97-109.

Eastwood, R., M. Lipton, and A. Newell. "Farm Size." Chapter 65. *Handbook of Agricultural Economics* Volume 4. Amsterdam: Elsevier, 2010.

Fan, S., & Chan-Kang, C. (2005). Is small beautiful? Farm size, productivity, and poverty in Asian agriculture. *Agricultural Economics*, *32*, 135-146.

Food and Agriculture Organization of the,United Nations *1970 world census of agriculture : analysis and international comparison of the results*. Rome: Food and Agriculture Organization of the United Nations; 1981.

———*1980 World Census of Agriculture: Methodological Review*. Rome: Food and Agriculture Organization of the United Nations, 1992.

———*2000 World Census of Agriculture: Analysis and International Comparison of the Results (1996-2005)*. Rome: Food and Agriculture Organization of the United Nations, 2013.

———*Report of the 1990 World Census of Agriculture: International Comparison and Primary Results by Country (1986-1995)*. Rome: Food and Agriculture Organization of the United Nations, 1997.

———*World Programme for the Census of Agriculture 2020*. Rome: FAO, 2015.

Gholamreza, H., W. Griffiths, J. Brice, D.S.P. Rao, and D. Chotikapanich "Inference for Income Distributions Using Grouped Data." *Journal of Business & Economic Statistics* 30 (2012):563-75.

Hall, B.H. "Empirical Analysis of the Size Distribution of Farms: Discussion." *American Journal of Agricultural Economics* 69 (1987):486-7.

Heltberg, R. "Rural Market Imperfections and the Farm Size— Productivity Relationship: Evidence from Pakistan." *World Development* 26 (1998):1807-26.

Ju, X., B. Gu, Y. Wu, and J.N. Galloway "Reducing China's fertilizer use by increasing farm size." *Global Environmental Change* (2016): 26-32.

Lowder, S.K., J. Skoet, and T. Raney "The Number, Size, and Distribution of Farms, Smallholder Farms, and Family Farms Worldwide." *World Development 87* (2016a): 16-29.

——— Raney "The Number, Size, and Distribution of Farms, Smallholder Farms, and Family Farms Worldwide." Web Appendix A: Supplementary Data. *World Development 87*. (2016b): 16-29. Available at: http://dx.doi.org/10.1016/.

Mottaleb, K.A., and S. Mohanty "Farm size and profitability of rice farming under rising input costs." *Journal of Land Use Science* 10 (2015):243-55.

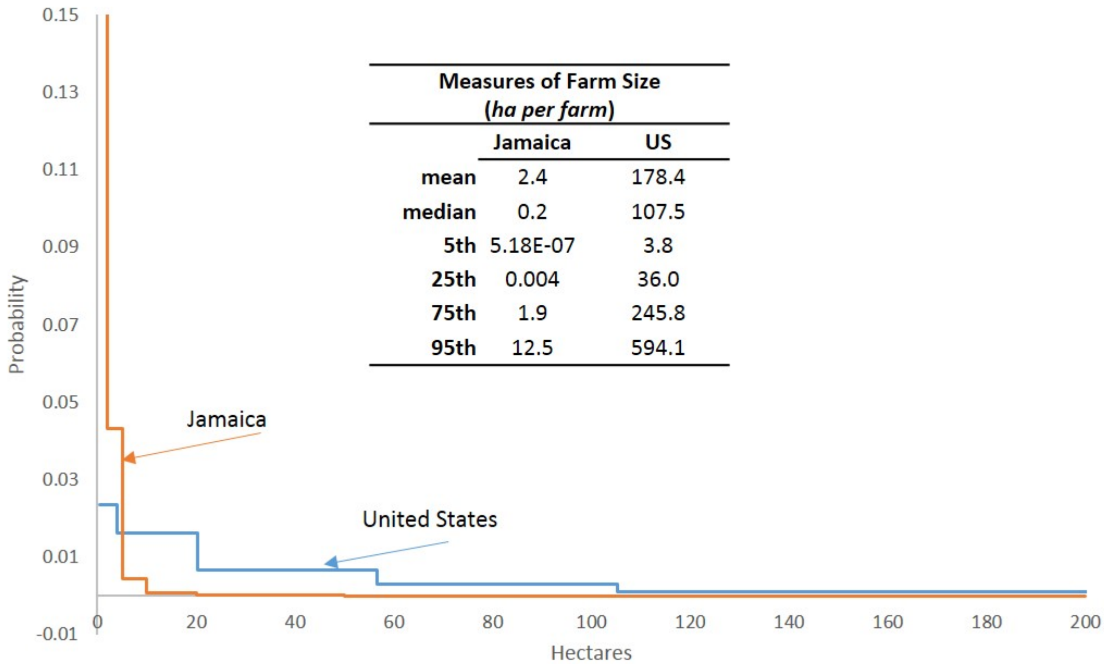
Sumner, D.A. "American Farms Keep Growing: Size, Productivity, and Policy." *Journal of Economic Perspectives* 28 (2014):147-66.

Van Zÿl, J. *The relationship between farm size and efficiency in South African agriculture*. Washington, DC: Washington, DC: World Bank, Agriculture and Natural Resources Dept., Office of the Director, 1995.

World Bank. *WDI online*. Washington, DC: World Bank Group, 2016.

## Tables & Figures

*Figure 1: United States vs Jamaica Farm Size Distributions, 2000*




| Measures of Farm Size (*ha per farm*) | | |
|---|---|---|
| | **Jamaica** | **US** |
| **mean** | 2.4 | 178.4 |
| **median** | 0.2 | 107.5 |
| **5th** | 5.18E-07 | 3.8 |
| **25th** | 0.004 | 36.0 |
| **75th** | 1.9 | 245.8 |
| **95th** | 12.5 | 594.1 |

*Source:* FAO (2013).
*Note:* For plotting purposes, truncated at 200 hectares as well as the probability level of 0.15.

