Understanding heterogeneity in Peruvian agriculture: A meta-frontier approach for analyzing technical efficiency

M. Espinoza[1]; R. Fort[1]; M. Morris[2]; A. Sebastian[2]; L. Villazón[1]

1: Group for the Analysis of Development (GRADE), None, Peru, 2: World Bank, Agriculture and Rural Development, United States of America

Corresponding author email: mespinoza@grade.org.pe

Abstract:

This paper describes an empirical application of a stochastic method for estimating meta-frontier production functions described by Huang, Huang, and Liu (2014) that has been little applied in the context of heterogeneity of production technology. We use nationally and regionally representative data to explore patterns of productivity within and between regions of Peru and to understand factors that influence technical efficiency. We estimate a stochastic national meta-frontier that allows for a comparison between different systems in relation to the agricultural sector as a whole. Our results support the view that even though total factor productivity (TFP) and technical efficiency (TE) measured at national level are rising steadily, significant differences persist within the country. The extreme heterogeneity of Peru's agricultural sector is reflected in different realities between its regions and farming systems. Levels of productivity technical efficiency differ not only between regions, but also within regions, because farmers located within the same region are not equally productive or efficient.

Acknowledgement:

JEL Codes: O47, O12

#1079

Understanding heterogeneity in Peruvian agriculture:

A meta-frontier approach for analyzing technical efficiency

Abstract

This paper describes an empirical application of a stochastic method for estimating meta-frontier production functions described by Huang, Huang, and Liu (2014) that has been little applied in the context of heterogeneity of production technology. We use nationally and regionally representative data to explore patterns of productivity within and between regions of Peru and to understand factors that influence technical efficiency. We estimate a stochastic national meta-frontier that allows for a comparison between different systems in relation to the agricultural sector as a whole. Our results support the view that even though total factor productivity (TFP) and technical efficiency (TE) measured at national level are rising steadily, significant differences persist within the country. The extreme heterogeneity of Peru's agricultural sector is reflected in different realities between its regions and farming systems. Levels of productivity technical efficiency differ not only between regions, but also within regions, because farmers located within the same region are not equally productive or efficient.

1. Introduction

Empirical research on productivity growth and the role of technical efficiency in agriculture has continued to grow over the last few decades. A particular focus has been on understanding heterogeneity in the sources of growth and causes of inefficiency both between and within countries. Understanding productivity growth stemming particularly from technological change and its implications for shifting farmers closer to a production possibilities frontier is essential for boosting agricultural output to meet global demand for food amidst rapid population growth.

More recently, analysis of productivity and heterogeneity has focused on the relationship between farm size and productivity, specifically Total Factor Productivity (TFP)[1] (Key, 2017; Sheng and Chancellor, 2017; Kagin et al., 2016; Helfand and Taylor, 2017). This analysis has been facilitated by the development of new econometric techniques to analyze such relationships including stochastic frontier analysis (SFA) and data envelopment analysis (DEA) (see Coelli et al. 2005). Insights derived through the use of these analytical techniques could also provide countries with policy recommendations for boosting regional productivity conditional on differences in agricultural systems driven by differences in agro-climatic conditions. Such recommendations would provide useful insights into factors that could mitigate existing productive heterogeneity. A related underlying issue that may be of interest to policy makers, that has been little examined to date, is to examine the tension between promoting measures to boost productivity growth for farmers in a region and promoting measures to drive productivity growth at the national level. Empirically identifying the difference between the extent to which the former could move farmers from a region closer to the regional possibilities frontier relative to the extent to which the latter

---

[1] Significantly more studies assess the relationship between farm size and productivity using crop yields (land productivity) as a measure of productivity. In the past, many studies found an inverse relationship (for example, see Sen 1962, 1966; Carter, 1984; Eswaran and Kotwal, 1985, 1986; Barrett, 1996; Alvarez and Arias, 2004; Barrett et al., 2010; Carletto et al., 2013; Kagin et al., 2015). Van Zyl et al. (1995) and Eastwood et al. (2010) for a review of the theoretical and empirical evidence surrounding the farm size-productivity debate. Other recent work suggests that the relationship is a U-shaped (for example, see Foster and Rosenzweig (2017), Kevane, 1996; Zaibet and Dunn, 1998).

could push out the national productivity frontier could provide interesting insights into productive efficiency.

With the aim of shedding light on the scope of productive inequality in Peru, this paper estimates the degree of heterogeneity in crop production efficiency levels both within regions and between regions, and it explores the main factors that influence technical efficiency in different agricultural systems. Our empirical approximation follows a meta-frontier approach that allows the estimation of comparable technical efficiency (TE) measures for farms or firms operating in different productive systems with variation in terms of climate, geography, and soil type. Such an analysis has never been done for Peru, and the results provide information of potential interest to policy makers.

Our empirical analysis is one of the first applications of a new econometric stochastic method proposed by Huang, Huang, & Liu (2014) for estimating meta-frontier production functions in contexts of heterogeneity in production technology. To date only one known study has applied this method to the agricultural sector. Melo-Becerra & Orozco-Gallo (2017) use a sample of farmers located in different production systems in Colombia. Our study contributes to this literature by analyzing the Peruvian case and using a considerably larger data set that is regionally and nationally representative, which, unlike the Colombian study, allow us to map crop production efficiency levels at more disaggregated levels of analysis.

This paper also seeks to contribute to the narrow body of literature on efficiency in Peruvian agriculture. Some of the studies that have analyzed this issue focused on specific valleys and crops (De Los Ríos, 2006, Trivelli, Escobal and Revesz, 2006) that do not allow a comprehensive view of efficiency trends throughout the country. Although other studies such as Jacoby (1991), has a larger scope as it studies a sample of farmers from the Sierra region, the study focused only on partial measures of productivity such as labor productivity. Finally, a more recent study from Galarza and Díaz (2015) estimates Total Factor Productivity measures for a representative sample of agricultural producers from the National Survey of Strategic Programs 2011 and 2012, which allow them to assess the level of heterogeneity in agricultural productivity. Galarza and Diaz's paper, however, implicitly assumes a single production function for the whole sector, which does not result appropriate given the marked contextual and technological differences observed between the regions of the country. Furthermore, Galarza and Díaz study does not analyze the causes that explain the inefficiency use of resources, an aspect than could be very useful for evidence-based policy recommendations. Our study, thus, is the first that analyzes productive efficiency and it determinants across natural regions of Peru, while at the same assesses the heterogeneity in efficiency levels at the national level.

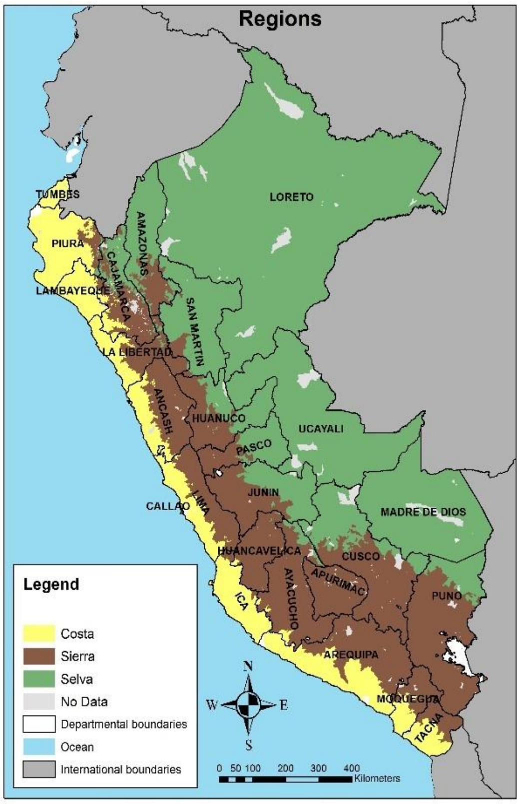
The structure of this paper is as follows. Section 2 gives some background on Peruvian agriculture in Costa, Sierra and Selva regions, and provides some initial indicators on productivity differentials across the three regions. Section 3 presents the methodological framework, which includes a brief explanation of key concepts such as efficiency and productivity, and details the Stochastic Meta-frontier approach followed in this study. Section 4 describes the data and provides some descriptive statistics for the sample. Section 5 presents the results and Section 6 concludes with a number of implications for agricultural and rural policy.