*Table 1: Example of Disparity in Reported Farm Size Cohorts*

| | | **Distrib data** | **Distribution Cohort** | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Data Year** | **Number of farms** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** |
| Australia | 1970 | 249,485 | <1 | 1 - 2 | 2 - 5 | 5 - 10 | 10 - 20 | 20 - 50 | 50 - 100 | 100 - 200 | 200 - 500 | 500 - 1000 | >1000 |
| Austria | 1970 | 362,216 | <1 | 1 - 2 | 2 - 5 | 5 - 10 | 10 - 20 | 20 - 50 | 50 - 100 | 100 - 200 | >200 | | |
| Guatemala | 2000 | 830,684 | 0.4 - 0.7 | 0.7 - 1.4 | 3.5 - 7.1 | 7.1 - 22.6 | 22.6 - 45.2 | 45.2 - 452 | 452 - 903.2 | 903.2 - 2258 | 2258 - 4516 | 4516 - 9032 | >9032 |
| Ireland | 2000 | 141,500 | 0 - 2 | 2 - 5 | 5 - 10 | 10 - 20 | 20 - 30 | 30 - 50 | 50 - 100 | >100 | | | |
| Republic of Korea | 1970 | 2,401,211 | <1 | 1 - 2 | >2 | | | | | | | | |
| Lesotho | 1970 | 176,651 | <1 | 1 - 2 | 2 - 5 | >5 | | | | | | | |
| Luxembourg | 1990 | 3,803 | <1 | 1 - 2 | 2 - 5 | 5 - 10 | 10 - 20 | 20 - 50 | 50 - 100 | 100 - 200 | | | |
| Rwanda | 1990 | 1,111,897 | <0.25 | 0.25 - 0.5 | 0.5 - 0.75 | 0.75 - 1 | 1 - 1.5 | 1.5 - 2 | >2 | | | | |
| Tanzania | 1970 | 2,434,425 | <1 | 1 - 2 | 2 - 5 | 5 - 10 | 10 - 20 | >20 | | | | | |
| United States | 1990 | 2,087,759 | 0.4 - 4 | 4 - 20.2 | 20.2 - 56.7 | 56.7 - 105.2 | 105.2 - 202.4 | 202.4 - 404.7 | 404.7 - 809.4 | >809.4 | | | |
| FAO Classifications | | | 0-1 | 1-2 | 2-5 | 5-10 | 10-20 | 20-50 | 50-100 | 100-200 | 200-500 | 500-1000 | >1000 |

*Source:* FAO (1981, 1997, 2013).

*Note:* Unit of measure is hectare.

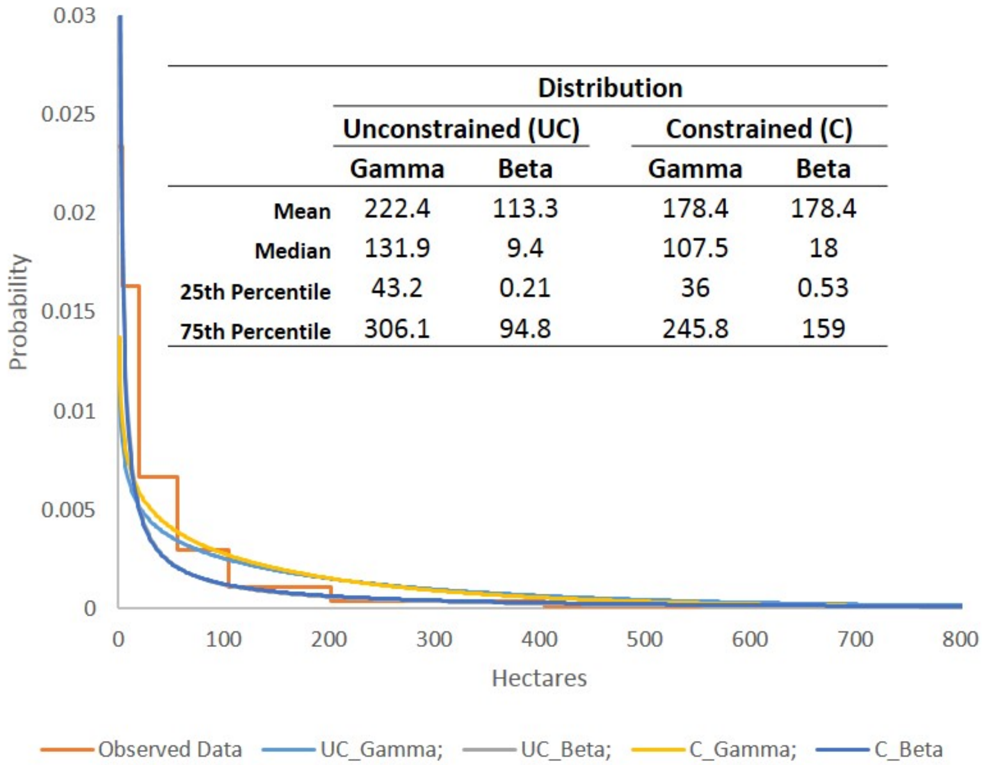
*Table 2: Summary of InSTePP Distributions and Lowder et al. (2016) Means*

| | **Census Years** | | | | | |
|---|---|---|---|---|---|---|
| | **InSTePP Distribution Data** | | | **Lowder Mean Size Data** | | |
| **Income Class** | **1970** | **1990** | **2000** | **1970** | **1990** | **2000** |
| **High** | 27 (*0*) | 21 (*4*) | 22 (*5*) | 23 | 21 | 20 |
| **Upper Middle** | 16 (*0*) | 13 (*3*) | 18 (*2*) | 12 | 10 | 11 |
| **Lower Middle** | 20 (*0*) | 17 (*2*) | 31 (*2*) | 19 | 17 | 22 |
| **Low** | 17 (*0*) | 11 (*5*) | 21 (*12*) | 12 | 11 | 9 |
| **Total** | 80 (*0*) | 62 (*14*) | 92 (*21*) | 66 | 59 | 62 |