2. Agriculture in Peru

Peru's heterogeneous agriculture sector comprises a mix of subsistence-oriented, transitional, and commercial farms that vary in size, production technology, input use, and degree of market integration.[2]

[2] About 88% of all farms (2 million) are small farms (< 10 ha), many of them family operated and subsistence oriented, while about 1% of the rest are large (>100 ha) using primarily hired labor and modern inputs for mostly commercial farming. The remaining mid-sized farms use a mix of labor.

Figure 1. National Regions of Peru



Peru is endowed with a complex geography characterized by a range of agro-climatic conditions that support many different types of agriculture. The massive Andean cordillera divides the country into three natural regions: the western Costa (arid coastal plains representing 24 percent of the agricultural land and home to about 16 percent of the farming households), the central Sierra (highlands representing 46 percent of the agricultural land and home to about 64 percent of the farming households), and the eastern Selva (low-lying Amazon rainforest representing 30 percent of the agricultural land and home to about 20 percent of the farming households).

Agriculture in the Costa region is dynamic, highly productive and relatively well integrated into local and/or international markets. Many farmers cultivate non-traditional high-value export commodities, particularly vegetables and fruits; and most of them have access to irrigation and improved technologies.[3]

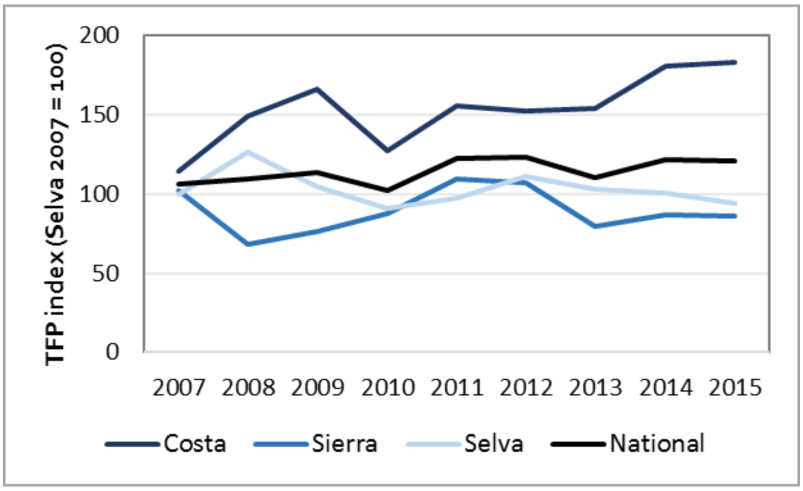
Agriculture in the Sierra region features static and generally unproductive agricultural systems that tend to be poorly integrated into the market, in which production of staples (potatoes, wheat, maize, quinoa, etc.) is combined with livestock keeping. Many farmers make limited use of improved technologies, including seed of modern crop varieties,

[3] According to the 2012 Agricultural Census, 90 percent of farmers in Costa have access to irrigated land, 41 percent uses improved seeds, more than 70 percent uses fertilizers, and 52 percent uses mechanization.

fertilizers, and irrigation; and mechanization is often infeasible on the steep mountain slopes.[4] Many farmers use family labor as a way to compensate for the low intensity of input use and technology adoption, achieving minimum levels of production that allow them to self-consume and survive.

Agriculture in the Selva region focuses to a large extent on production of tree crops including coffee, cocoa and palm oil, but also in more traditional crops such as banana, yucca, and rice. In recent years production growth has come mainly from expansion of the land frontier, rather than through increases in input use and technology improvements (World Bank, 2017).[5]

Figure 2. TFP growth index by natural regions, Peru, 2007-2015



Source: World Bank (2007)

The regional differences in agro-climatic conditions and farming systems are reflected in different levels of agricultural productivity. A recent report from the World Bank (2017) presents Total Factor Productivity (TFP) measures at national and sub-national levels calculated using data from the Peruvian National Household Survey (ENAHO) for the years 2007-2015. The measures support the widely held view that agricultural productivity has been growing rapidly in the Costa region while essentially stagnating in the Sierra and

---

4 About 60% of farmers do not have access to irrigation, about 20% uses mechanization, and less than 50 percent of farmers make use of fertilizers or insecticides (2012 Agricultural Census).

5 Because of the abundance of rainfall in the region, more than 90% percent of farmers do not have access to irrigation, but only 11 percent of farmers make use of improved seeds, 20 percent uses fertilizers, and 4 percent employ mechanization.

Selva regions. As a result, inter-regional productivity gaps have widened over time (Figure 2).

3. Empirical Methodology

3.1 Stochastic Frontier Analysis (SFA)

Performance in productive sectors is often evaluated using absolute measures based on output, such as the level of production or gross value added. While useful, such measures do not take into account all the inputs and resources utilized in the process of productive transformation and therefore do not consider the efficiency of economic activities.

A productive activity is technically inefficient whenever a higher level of output is attainable for the given inputs (output-oriented TE), or whenever the same level of output can be produced using fewer inputs (input-oriented TE). The choice between using output-oriented TE measures and input-oriented TE measures hinges on how producer behavior is conceptualized. Output oriented TE measures are appropriate when producers are attempting to maximize output with a given set of inputs, while input oriented TE measures are appropriate when producers are attempting to minimize inputs use. In agriculture, most papers analyzing technical efficiency use output-oriented TE measures, which is the approach we use here (Bravo-Ureta et al., 2007; Kumbhakar et al., 2015).

Output-oriented TE measures assume inputs are exogenous. This assumption is often reasonable for analysis of developing country agriculture, since most farmers in developing countries lack the resources to apply inputs at recommended levels (Crawford et al., 2003; Ayenew et al.2017). In the presence of resource constraints, they therefore seek to improve efficiency by maximizing output given a (limited) set of inputs, rather than seeking to minimize input use for a given level of output. In the Peruvian context, resource constraints are very common: because of the underdeveloped land market, imperfections in the financial system, and limited availability of credit, many producers have limited flexibility to adjust inputs.

Examples of studies that have analyzed technical efficiency in agriculture have assumed exogeneity in input choice include the meta-frontier studies conducted in Colombia (Melo-

Becerra and Orozco-Gallo, 2017), New Zealand (Jiang and Sharp, 2015), and Argentina, Chile and Uruguay (Moreira and Bravo-Ureta, 2010), as well as many studies in the standard stochastic frontier literature (for summaries, see Ayenew et al., 2017; Seymour, 2017). In addition, a number of authors have used instrumental variables approaches to test the exogeneity assumption; this work has consistently failed to detect evidence of significant bias due to endogeneity (Wouterse, 2016; Sonoda and Mishra, 2015).