*Sources:* Lowder et al. (2016b); FAO (1981, 1997, 2013); InSTePP (2016)

*Note:* () indicates distributional data that reports either the number of farms or the total area in each cohort but not both, numbers represent number of countries with data

*Figure 2: Fitted Distributions for the United States, 2000*




*Source:* FAO (2013).
*Note*: UC indicates unconstrained; C indicates constrained estimation. Truncated at .03 and 800.

*Table 3: Fitting Problems and Causes*

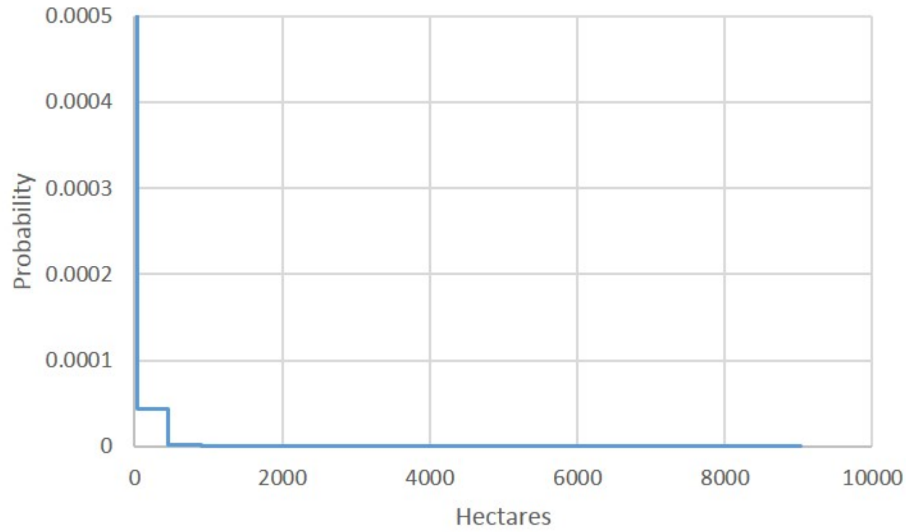
| Country | Year | Cause |
|---|---|---|
| Finland | 1970 | Incorrect area or number of farms in cohort(s)[b] |
| France | 1970 | Incorrect area or number of farms in cohort(s)[b] |
| Germany, Fed Rep | 1970 | Incorrect area or number of farms in cohort(s)[b] |
| Israel | 1970 | Incorrect area or number of farms in cohort(s)[b] |
| Norway | 1970 | Incorrect area or number of farms in cohort(s)[b] |
| Sweden | 1970 | Incorrect area or number of farms in cohort(s)[b] |
| Finland | 1990 | Incorrect area or number of farms in cohort(s)[b] |
| Korea, Rep of | 1990 | Incorrect area or number of farms in cohort(s)[b] |
| Thailand | 1990 | Incorrect area or number of farms in cohort(s)[b] |
| Cape Verde | 2000 | Missing Data |
| Denmark | 2000 | Incorrect area or number of farms in cohort(s)[b] |
| Finland | 2000 | Incorrect area or number of farms in cohort(s)[b] |
| Georgia | 2000 | Incorrect area or number of farms in cohort(s)[b] |
| Guatemala | 2000 | Unusual Distribution[a] |
| Guinea | 2000 | Missing Data |
| Norway | 2000 | Incorrect area or number of farms in cohort(s)[b] |
| Poland | 2000 | Incorrect area or number of farms in cohort(s)[b] |
| Portugal | 2000 | Incorrect area or number of farms in cohort(s)[b] |
| Serbia | 2000 | Incorrect area or number of farms in cohort(s)[b] |
| Sweden | 2000 | Incorrect area or number of farms in cohort(s)[b] |
| Togo | 2000 | Missing Data |

*Note:* Missing data refers to distributions with gaps in cohorts or data points.

[a]This instance is explicitly discussed in section (4.1) Estimation Results

[b]See footnote 7 for explanation of this issue

*Figure 3: Guatemala Farm Size Distribution, 2000*




*Source:* FAO (2013).