Following Farrell (1957), TE can be defined as the actual productivity of a firm relative to maximum productivity attainable. The actual productivity of a firm can be measured as the ratio of the output(s) that it produces to the input(s) it uses. The maximum productivity attainable is revealed in the production frontier achieved by the best-performing firms. Measuring the efficiency of an individual firm therefore involves measuring the distance from the actual productivity of that firm to the frontier.

TE is typically assessed through the estimation of a production frontier function. The production frontier function, which is unobservable, is often viewed as stochastic, giving rise to the problem of how to estimate efficiency relative to the stochastic frontier when we can estimate only the “deterministic” component of the frontier—the component explained by inputs. Stochastic Frontier Analysis (SFA), developed by Aigner, Lovell, and Schmidt (1977) and Meeusen and van den Broeck (1977), is a standard approach that can be used to estimate the stochastic frontier and assess efficiency relative to the frontier for individual firms (Coelli et al. 2015).

The standard cross-sectional SFA model (Battese and Coelli, 1995) for production unit "i" in region "j" can be written as:

$$Y_{ji} = f^{j}(X_{ji})\, e^{V_{ji}-U_{ji}},$$

$$j=1,2, \ldots , \mathrm{J}\ ;\ i=1,2, \ldots , \mathrm{N_j}$$

where Y_{ji} is the output produced by production unit "i" in group or region "j"; X_{ji} is a vector of inputs; V_{ji} is assumed to be distributed as $i.i.d\ N(0, \sigma_V^2)$ and captures statistical noise (exogenous shocks beyond farmers’ control and measurement error) and U_{ji} is a nonnegative term measuring technical efficiency of farm "i" from region "j". Following

Caudill et al. (1995), $U_{ji}$is assumed to be distributed half-normal $N^{+}(0, \sigma^2_{u_{ji}})$, with a variance $\sigma^2_{u_{ji}} = exp(Z'_{ji}\delta_j)$, which is a function of exogenous determinants of technical efficiency, captured by a vector Z_{ji}, and a vector of unknown parameters δ_j.

Using on this model, it is possible to predict U_{ji} and calculate technical efficiency for production unit "i" in region "j" as $TE_i^j = exp(-U_{ji})$ (Jondrow et al., 1982).

3.2 Meta-Frontier Analysis (MFA)

A disadvantage of standard SFA models is that it is not possible to compare the TE of producers who use different technologies, because they do not operate under the same production frontier. This complication applies to the present study, because farmers in different regions of Peru have evolved distinct production technologies in response to widely divergent agro-climatic conditions. Spatial differences in temperature, rainfall, soil characteristics, and other production factors determine the input-output combinations that are possible for each region and, therefore, result in different production frontiers (O'Donnell et al. 2008). If TE is estimated independently for each for each region, the results cannot be compared directly because of the heterogeneity of the production technology between regions.

Recognizing this problem, Battese and Rao (2002) propose a meta-frontier approach for the calculation of TE measures that allow for comparisons among groups with different technologies. The measures they propose can be decomposed into group-specific TE measures and technology gap ratios. The approach is complemented by Battese et al. (2004) and O'Donnell et al. (2008), who use a two-step procedure for estimating the meta-frontier. In the first stage, standard SFA techniques are used to estimate the specific frontier of each group; in the second, Data Envelopment Analysis (DEA) is used to estimate the meta-frontier.

Since nonparametric methods as DEA lack of statistical inference and do not allow identification of the sources of variation among groups, we prefer a new approach proposed by Huang et al. (2014) which uses stochastic frontier techniques in the second stage, ensuring statistical properties of the stochastic frontier in the estimation.

Huang et al. (2014) depart from the estimation of group frontiers. The meta-frontier production function is defined as $f^M(X_{ji})$ and includes the group frontiers $f^j(X_{ji})$ estimated previously. It is expressed by the following relationship:

$$f^j(X_{ji};\beta_j) = f^M(X_{ji})e^{V_{ji}^M - U_{ji}^M}$$

The technology gap ratio, which represents the distance from the production frontier jth to the meta-frontier, is defined as follows:

$$TGR_i^j = e^{-U_{ji}^M}.$$

U_{ji}^M represents the technological gap ratio and is assumed to be distributed as half-normal $N^+(0,\sigma_{u_{ji}^M}^2)$ with variance $\sigma_{u_{ji}}^2 = exp(Z_{ji}^M \partial_j)$, which is a function of technology environment factors Z_{ji}^M.

The meta-frontier must be equal or greater than the group frontiers: $f^M(.) \geq f^j(.)$, which in turn implies that the technological gap ratio (TGR) must be less than or equal to 1.

The presence of V_{ji}^M in the meta-frontier equation is essential to ensure the stochastic nature of the metaproduction function. This stochastic element poses a problem in the meta-frontier estimation, however. Since $ln\hat{f}^j(X_{ji})$ is the maximum-likelihood estimator of the group frontiers, V_{ji}^M is asymptotically normally distributed with zero mean. Since it contains the residuals from the estimation of the group frontiers, however, it may not be independently and identically distributed (i.i.d.). Therefore, the stochastic frontier likelihood function associated with the meta-frontier equation (assuming i.i.d. in V_{ji}^M) is referred to as the quasi-likelihood function. However, the derived quasi-maximum likelihood estimator has invalid standard errors that have to be adjusted to account for the heteroscedasticity (Huang et al. 2014).

Measures of the efficiency of producers with respect to the meta-frontier are obtained by calculating the product of the estimated technological gap ratio (TGR) and the estimated individual farm's technical efficiency (TE). The resulting estimates of metatechnical efficiency (MTE) can be compared across firms and between regions.

$$\mathrm{MTE}_\mathrm{i}^\mathrm{j} = \mathrm{TGR}_\mathrm{i}^\mathrm{j}\ x\ \mathrm{TE}_\mathrm{i}^\mathrm{j}$$

3.3 Empirical Specification

Estimating the meta-frontier requires two stochastic frontier regressions. Following Huang et al. (2014), the equations to be estimated are:

$$ln\, Y_{ji} = ln\, f^{j}(X_{ji}) + V_{ji} - U_{ji} \qquad (1)$$

$$ln\hat{f}^{j}(X_{ji}) = lnf^{M}(X_{ji}) + V_{ji}^{M} - U_{ji}^{M} \qquad (2)$$

$$j=1,2, \dots, J\, ;\, i=1,2, \dots, N_j$$

Equation 1 represents the estimation of the group frontiers and it is estimated "j" times, one for each region. X_{ji} represents the vector of inputs (named above); and Y_{ji} define the output variable (the crop production value). Equation 2 represents the estimation of the meta-frontier, the estimated values $ln\hat{f}^{j}(X_{ji})$ obtained in the equation 1 are then pooled and used as dependent variable and the vector of inputs X_{ji} as independent variables. We use a flexible functional form for the production function, such as the translog, since it is usually applied in agricultural efficiency studies and can be easily adapted to any underlying production frontier (Bravo-Ureta et al., 2007; Seymour, 2017).

We adopt a translog production function for a single output and five inputs. Following Battese (1997), we include a dummy variable, which takes the value of 1 if farmer do not apply a determinate input in the production process, and 0 otherwise[6]. By using a dummy variable, we avoid potential biases in the parameter estimation due to the incidence of zero observations. We also include department fixed effect variables to account for any unobservable characteristics not in the model.

As mentioned previously, U_{ji} is assumed to be distributed half-normal $N^{+}(0, \sigma_{u_{ji}}^{2})$, and its variance $\sigma_{u_{ji}}^{2} = exp(Z_{ji}'\delta_j)$ is a function of exogenous determinants of technical efficiency captured by Z_{ji} vector and δ_j vector of unknown parameters to be estimated

[6] It is common in less developed countries to observe production processes that do not utilize all available inputs, resulting in zero values. As the estimation of production functions requires logarithmic transformations, such values present a problem if not properly dealt with.