*Table 4: Summary of Distribution Selection Criteria*

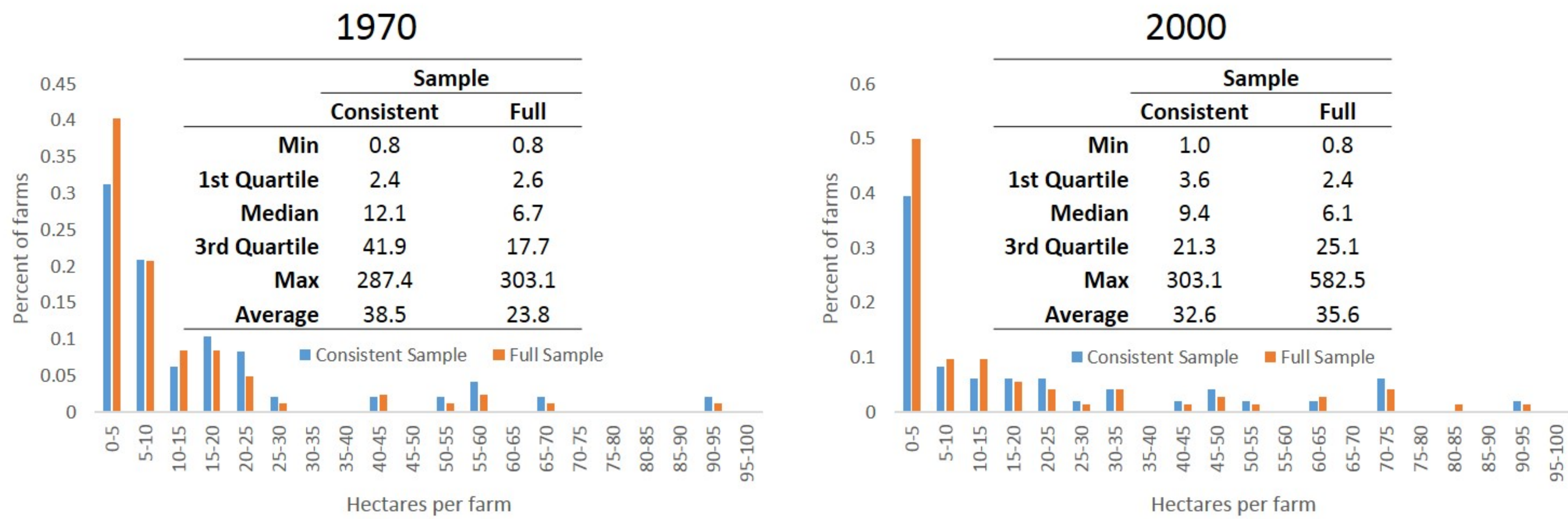
| | | Criteria | | | | | |
|---|---|---|---|---|---|---|---|
| | | | Estimated Means Within _% of the Observed Mean | | | | |
| | Successful Fit Rate | Estimates with Lowest Objective Function Value | 10% | 20% | 30% | 50% | 75% |
| **Unconstrained** | | *Percent* | | | | | |
| **Beta** | 99.4 | 3.1 | 0.0 | 18.6 | 24.2 | 30.2 | 36.3 |
| **Gamma** | 99.4 | 96.9 | 13.9 | 52.1 | 60.9 | 69.3 | 77.7 |
| **Constrained** | | | | | | | |
| **Beta** | 96.9 | 7.8 | na | na | na | na | na |
| **Gamma** | 98.9 | 92.2 | na | na | na | na | na |

*Note:* Successful fit rate does not include distributions that were not successful for any of the 4 options (constrained/unconstrained gamma/beta)

[na]Indicates not applicable as the estimated mean size is always equal to the observed mean size in the constrained approach

*Note:* Successful fit rate does not include distributions that were not successful for any of the 4 options (constrained/unconstrained gamma/beta)

*Figure 4: Frequency Distribution of Averages, 1970 versus 2000*




*Note:* Plots indicate share of countries that fall within a particular *average* farm-size cohort. The census consistent sample includes 48 countries. The full sample includes 81 countries in 1970 and 77 countries in 2000. Cohorts are inclusive at the low end and exclusive at the high end (e.g., [0, 5), [5, 10), etc.)

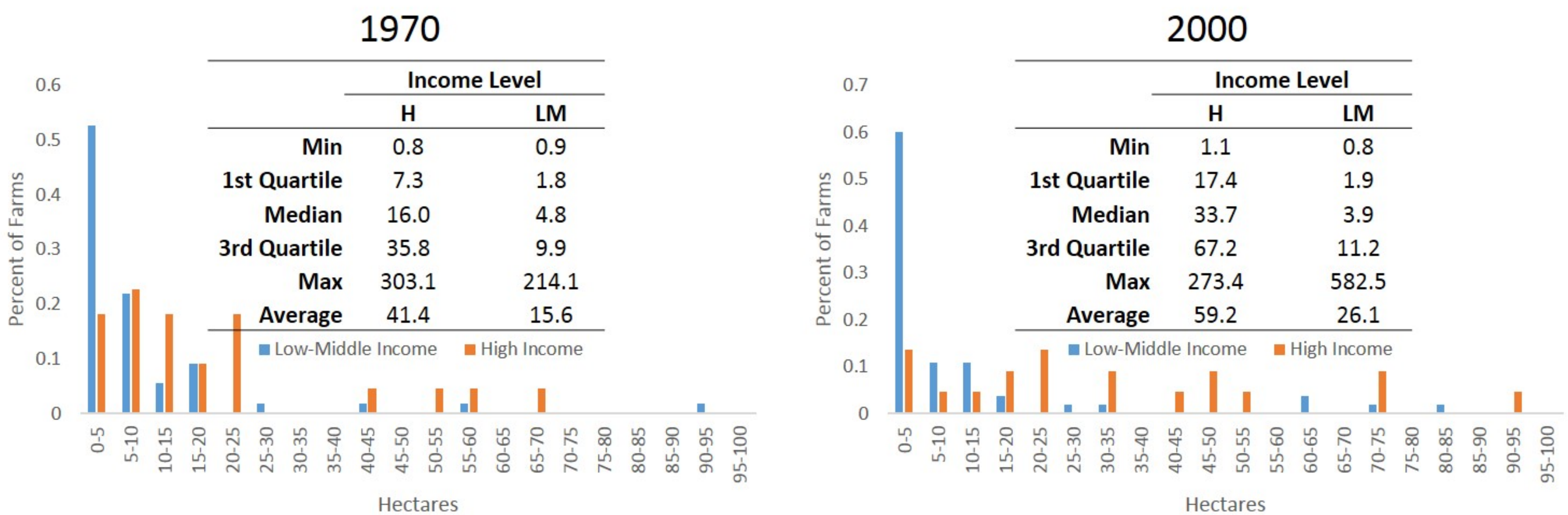
*Table 5: Summary of the Distribution of Average Farm Size*

| | Census Year | | | | | |
|---|---|---|---|---|---|---|
| | 1970 | | | 2000 | | |
| | H | L-M | All | H | L-M | All |
| **Sample Size (countries)** | 26 | 55 | 81 | 22 | 55 | 77 |
| **Share of Averages <2 ha[a]** | 0.08 | 0.27 | 0.21 | 0.09 | 0.27 | 0.22 |
| **Share of Averages >100 ha[a]** | 0.12 | 0.04 | 0.06 | 0.14 | 0.04 | 0.06 |
| **Standard Deviation (ha)** | 69.94 | 36.60 | 50.79 | 73.20 | 86.96 | 84.15 |
| **Average Farm Size (ha)** | 41.35 | 15.56 | 23.84 | 59.25 | 26.10 | 35.57 |

*Note*: H indicates high-income countries, L-M indicates low-middle income countries and All indicates the full sample.

[a]Represents the share of countries with average farm sizes less than 2 ha or greater than 100 ha.

*Figure 5: Frequency Distribution of Averages by Income Level, 1970 and 2000*




**1970**

| | Income Level | |
|---|---|---|
| | H | LM |
| Min | 0.8 | 0.9 |
| 1st Quartile | 7.3 | 1.8 |
| Median | 16.0 | 4.8 |
| 3rd Quartile | 35.8 | 9.9 |
| Max | 303.1 | 214.1 |
| Average | 41.4 | 15.6 |

**2000**

| | Income Level | |
|---|---|---|
| | H | LM |
| Min | 1.1 | 0.8 |
| 1st Quartile | 17.4 | 1.9 |
| Median | 33.7 | 3.9 |
| 3rd Quartile | 67.2 | 11.2 |
| Max | 273.4 | 582.5 |
| Average | 59.2 | 26.1 |

*Notes:* See Figure 4.

*Table 6: Differing Trends in Average Farm Size at Different Income Levels, 1970 to 2000*

| | Share of Countries that Exhibited | | |
|---|---|---|---|
| | **Increase in Average Farm Size** | **Decrease in Average Farm Size** | **Sample Size** |
| **High-income** | 0.81 | 0.19 | 21 |
| **Low-Middle Income** | 0.30 | 0.7 | 27 |
| **Total** | 0.52 | 0.48 | 48 |

*Table 7: Changes in Select Distribution Statistics, 1970 versus 2000*

| | Number of Countries | | | | | | Share of Countries | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Increased | | | Decreased | | | Increased | | | Decreased | | |
| | **W** | **H** | **L-M** | **W** | **H** | **L-M** | **W** | **H** | **L-M** | **W** | **H** | **L-M** |
| **Measures of Central Tendency** | | | | | | | | | | | | |
| **Mean** | 11 | 7 | 4 | 19 | 3 | 16 | 0.37 | 0.70 | 0.20 | 0.63 | 0.30 | 0.80 |
| **Median** | 12 | 6 | 6 | 18 | 4 | 14 | 0.40 | 0.60 | 0.30 | 0.60 | 0.40 | 0.70 |
| **Measures of Dispersion** | | | | | | | | | | | | |
| **St Dev.** | 14 | 7 | 7 | 16 | 3 | 13 | 0.47 | 0.70 | 0.35 | 0.53 | 0.30 | 0.65 |
| **Gini Coefficient** | 16 | 6 | 10 | 14 | 4 | 10 | 0.53 | 0.60 | 0.50 | 0.47 | 0.40 | 0.50 |
| **Measures of Symmetry** | | | | | | | | | | | | |
| **Skewness** | 16 | 6 | 10 | 14 | 4 | 10 | 0.53 | 0.60 | 0.50 | 0.47 | 0.40 | 0.50 |
| **Percentiles** | | | | | | | | | | | | |
| 5th | 14 | 7 | 7 | 16 | 3 | 13 | 0.47 | 0.70 | 0.35 | 0.53 | 0.30 | 0.65 |
| 25th | 13 | 6 | 7 | 17 | 4 | 13 | 0.43 | 0.60 | 0.35 | 0.57 | 0.40 | 0.65 |
| 75th | 14 | 7 | 7 | 16 | 3 | 13 | 0.47 | 0.70 | 0.35 | 0.53 | 0.30 | 0.65 |
| 95th | 12 | 7 | 5 | 18 | 3 | 15 | 0.40 | 0.70 | 0.25 | 0.60 | 0.30 | 0.75 |

*Note*: H indicates the high-income countries sampled, L-M indicates the low-middle income countries sampled and W indicates the combination of both groups.