In Equation 2, the meta-frontier $lnf^M(X_{ji})$ is estimated using the pooled estimates of the regional frontiers $ln\hat{f}^j(X_{ji})$. U_{ji}^M represents the technological gap ratio and its variance $\sigma_{u_{ji}}^2 = exp(Z_{ji}^M \partial_j)$. Since this second-step estimation is still based on a stochastic frontier regression, the technology gaps represented by the one-sided term can be further specified as a function of environmental variables Z_i^M beyond the control of firms and a vector of unknown parameters to be estimated (∂_j). Here we also use a translog functional form for the metaproduction function and include dummy variables for inputs not used in the production process.

4. Data

The data used for this analysis come from the 2015 National Agricultural Survey (Encuesta Nacional Agropecuaria – ENA), which was carried out by the National Institute of Statistics and Informatics (INEI). ENA, which uses the 2012 National Agricultural Census (CENAGRO) as sampling frame, will be repeated every year; for this reason, it is intended to become the official source for agricultural data that can be used to estimate the effects of public programs.

The 2015 ENA included about 28,000 observations. The sampling strategy for ENA is designed to ensure representativeness at national and departmental level. ENA recognizes two types of producers—small-scale and medium producers, and large producers—and uses a different (although quite similar) questionnaire for each type. We exclude large producers and producers who engaged exclusively in livestock production activities, i.e., we consider only small-scale and medium producers dedicated to crop production, including those who carry out livestock activities at the same time (89% of the sample). We also exclude producers who did not report the use of any input, as well as those who did not report quantity of production. Our final sample included 23,686 farms.

The ENA database includes detailed information about agricultural output. Quantities, prices, and value added by crop in each plot of the agricultural unit are recorded. With respect to inputs, ENA collects detailed information about quantities and expenditures by crop for a set of inputs, including seeds, organic manure, fertilizers and pesticides. ENA also collects information on farm-level aggregate expenditures for the following inputs: land rental, permanent and temporary workers, water, technical assistance, purchase and rental of agricultural equipment and machinery, fuel purchase, among others.

We used ENA to construct our input and output variables. Similar to most studies that consider multi-crop producers (Melo-Becerra & Orozco-Gallo, 2017; Wouterse, 2016; Jiang et al., 2015), we used as the dependent variable the value added of crop production (including byproducts and derivatives). We considered five inputs as explanatory variables, namely 'Land' (total area under cultivation in hectares); 'Labor' (number of days worked on the farm by hired workers and family members); 'Intermediate Inputs' (sum of the real cost of fertilizers, pesticides, seeds and manure); 'Capital' (the real cost of machinery and equipment rental and purchase); and 'Other Inputs' (including irrigation, technical assistance, and other input expenses). To the extent possible, we deflated the input expenses (for intermediate inputs, capital, and others) by using regional average prices to avoid endogeneity problems with input prices due to the presence of market power relations.

ENA collects information on how many members of the producer's family work in farm activities, but it does not collect information on hours or number of days worked. Since family labor is a critical input in Peruvian agriculture, a quantity measure was constructed using ENAHO 2015. This measure was then weighted according to the number of family members that reported working on the farm in ENA 2015. Following Coelli and Battese (1996), the number of days worked was converted to male equivalent days by discounting female days (conversion factor = 0.75) and child days (conversion factor = 0.5). The number of days worked by the family members was added to the number of days worked by the hired workers.

Table 1 shows the mean and standard deviations of the output and input variables used for the empirical analysis, disaggregated by region.

Table 1. Output and input use by regions

Variable	Description	Costa		Sierra		Selva	
		mean	sd.	mean	sd.	mean	sd.
		Output					
Production value	Crop production value (includes byproducts and derivatives)	27365	113003	4852	15368	14867	27935
		Inputs					
Intermediate inputs	Total expenditure in manure, fertilizers, pesticides, and seeds.	4973	14959	731	2573	831	3902
Labor	Total labor (in man days)	276	837	194	134	227	194
	Hired labor (in man days).	138	835	21	104	66	177
	Family labor (in man days)	138	81	173	84	161	77
Land	Cultivated Land (hectares).	2.56	5.57	0.93	2.42	4.71	17.7
Capital	Total expenditure on purchase, rental and maintenance of equipment and machinery	554	3166	29	397	183	1060
Other inputs	Total expenditure on land rental, technical assistance, irrigation and other inputs.	3293	12158	513	2949	316	2106

Source: Author's calculations

Sierra has, on average, the lowest production value per farm. In contrast, Costa has, on average, the highest production value per farm. Moreover, Costa farmers spend more on intermediate inputs and capital than Sierra and Selva farmers. Costa farmers use more hired labor than Sierra and Selva farmers, who on average use more family labor. Costa farmers use greater amounts of technological inputs than Sierra or Selva farmers, whose production relies heavily on the use of family labor. In addition to the marked inter-regional differences, the wide dispersion in output and most input variables suggests the presence of significant heterogeneity among farmers within each region.

We first perform stochastic frontier regressions by regions. Among the exogenous determinants of technical efficiency, we consider variables related to mobile connectivity (cell phone coverage in the district), electrification (share of households with electricity access in the district), irrigation (irrigation coverage in the district), and credit access (% of households in the district). The ENA data do not include information on farm micro-location, so the center of the conglomerate (a sub-district level of aggregation) where farms are located was used as a proxy of farm location. Using Open Street Map, we calculated the road distance and the road travel time from the conglomerate center to a city with population of 50,000 inhabitants or more.

We included also a dummy variable for land titling (which takes a value of 1 if farmer has a title and a value of 0 if the farmer does not have a title), a variable that accounts for nonagricultural income opportunities (share of farmers with non-agricultural income in the district), and a variable for crop diversification (average crop Herfindahl index of farms at the district level). The average number of plots in the district was included as a proxy for land fragmentation. Our variables at district level came mainly from 2012 National Agricultural Census-CENAGRO.

Other variables include the membership in an association, a dummy variable that indicates whether the farmer also engages in livestock production activities, access to extension and advisory services, and access to agricultural information. All of these variables were measured at the farm level. Finally, several variables were included relating to household socioeconomic characteristics (age, gender, education level of the household head and household size). Most of these variables come from the 2015 ENA. Table 2 presents information about the variables and their sources.

Table 2. Exogenous determinants of in(efficiency)

Variable	Description	Source	Costa mean	Costa sd.	Sierra mean	Sierra sd.	Selva mean	Selva sd.
Mobile connectivity	District cell phone coverage (%)	MINTRA 2012	92%	13%	65%	25%	57%	27%
Access to markets	Distance to nearest city with population above 50k	Open Street Map	47.6	44	144.5	83.0	148.8	102.5
Electrification	Share of households with electricity access in the district (%)	SISFOH 2012	46%	23%	35%	25%	17%	15%
Irrigation	Irrigation coverage in the district (%)	CENAGRO 2012	93%	14%	51%	39%	4%	13%
Credit access	Credit access in the district (% of HH)	CENAGRO 2012	24%	17%	6%	8%	10%	10%
Titling	Land titling dummy	ENA 2015	49%	50%	17%	38%	23%	42%
Nonagricultural income opportunities	Share of farmers with non-agricultural income in the district (%)	CENAGRO 2012	22%	13%	23%	18%	18%	18%
Crop diversification	Average crop Herfindahl Index of farms at the district level (0 to 1)	CENAGRO 2012	0.83	0.09	0.66	0.11	0.69	0.10
Land fragmentation	Average number of plots in the district	CENAGRO 2012	1.49	0.27	2.95	1.36	1.37	0.24
Association	If farmer belongs to a producer association (dummy, yes=1).	ENA 2015	9%	29%	4%	20%	14%	35%
Livestock	If farmer realize livestock activities (dummy, yes=1)	ENA 2015	51%	50%	85%	36%	78%	41%
Extension and advisory services	If farmer receives extension and advisory services (dummy, yes=1)	ENA 2015	12%	33%	5%	21%	15%	35%
Access to information	If farmer receives agricultural information (dummy, yes=1)	ENA 2015	43%	50%	40%	49%	60%	49%
Household size	Number of household members	ENA 2015	3,45	1.89	3.62	2.01	4.09	2.19
Gender	Female producer (dummy, yes=1)	ENA 2015	24%	43%	31%	46%	19%	39%
Age	Age of producer	ENA 2015	56.42	14.49	52.62	15.52	46.59	14.19
Education	% of farmers with secondary education or better in the district	CENAGRO 2012	37%	14%	22%	12%	21%	8%

Source: Author's calculations

As shown in Table 2, the Costa region is more favorably endowed than the Sierra and Selva regions with respect to infrastructure and connectivity. Districts in the Costa region have better mobile connectivity, electrification, irrigation coverage, and credit access. Moreover, farms in the Costa region are located closer to urban population centers, which suggests greater access to markets. In the Sierra region there is greater diversification of crops (lower Herfindahl concentration index) and more livestock activities. Participation in non-agricultural activities has similar importance across all regions.

In the Sierra region, landholdings are more fragmented than in the Selva and Costa regions. Additionally, Sierra producers show low rates of membership in producer associations, are less likely to have received extension and advisory services, and have less access to information. With respect to household socioeconomic characteristics variables, there is a greater presence of producer households with female household heads and a lower percentage of households with household heads with at least complete secondary education in the Sierra.

Following the estimation of regional frontiers, a production meta-frontier was estimated. This second stage estimation included as explanatory variables a number of exogenous factors that are hypothesized to contribute to technological gap ratios (i.e., the difference between a given technology with respect to the best regional technology). Huang et al. (2014) refer to these variables as environmental factors beyond the control of firms. Because the production possibilities in the agricultural sector of each region heavily depend on the endowment of natural resources and climate, we considered climate and soil variables as determinants of the technology gap-ratios. We included a district-level soil quality proxy variable that came from 2012 CENAGRO, as well as precipitation and temperature indicators at district level that were constructed from information from the Terrestrial Air Temperature: 1900-2014 Gridded Monthly Time Series (V 4.01 added 5/1/15) (Matsuura and Willmott, 2009). Finally, we included a variable of average altitude of the SEA (a sub-district level of aggregation in CENAGRO 2012 and ENA 2015). Table 3 presents descriptions of these variables.

Rainfall in the Costa region is clearly lower than in the other two regions, which points to a greater importance of irrigation for agriculture in this region. In the Sierra region, the greater temperature variability and the higher altitude at which many farms are located provide an indication of the strong influence of climatic and geographical conditions in determining the characteristics of agriculture, which in turn may play a role in influencing technical efficiency in this region is compared to the Costa and Selva regions.

Table 3. Determinants of technology-gap ratios

Variable	Description	Source	Costa		Sierra		Selva	
			mean	sd.	mean	sd.	mean	sd.
Soil quality	Share of non-cultivated agricultural land by soil quality factors (district level)	CENAGRO 2012	0.6%	1.5%	0.2%	1.2%	0.3%	0.9%
Rainfall	Log (Average Rainfall during 2010-2014) (district level)	TAT&P*	2.75	1.07	4.42	0.51	5.25	0.38
Temperature	Log (Average Temperature during 2010-2014) (district level)	TAT&P	2.94	0.26	2.27	0.64	3.22	0.16
Altitude	Average altitude in meters above the sea of centers from SEAs to which farms belong.	CENAGRO 2012	281	408	3225	636	681	568

Source: Author's calculations

* Terrestrial Air Temperature and Precipitation: 1900-2008 Gridded Monthly Time Series, Version 1.02 (Matsuura and Willmott (2009))

5. Results

5.4 Stage 1: Regional production frontiers

We began our investigation by evaluating the relevance of assuming that there are different frontiers for each region. If a common frontier exists (implying that the same technology is used in all three regions), it would suffice to estimate the common frontier, and there would be no need to go to the second stage of estimating a meta-frontier based on the different frontiers of each region.

We used a likelihood-ratio test to evaluate the null hypothesis that stochastic frontier models for the three regions are the same for all farms in Peru.[7] The statistical value of the LR test is 2256 and is highly significant, concluding that the stochastic frontier for the three regions is not the same and has to be independently estimated.

[7] The LR statistic is $\lambda = -2\{\ln[L(H_0)] - \ln[L(H_1)]\}$, where $\ln[L(H_0)]$ is the value of the log-likelihood function for the frontier estimated pooling data of farms from all regions and $\ln[L(H_1)]$ is the sum of the values of the log-likelihood functions for the three regional frontiers.

Table 4. Hypothesis testing for the adopted model by Regions

Model	Costa					Sierra					Selva				
	Ln L(H_0)	Λ	$\chi^2_{0.95}$	d.f.	Decision	Ln L(H_0)	λ	$\chi^2_{0.95}$	d.f.	Decision	Ln L(H_0)	Λ	$\chi^2_{0.95}$	d.f.	Decision
Translog Best Specification	-4,577					-12,984					-4,583				
No Inefficiency effects	-4,861	568	29.55	19	Rejected	-13,212	456	29.55	19	Rejected	-4,739	312	29.55	19	Rejected
Cobb-Douglas	-4,643	132	24.38	15	Rejected	-13,159	350	24.38	15	Rejected	-4,732	298	24.38	15	Rejected
No department fixed effects	-4,767	380	19.05	10	Rejected	-13,489	1010	28.27	18	Rejected	-4,659	152	19.05	11	Rejected

Source: Author's calculations

Stochastic frontiers specific to each region were estimated using the approach described by Battese and Coelli (1995) using a maximum-likelihood estimation (see Equation 1). Several likelihood-ratio tests were used to guide the choice of the most appropriate specification (see results in Table 4). The null hypothesis that the technical inefficiency effects were not present in a given region was rejected for all regions. Likewise, the null hypothesis that the Cobb-Douglas frontier is an adequate representation of the data was strongly rejected. In addition, the null hypothesis that department fixed effects were not present was rejected for all regions.

Table 5. Input Elasticities by Regions, Translog Production Function Model

Elasticities	Costa		Sierra		Selva	
	mean	s.d.	mean	s.d.	mean	s.d.
Intermediate Inputs	0.18	0.01	0.23	0.01	0.01	0.01
Labor	0.08	0.02	0.09	0.02	0.12	0.02
Land	0.66	0.02	0.68	0.01	0.69	0.01
Capital	0.00	0.05	-0.07	0.08	0.00	0.05
Other inputs	0.06	0.01	0.03	0.01	0.02	0.03
Returns to scale	0.98	0.06	0.95	0.08	0.84	0.06

Source: Author's calculations

Note: Elasticities standard errors are bootstrapped using 300 replications.