**Appendix 1**

Table 1: Summary of Distribution Data by Country, Year, Income Level and Region

| | Census Year | | | | | | World Bank Classification | |
|---|---|---|---|---|---|---|---|---|
| Country | 1970 | | 1990 | | 2000 | | Income Level | Region |
| Albania | | | x | | x | | LM | Europe & Central Asia |
| Algeria[1] | x | | | | x | | UM | Middle East & Northern Africa |
| American Samoa[1] | x | L | x | L | x | L | H | H |
| Argentina | | | x | L | | L | UM | Latin America & the Caribbean |
| Australia | x | | x | | x | | H | H |
| Austria | x | L | x | L | x | L | H | H |
| Bahamas | | | x | L | | | H | H |
| Bahrain | x | L | | | | | H | H |
| Bangladesh | | | | | | L | L | South Asia |
| Barbados | | L | x | L | | | UM | Latin America & the Caribbean |
| Belgium[1] | x | L | x | L | x | L | H | H |
| Belize | | L | | | | | LM | Latin America & the Caribbean |
| Botswana[1] | x | | | | x | | LM | Sub-Saharan Africa |
| Brazil[1] | x | L | x | L | x | L | UM | Latin America & the Caribbean |
| Burkina Faso | | | x | | | | L | Sub-Saharan Africa |
| Cabo Verde | | | | L | x | L | LM | Sub-Saharan Africa |
| Cameroon | x | | | | | | LM | Sub-Saharan Africa |
| Canada | x | L | x | L | x | L | H | H |
| Central African Republic | x | L | | | | | L | Sub-Saharan Africa |
| Chad | x | | | | | | L | Sub-Saharan Africa |
| Chile | | | | | x | L | LM | Latin America & the Caribbean |
| China | | | | | x | | L | East Asia & the Pacific |
| Colombia[1] | x | L | x | L | x | L | LM | Latin America & the Caribbean |
| Congo | x | L | x | | | | LM | Sub-Saharan Africa |
| Cook Islands | | | x | L | x | L | * | East Asia & the Pacific |
| Costa Rica | x | L | | | | | LM | Latin America & the Caribbean |
| Côte d'Ivoire[1] | x | L | | | x | L | LM | Sub-Saharan Africa |
| Croatia | | | | | x | | LM | Europe & Central Asia |
| Cyprus | | | x | L | x | L | UM | Europe & Central Asia |
| Czech Republic | | | | | x | | LM | Europe & Central Asia |
| Czechoslovakia | x | | | | | | UM | Europe & Central Asia |
| Denmark | x | L | x | L | x | L | H | H |
| Djibouti | | | x | | | | LM | Middle East & Northern Africa |
| Dominica | | | x | | | | LM | Latin America & the Caribbean |
| Dominican Republic | x | L | | | | | LM | Latin America & the Caribbean |
| Ecuador[1] | x | L | | | x | L | LM | Latin America & the Caribbean |
| Egypt | | | | L | x | L | LM | Middle East & Northern Africa |

| | Census Year | | | | | | World Bank Classification | |
|---|---|---|---|---|---|---|---|---|
| **Country** | **1970** | | **1990** | | **2000** | | **Income Level** | **Region** |
| El Salvador | **x** | **L** | | | | | LM | Latin America & the Caribbean |
| Estonia | | | | | **x** | | UM | Europe & Central Asia |
| Ethiopia | | | **x** | **L** | **x** | **L** | L | Sub-Saharan Africa |
| Fiji | **x** | **L** | **x** | **L** | | | LM | East Asia & the Pacific |
| Finland | **x** | **L** | **x** | **L** | **x** | **L** | H | H |
| France | **x** | **L** | **x** | **L** | **x** | **L** | H | H |
| French Guiana | | | **x** | **L** | **x** | **L** | * | Latin America & the Caribbean |
| Gabon | **x** | | | | | | UM | Sub-Saharan Africa |
| Georgia | | | | | **x** | | LM | Europe & Central Asia |
| Germany | **x** | **L** | **x** | **L** | **x** | **L** | H | H |
| Ghana | **x** | | **x** | | **x** | | L | Sub-Saharan Africa |
| Greece[1] | **x** | **L** | **x** | **L** | **x** | **L** | UM | Europe & Central Asia |
| Grenada | | | **x** | **L** | | | LM | Latin America & the Caribbean |
| Guadeloupe | **x** | **L** | **x** | **L** | | **L** | * | Latin America & the Caribbean |
| Guam[1] | **x** | **L** | **x** | **L** | **x** | **L** | H | H |
| Guatemala | | | | | **x** | **L** | LM | Latin America & the Caribbean |
| Guinea | | | **x** | **L** | **x** | | L | Sub-Saharan Africa |
| Haiti | **x** | | | | | | L | Latin America & the Caribbean |
| Honduras | **x** | **L** | **x** | **L** | | | LM | Latin America & the Caribbean |
| Hungary | **x** | **L** | | | **x** | **L** | UM | Europe & Central Asia |
| India[1] | **x** | **L** | **x** | **L** | **x** | **L** | L | South Asia |
| Indonesia | **x** | **L** | | **L** | **x** | **L** | L | East Asia & the Pacific |
| Iran (Islamic Republic of) | | | **x** | **L** | **x** | **L** | UM | Middle East & Northern Africa |
| Iraq | **x** | **L** | | | | | UM | Middle East & Northern Africa |
| Ireland[1] | **x** | **L** | **x** | **L** | **x** | **L** | H | H |
| Israel | **x** | **L** | | **L** | | | H | H |
| Italy[1] | **x** | **L** | **x** | **L** | **x** | **L** | H | H |
| Jamaica[1] | **x** | **L** | | | **x** | **L** | LM | Latin America & the Caribbean |
| Japan | **x** | **L** | | **L** | | **L** | H | H |
| Jordan | | **L** | | | **x** | **L** | LM | Middle East & Northern Africa |
| Kenya | **x** | **L** | | | **x** | | L | Sub-Saharan Africa |
| Kuwait | **x** | | | | | | H | H |
| Kyrgyzstan | | | | | **x** | | LM | Europe & Central Asia |
| Laos | | | | | **x** | | L | East Asia & the Pacific |
| Latvia | | | | | **x** | | UM | Europe & Central Asia |
| Lebanon[1] | **x** | **L** | | | **x** | **L** | LM | Middle East & Northern Africa |
| Lesotho | **x** | **L** | **x** | **L** | **x** | | L | Sub-Saharan Africa |
| Liberia | **x** | | | | | | L | Sub-Saharan Africa |
| Libya | | | **x** | | | | UM | Middle East & Northern Africa |
| Libyan Arab Jamahiriya | | **L** | | **L** | | **L** | | |