Under the Battese and Coelli approach, technical (in)efficiency and its exogenous determinants are estimated simultaneously through a stochastic production function estimation. Table 5 shows the input elasticity estimates by region (see results for input coefficient estimates of the Translog production function in Table A1).

Table 6. Stochastic Production Frontier (SPF) model results, Inefficiency coefficients

Inefficiency components	Costa		Sierra		Selva	
	Coef.	**S.E.**	**Coef.**	**S.E.**	**Coef.**	**S.E.**
Association dummy	-0.4964	(0.3028)	-0.1953	(0.1580)	-0.3614***	(0.0924)
Livestock dummy	-0.0122	(0.1289)	-0.1133	(0.0770)	0.1340*	(0.0698)
Cell phone coverage in the district	-0.0087	(0.4536)	0.0565	(0.1189)	0.0852	(0.1183)
Electric power coverage in the district	-0.5652*	(0.2914)	-0.7014***	(0.1301)	0.0304	(0.2180)
Distance to nearest city with population above 50k	0.0025	(0.0018)	0.0012***	(0.0004)	-0.0009***	(0.0003)
Extension and advisory services dummy	-0.5362*	(0.3221)	-0.8723***	(0.2330)	-0.4583***	(0.0929)
Access to information dummy	-0.7681***	(0.1661)	-0.6339***	(0.0665)	-0.0627	(0.0564)
Non-agricultural income opportunities in the district	1.5683***	(0.5309)	-0.3256**	(0.1604)	0.5008***	(0.1824)
Average number of plots in the district	1.1905***	(0.1969)	0.0749***	(0.0213)	-0.2371*	(0.1403)
Average Herfindahl index in the district	0.0950	(0.7854)	1.6386***	(0.2793)	-0.1956	(0.3150)
Irrigation coverage in the district	-1.3298***	(0.4078)	-0.6927***	(0.0907)	0.2324	(0.2534)
Credit access in the district	-2.5002***	(0.5350)	-0.9084**	(0.4431)	-1.5901***	(0.3471)
Titling dummy	-0.5982***	(0.1256)	-0.1141	(0.0758)	0.0312	(0.0707)
Number of household members	0.0698**	(0.0337)	0.0181	(0.0170)	-0.0059	(0.0141)
Female producer dummy	0.3733***	(0.1287)	0.2112***	(0.0559)	0.0741	(0.0693)
Age of producer	-0.0002	(0.0045)	0.0056***	(0.0018)	0.0026	(0.0020)
% farmers with at least secondary education in the dist.	-5.1254***	(1.3376)	-1.0106***	(0.3212)	-0.3633	(0.4570)
Constant	-0.2756	(0.7822)	-2.1863***	(0.3031)	-0.0043	(0.3305)
Department dummies	YES		YES		YES	
N	4731		14003		4952	
Chi^2	25042.2		53466.1		15956.4	
Log likelihood	-4,577		-12,984		-4,583	

Source: Author's calculations
∗∗∗99%, ∗∗95%, ∗90% confidence level

Table 6 shows results on the exogenous determinants of technical efficiency by region. Two factors are strongly and significantly associated with impacting levels of (in)efficiency in all regions. Access to technical assistance (extension and advisory services) and access to credit both have a significant impact on reducing inefficiency. Income diversification opportunities, measured as the share of farmers with non-agricultural income, significantly increases inefficiency in the Costa and Selva regions, while reducing inefficiency in the Sierra. Non-agricultural income opportunities could be contributing to increased productivity in the Sierra region by absorbing the oversupply of family labor engaged in agriculture, or by providing economic resources for input purchases or land-related investments, thus alleviating credit constrains. The finding that income diversification is associated with greater inefficiency in the Costa and Selva regions suggests a diversion effect, whereby rural households with non-agricultural income opportunities devote less attention to their farming activities.

In addition, other factors are associated with decreased inefficiency in each region. For example, in the Costa region access to market information, irrigation, electric power coverage and land titling reduce inefficiency. Moreover, the effect of titling is more pronounced in the Costa region compared to the other two regions, most likely because restrictions in credit and land markets are not as important as in the other regions, which allows titling to be more effective. A similar situation occurs with irrigation, whose effect on reducing inefficiency is greater, due to the desert climate and the scarcity of rainfall in the Costa region.

In the Sierra region, land fragmentation is associated with increased inefficiency. Electric power coverage, market integration (measured by distance to a large city with 50,000 or more inhabitants) and access to market information are associated with reduction in inefficiency in this region. Crop diversification (measured using a Herfindahl index), also appears to be important in reducing inefficiency in the Sierra region, a result that is not found in the other two. This supports the conventional wisdom that having a mixed crop portfolio provide a way for farmers in the Sierra region to hedge against risk.

In the Selva region, several factors in addition to the cross-cutting factors are associated with decreased inefficiency. Land fragmentation appears to be an important driver of efficiency, and may be related to the low average number of plots in the Selva region, and to the extensification strategy that predominates here. Membership in a cooperative or producer association is especially important in explaining reduced inefficiency, which points to the benefits derived from participating in the important commercial value chains found in this region, such as coffee, cocoa, or palm oil.

5.5 Stage 2: Meta-frontier production function

The meta-frontier production function $lnf^M(X_{ji})$ was estimated using the pooled estimates of the regional frontiers $ln\hat{f}^j(X_{ji})$ through a stochastic production function regression using the Battese and Coelli approach (1995) (Huang et al., 2014).

Similar to the first stage, several likelihood-ratio tests were used to guide the choice of the most appropriate specification for the meta-frontier (see Table A3 in the Appendix). The null hypothesis that technological gap-ratios effects were not present was rejected. Likewise, the null hypothesis that a Cobb-Douglas meta-frontier is an adequate representation of the data was strongly rejected.

The concept of the meta-frontier is based on the existence of a metaproduction function that envelops all regional production functions (Hayami and Ruttan, 1971). The distance from the regional frontiers (the technology gap ratio) to the meta-frontier is assumed to be a function of environmental variables, which we argue determine the production possibilities of firms in each region. We assume that although there is a common technological environment in Peruvian agriculture, there are also structural differences that depend on the natural characteristics of each region and that will not be covered even if the efficiency of all the firms in the regions is maximized to its higher level. This assumption is similar to the one used in the study that applies stochastic meta-frontier techniques to the agricultural sector of Colombia (Melo-Becerra and Orozco-Gallo, 2017). Here, we consider soil quality, rainfall, and temperature as environmental variables that determinates the technological environment in which agriculture operates in each region.

Table 7 shows the estimates of inputs elasticities of the meta-frontier (Estimated parameters of input coefficient for the metafrontier are presented in Table A3 in the Appendix), and the coefficients and standard errors of the estimated parameters of the environmental variables as determinants of technology gap ratios.

The results show that higher rainfall levels are associated with larger technology gaps between the regional frontiers and the meta-frontier. Larger technology gaps also are associated with lower temperatures and higher altitude. These findings suggest that climatic conditions in the environments in which agriculture is carried out influence the technological possibilities and frontiers of each region.

Table 7. Meta-frontier input elasticities and determinants of technology gap ratios

Parameters	**mean/coef.**	**s.e.**
Elasticities		
Intermediate Inputs	0.14	0.005
Labor	0.09	0.004
Land	0.73	0.003
Capital	0.06	0.014
Other inputs	0.03	0.002
Returns to scale	1.04	0.014
Environmental Variables		
Soil quality	-0.9852	(0.7461)
Rainfall	0.2349***	(0.0632)
Temperature	-0.1585***	(0.041)
Altitude	0.0007***	(0.000)
Constant	-4.8809***	(0.4640)
Log-Likelihood	-4,380	
Observations	23686	

Source: Author's calculations

***99%, **95%, *90% confidence level.

Notes: Elasticities std. errors are bootstrapped using 300 replications.

Table 8 presents the technical efficiency measures (TE) derived from the frontiers of each region, the technology gap ratios (TGR) corresponding to the distance from the regional frontiers to the meta-frontier, and the meta-frontier's technical efficiency (MTE), which measures the distance from farms to the meta-frontier.

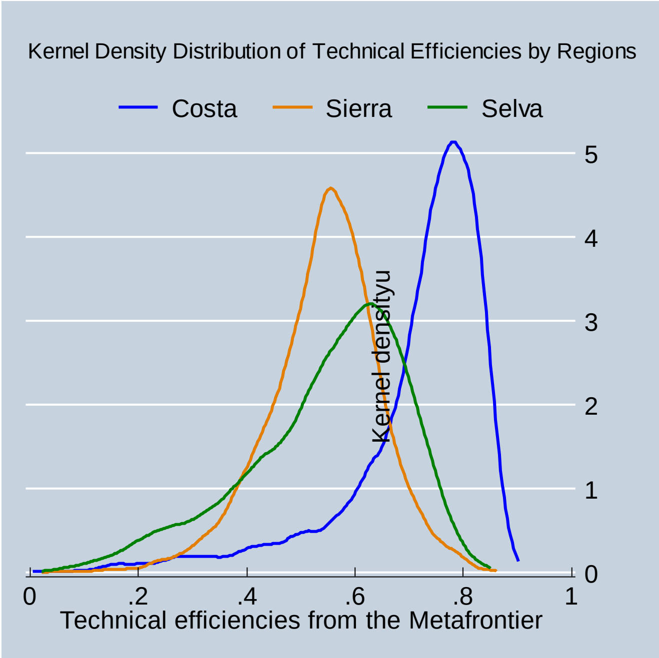
Table 8. Technical efficiency and meta-efficiency by regions

	Technology gap ratio (TGR)		Technical efficiency by region (TE)		Technical efficiency from the Meta-frontier (MTE)	
	Coef.	**S.E.**	**Coef.**	**S.E.**	**Coef.**	**S.E.**
By Regions						
Costa	0.93	0.01	0.77	0.15	0.71	0.14
Sierra	0.75	0.07	0.72	0.10	0.54	0.10
Selva	0.90	0.02	0.61	0.16	0.55	0.15

Source: Author's calculations

As mentioned previously, MTE estimates are comparable, since they measure the distance from the farm to the common meta-frontier. On average, farms in the Costa region are located much closer to the meta-frontier than farms in the Selva and Sierra regions. An interesting finding is that although the technological environment of the Sierra region is more underdeveloped than that of the Selva region, per the TGR, farms in the Selva region are located far below their regional frontier due to a larger distribution of technical efficiencies, determining at the end similar levels of meta technical efficiency in the farms of both regions. This can be better seen in the following graph, which shows the distribution of the farms of the three regions according to their efficiency levels in relation to the meta-frontier (MTE).

Figure 3. Kernel density distribution of technical efficiency, by region

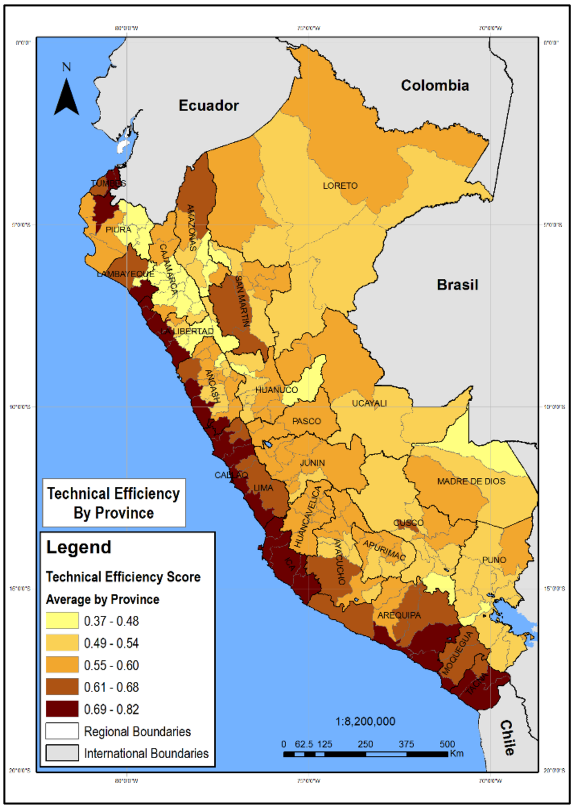


Source: Author's calculations

Because they are comparable, the meta-efficiency measures allow us to discern spatial variability in productive efficiency (Figure 4). The results are very consistent with what was expected. Generally speaking, farmers located in most of the Costa provinces are more efficient than those located in most of the Sierra and Selva provinces. Exceptions to this general patter involve locations outside the Costa region that are known for their dynamic and highly productive agricultural systems. Coffee and cocoa growers in some parts of San Martin, the famous fertility of the Urubamba valley in Cusco, the modern agriculture of Arequipa, and the new areas with olives and grapes in Moquegua are some examples of provinces outside the Costa region where efficiency levels are among the highest in the country.

In line with recent studies analyzing the Peruvian agriculture (World Bank, 2007), the most efficient producers in the Costa region are those cultivating high-value export crops like asparagus, mangoes, organic bananas, or avocados, and also others growing crops related to local agro industries like rice, sugar cane and vid. In the Sierra, the most efficient farmers cultivate avocado and alfalfa, mainly in the southern part; while in Selva, coffee, cocoa, and banana producers tend to be the most efficient in the region. In contrast, farmers from provinces of the north Sierra, belonging to the departments of Piura, Cajamarca, La Libertad, and some provinces from Amazonas and Huanuco departments show the lowest levels of efficiency at the national level. In this case, these are farms that perform subsistence agriculture and self-consumption, usually based in potatoes, corn and wheat crops (World Bank, 2007).

Figure 4. Technical efficiency by province, Peru



Source: Authors' calculation

6. Final thoughts

Our results support the widely held view that even though TFP and TE measured at national level are rising steadily, significant differences persist within the country. The extreme heterogeneity of Peru's agricultural sector is reflected in different realities between its regions and farming systems, and characteristics vary by natural region. Levels of productivity technical efficiency differ not only between regions, but also within regions, because farmers located within the same region are unlikely to be equally productive or efficient.

Our results are consistent with those of many studies that have documented considerable heterogeneity among farms in terms of scale, technology adoption, and efficiency in the use of resources (Trivelli, Escobal and Revesz, 2006). Many farms in Peru make limited use of improved technologies, including seed of modern crop varieties, fertilizers, machinery, and irrigation; which in turn results in low yields and profits. The heterogenous productive performance of farmers within and between regions appears to be driven by observed disparities in terms of pre-existing endowments, the presence of public infrastructure and services, imperfection in relevant markets (i.e., land and credit), and access to information and technology. In sum, the positive overall performance of the sector hides important inefficiencies that undermine the well-being of many segments of rural producers.

Our results of the meta-frontier estimation give a more accurate vision of these differences. Farms in the Costa region are on average located much closer to the meta-frontier than farms in the Selva and Sierra regions. But, although the technological environment of the Sierra region is more underdeveloped than that of the Selva, farms here are located far below their regional frontier due to a larger distribution of technical efficiencies, determining at the end similar levels of meta technical efficiency in the farms of both regions. On a more disaggregated look, we find some exceptions to this general pattern involving locations outside the Costa region that are known for their dynamic and highly productive agricultural systems.