| | Census Year | | | | | | World Bank Classification | |
|---|---|---|---|---|---|---|---|---|
| Country | 1970 | | 1990 | | 2000 | | Income Level | Region |
| Lithuania | | | | | **x** | | UM | Europe & Central Asia |
| Luxembourg[1] | **x** | **L** | **x** | **L** | **x** | **L** | H | H |
| Madagascar | | | | | | **L** | L | Sub-Saharan Africa |
| Malawi | **x** | **L** | | **L** | | | L | Sub-Saharan Africa |
| Mali | | | | | **x** | | L | Sub-Saharan Africa |
| Malta[1] | **x** | **L** | | | **x** | **L** | UM | Europe & Central Asia |
| Martinique | | | **x** | **L** | | **L** | * | Latin America & the Caribbean |
| Mexico | **x** | **L** | **x** | **L** | | | LM | Latin America & the Caribbean |
| Morocco | | | | | **x** | **L** | LM | Middle East & Northern Africa |
| Mozambique | | | | | **x** | | L | Sub-Saharan Africa |
| Myanmar | | | **x** | **L** | **x** | **L** | L | East Asia & the Pacific |
| Namibia | | | | | **x** | | LM | Sub-Saharan Africa |
| Nepal[1] | **x** | **L** | **x** | **L** | **x** | **L** | L | South Asia |
| Netherlands[1] | **x** | **L** | **x** | **L** | **x** | **L** | H | H |
| New Caledonia | | | | **L** | | **L** | UM | East Asia & the Pacific |
| New Zealand | **x** | **L** | | **L** | **x** | **L** | H | H |
| Nicaragua | | | | | **x** | **L** | LM | Latin America & the Caribbean |
| Niger | | | | | **x** | | L | Sub-Saharan Africa |
| Nigeria | | | | | **x** | | L | Sub-Saharan Africa |
| Northern Mariana Islands | | | **x** | **L** | **x** | **L** | LM | East Asia & the Pacific |
| Norway | **x** | | **x** | | **x** | | H | H |
| Pacific Islands | **x** | | | | | | * | East Asia & the Pacific |
| Pakistan[1] | **x** | **L** | **x** | **L** | **x** | **L** | L | South Asia |
| Panama[1] | **x** | **L** | **x** | **L** | **x** | **L** | UM | Latin America & the Caribbean |
| Paraguay | | | **x** | **L** | | | LM | Latin America & the Caribbean |
| Peru | **x** | **L** | **x** | **L** | | | LM | Latin America & the Caribbean |
| Philippines[1] | **x** | **L** | **x** | **L** | **x** | **L** | LM | East Asia & the Pacific |
| Poland | **x** | **L** | | **L** | **x** | **L** | LM | Europe & Central Asia |
| Portugal | **x** | **L** | **x** | **L** | **x** | **L** | UM | Europe & Central Asia |
| Puerto Rico[1] | **x** | | **x** | | **x** | | UM | Latin America & the Caribbean |
| Qatar | | | | | **x** | | H | H |
| Republic of Korea | **x** | **L** | **x** | **L** | | | UM | East Asia & the Pacific |
| Réunion | **x** | **L** | **x** | **L** | **x** | **L** | * | Sub-Saharan Africa |
| Romania | | | | | **x** | | UM | Europe & Central Asia |
| Rwanda | | | **x** | | **x** | | * | * |
| Saint Kitts and Nevis | | | **x** | | | | UM | Latin America & the Caribbean |
| Saint Lucia[1] | **x** | **L** | **x** | **L** | **x** | **L** | LM | Latin America & the Caribbean |
| Saint Vincent and the Grenadines | | | **x** | **L** | **x** | **L** | LM | Latin America & the Caribbean |
| Samoa | | | | **L** | **x** | **L** | LM | East Asia & the Pacific |
| Saudi Arabia | **x** | **L** | | | | **L** | H | H |

| | Census Year | | | | | | World Bank Classification | |
|---|---|---|---|---|---|---|---|---|
| **Country** | **1970** | | **1990** | | **2000** | | **Income Level** | **Region** |
| Senegal | | | | | **x** | **L** | LM | Sub-Saharan Africa |
| Serbia | | | | | **x** | | UM | Europe & Central Asia |
| Sierra Leone | **x** | **L** | | | | | L | Sub-Saharan Africa |
| Singapore | **x** | | | | | | H | H |
| Slovakia | | | | | **x** | | LM | Europe & Central Asia |
| Slovenia | | | **x** | **L** | | **L** | UM | Europe & Central Asia |
| South Africa | | | **x** | | **x** | | * | * |
| Spain | | **L** | **x** | **L** | **x** | **L** | H | H |
| Sri Lanka[1] | **x** | **L** | | | **x** | **L** | L | South Asia |
| Suriname | **x** | **L** | | | | | UM | Latin America & the Caribbean |
| Swaziland | **x** | **L** | | | | | LM | Sub-Saharan Africa |
| Sweden | **x** | **L** | | | **x** | **L** | H | H |
| Switzerland | **x** | **L** | **x** | **L** | | | H | H |
| Syria | **x** | **L** | | | | | LM | Middle East & Northern Africa |
| Tanzania | **x** | **L** | | **L** | **x** | **L** | L | Sub-Saharan Africa |
| Thailand | | | **x** | **L** | **x** | **L** | LM | East Asia & the Pacific |
| Togo | **x** | **L** | | | **x** | **L** | L | Sub-Saharan Africa |
| Tonga | | | | | | **L** | LM | East Asia & the Pacific |
| Trinidad and Tobago | | | | | **x** | **L** | UM | Latin America & the Caribbean |
| Tunisia | | | | | **x** | **L** | LM | Middle East & Northern Africa |
| Turkey | | | **x** | | **x** | | LM | Europe & Central Asia |
| Uganda | | | **x** | **L** | | | L | Sub-Saharan Africa |
| United Kingdom[1] | **x** | **L** | **x** | **L** | **x** | **L** | H | H |
| United States of America[1] | **x** | **L** | **x** | **L** | **x** | **L** | H | H |
| United States Virgin Islands[1] | **x** | | **x** | | **x** | | H | H |
| Uruguay[1] | **x** | **L** | | **L** | **x** | **L** | UM | Latin America & the Caribbean |
| Venezuela[1] | **x** | **L** | | | **x** | **L** | UM | Latin America & the Caribbean |
| Viet Nam | | | | **L** | **x** | **L** | L | East Asia & the Pacific |
| Western Samoa | | | **x** | | | | * | * |
| Yemen | | | | | **x** | **L** | LM | Middle East & Northern Africa |
| Yugoslav SFR | **x** | | | | | | UM | Europe & Central Asia |
| Zaire | **x** | | | | | | * | * |
| Zambia | **x** | | **x** | | **x** | | L | Sub-Saharan Africa |

*Source:* Food and Agriculture Organization of the United States (1970, 1990, 2000), InSTePP Farm Size Data Series v 1.5 (2016), World Bank (2016)
*Notes:* To create traditional World Bank regions, the high-income class was used to replace the region and the 'Americas' classification was renamed 'Latin America & the Caribbean.'
[1]Denotes a country used in the consistent distribution sample, see footnote 13

* denotes no classification provided by World Bank, World Development Indicators (2016)

**L** denotes a Lowder et al. (means) observation

**x** denotes InSTePP distribution

## Appendix 2: Conditional Statistics Derivation

*Calculation of Gamma Conditional Average (Equation 16)*

$$\mu(x^l, x^u) = \int_{x^l}^{x^u} \frac{\frac{x}{\Gamma(k)\theta^k} x^{k-1} e^{-\frac{x}{\theta}}}{F(x^u|k,\theta) - F(x^l|k,\theta)} dx$$

$$= \frac{\Gamma(k+1)}{\Gamma(k)} \left( \frac{F(x^u|k+1,\theta) - F(x^l|k+1,\theta)}{F(x^u|k,\theta) - F(x^l|k,\theta)} \right) \theta \frac{\int_{x^l}^{x^u} \frac{1}{\Gamma(k+1)\theta^{k+1}} x^{k+1-1} e^{-\frac{x}{\theta}} dx}{F(x^u|k+1,\theta) - F(x^l|k+1,\theta)}$$

$$= \left( \frac{F(x^u|k+1,\theta) - F(x^l|k+1,\theta)}{F(x^u|k,\theta) - F(x^l|k,\theta)} \right) k\theta$$

*Calculation of Gamma Conditional Variance (Equation 17)*

$$\sigma^2(x^l, x^u) = \int_{x^l}^{x^u} \frac{\frac{x^2}{\Gamma(k)\theta^k} x^{k-1} e^{-\frac{x}{\theta}}}{F(x^u|k,\theta) - F(x^l|k,\theta)} dx$$

$$= \frac{F(x^u|k+2,\theta) - F(x^l|k+2,\theta)}{F(x^u|k,\theta) - F(x^l|k,\theta)} \frac{\Gamma(k+2)}{\Gamma(k)} \theta^2 \frac{\int_{x^l}^{x^u} \frac{1}{\Gamma(k+2)\theta^{k+2}} x^{k+2-1} e^{-\frac{x}{\theta}} dx}{F(x^u|k+2,\theta) - F(x^l|k+2,\theta)}$$

$$= \frac{F(x^u|k+2,\theta) - F(x^l|k+2,\theta)}{F(x^u|k,\theta) - F(x^l|k,\theta)} k(k+1)\theta^2$$

*Calculation of Beta Conditional Average (Equation 19)*

$$\mu(x^l, x^u) = \int_{x^l}^{x^u} x \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} dx$$

$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)} \frac{\Gamma(\alpha+1)}{\Gamma(\alpha+1+\beta)} \Big( F(x^u|\alpha+1,\beta) - F(x^l|\alpha+1,\beta) \Big) \frac{\int_{x^l}^{x^u} \frac{\Gamma(\alpha+1+\beta)}{\Gamma(\alpha+1)\Gamma(\beta)} x^{\alpha+1-1}(1-x)^{\beta-1} dx}{F(x^u|\alpha+1,\beta) - F(x^l|\alpha+1,\beta)}$$

$$= \frac{\alpha}{\alpha+\beta} \Big( F(x^u|\alpha+1,\beta) - F(x^l|\alpha+1,\beta) \Big)$$

*Calculation of Beta Conditional Variance (Equation 20)*

$$
\begin{aligned}
\sigma^2(x^l, x^u) &= \int_{x^l}^{x^u} x^2 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}\,dx \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)} \frac{\Gamma(\alpha+2)}{\Gamma(\alpha+2+\beta)} \Big( F(x^u|\alpha+2,\beta) \\
&\quad - F\big(x^l|\alpha+2,\beta\big)\Big) \frac{\int_{x^l}^{x^u} \frac{\Gamma(\alpha+2+\beta)}{\Gamma(\alpha+2)\Gamma(\beta)} x^{\alpha+2-1}(1-x)^{\beta-1}\,dx}{F(x^u|\alpha+2,\beta) - F\big(x^l|\alpha+2,\beta\big)} \\
&= \frac{(\alpha+1)\alpha}{(\alpha+1+\beta)(\alpha+\beta)} \Big( F(x^u|\alpha+2,\beta) - F\big(x^l|\alpha+2,\beta\big)\Big)
\end{aligned}
$$