7. References

Aigner, D., Lovell, C. K., & Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of econometrics*, *6*(1), 21-37.

Alvarez, A., & Arias, C. (2004). Technical efficiency and farm size: a conditional analysis. *Agricultural Economics*, *30*(3), 241-250.

Ayenew, H. Y., Estruch, E., Sauer, J., Abate-Kassa, G., Schickramm, L., & Wobst, P. (2017). Decent rural employment and farm production efficiency: empirical evidence from Tanzania and Ethiopia. *Agricultural Economics*, *48*(5), 587-596.

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*(1-3), 250-252.

Battese, G. E., & Coelli, T. J. (1995). A model for technical inefficiency effects in a stochastic frontier production function for panel data. *Empirical economics*, *20*(2), 325-332.

Battese, G. E., & Rao, D. P. (2002). Technology gap, efficiency, and a stochastic metafrontier function. *International Journal of Business and Economics*, *1*(2), 87.

Battese, G. E., Rao, D. P., & O'donnell, C. J. (2004). A metafrontier production function for estimation of technical efficiencies and technology gaps for firms operating under different technologies. *Journal of Productivity Analysis*, *21*(1), 91-103.

Bravo-Ureta, B. E., Solís, D., López, V. H. M., Maripani, J. F., Thiam, A., & Rivas, T. (2007). Technical efficiency in farming: a meta-regression analysis. *Journal of productivity Analysis*, *27*(1), 57-72.

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*, 254-261.

Carter, M. R. (1984). Identification of the inverse relationship between farm size and productivity: an empirical analysis of peasant agricultural production. *Oxford Economic Papers*, *36*(1), 131-145.

Caudill, S. B., Ford, J. M., & Gropper, D. M. (1995). Frontier estimation and firm-specific inefficiency measures in the presence of heteroscedasticity. *Journal of Business & Economic Statistics*, *13*(1), 105-111.

Coelli, T. J., & Battese, G. E. (1996). Identification of factors which influence the technical inefficiency of Indian farmers. *Australian journal of agricultural and resource economics*, *40*(2), 103-128.

Coelli, T. J., Rao, D. S. P., O'Donnell, C. J., & Battese, G. E. (2005). *An introduction to efficiency and productivity analysis*. Springer Science & Business Media.

Crawford, E., Kelly, V., Jayne, T. S., & Howard, J. (2003). Input use and market development in Sub-Saharan Africa: an overview. *Food Policy*, *28*(4), 277-292.

De los Ríos, C. (2006). Eficiencia técnica en la agricultura peruana (caso del algodón Tanguis en los valles de Huaral, Cañete y Chincha). *Debate Agrario,* 40-41, 141-168.

Eastwood, R., Lipton, M., & Newell, A. (2010). Farm size. *Handbook of agricultural economics*, *4*, 3323-3397.

Eswaran, M., & Kotwal, A. (1985). A theory of contractual structure in agriculture. *The American Economic Review*, *75*(3), 352-367.

Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society. Series A (General)*, *120*(3), 253-290.

Foster, A. & Rosenzweig, M. (2017). Input Transaction Costs, Mechanization and the Misallocation of Land: The Irrelevance of the the IR. In *USDA Farm Size and Productivity Conference*.

Galarza, F. B., & Díaz, J. G. (2015). \Total Factor Productivity in the Peruvian Agriculture: Estimation and Determinants. *Economía*, *38*(76), 77.

Hayami, Y., & Ruttan, V. W. (1971). *Agricultural development: an international perspective*. Baltimore, Md/London: The Johns Hopkins Press.

Helfand, S. M. ., & Taylor, M. P. . H. . (2017). The Inverse Relationship between Farm Size and Productivity: Refocusing the Debate. Paper to be presented at the 2017 Pacific Conference for Development Economics, Riverside, California. March 11, 2017.

Huang, C. J., Huang, T. H., & Liu, N. H. (2014). A new approach to estimating the metafrontier production function based on a stochastic frontier framework. *Journal of productivity Analysis*, *42*(3), 241-254.

Jacoby, H. G. (1991). Productivity of men and women and the sexual division of labor in peasant agriculture of the Peruvian Sierra. *Journal of Development Economics*, *37*(1-2), 265-287.

Jiang, N., & Sharp, B. (2015). Technical efficiency and technological gap of New Zealand dairy farms: a stochastic meta-frontier model. *Journal of Productivity Analysis*, *44*(1), 39-49.

Jondrow, J., Lovell, C. K., Materov, I. S., & Schmidt, P. (1982). On the estimation of technical inefficiency in the stochastic frontier production function model. *Journal of econometrics*, *19*(2-3), 233-238.

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.

Kevane, M. (1996). Agrarian structure and agricultural practice: Typology and application to Western Sudan. *American Journal of Agricultural Economics*, *78*(1), 236-245.

Key, N. (2017, February). Farm size and Productivity Growth in the United States Corn Belt. In *USDA Farm Size and Productivity Conference*.

Kumbhakar, S. C., Wang, H. J., & Horncastle, A. P. (2015). *A practitioner's guide to stochastic frontier analysis using Stata*. Cambridge University Press.

Matsuura, K., & Willmott, C. J. (2009). Terrestrial precipitation: 1900–2008 gridded monthly time series. *Center for Climatic Research Department of Geography Center for Climatic Research, University of Delaware*.

Meeusen, W., & van Den Broeck, J. (1977). Efficiency estimation from Cobb-Douglas production functions with composed error. *International economic review*, 435-444.

Melo-Becerra, L. A., & Orozco-Gallo, A. J. (2017). Technical efficiency for Colombian small crop and livestock farmers: A stochastic metafrontier approach for different production systems. *Journal of Productivity Analysis*, *47*(1), 1-16.

Moreira, V. H., & Bravo-Ureta, B. E. (2010). Technical efficiency and metatechnology ratios for dairy farms in three southern cone countries: a stochastic meta-frontier model. *Journal of Productivity Analysis*, *33*(1), 33-45.

O'Donnell, C. J., Rao, D. P., & Battese, G. E. (2008). Metafrontier frameworks for the study of firm-level efficiencies and technology ratios. *Empirical economics*, *34*(2), 231-25

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

Sen, A. K. (1966). Peasants and Dualism with or without Surplus Labor. *Journal of political Economy*, *74*(5), 425-450.

Seymour, G. (2017). Women's empowerment in agriculture: Implications for technical efficiency in rural Bangladesh. *Agricultural Economics*.

Sheng, Y., & Chancellor, W. (2017). Exploring the relationship between farm size and productivity—Evidence from the Australian grain industry. In *USDA Farm Size and Productivity Conference*.

Sonoda, T., & Mishra, A. (2015). Production Function Estimation Using Cross Sectional Data: A Partial Identification Approach. In *2015 AAEA & WAEA Joint Annual Meeting, July 26-28, San Francisco, California* (No. 205308). Agricultural and Applied Economics Association & Western Agricultural Economics Association.

Trivelli, C., Escobal, J., & Revesz, B. (2006). *Small Commercial Agriculture: dynamics and challenges in Peru*. Grupo de Análisis para el Desarrollo (GRADE) & Consorcio de Investigación Económica y Social (CIES).

Van Zyl, J., Binswanger, H. P., & Thirtle, C. G. (1995). *The relationship between farm size and efficiency in South African agriculture* (No. 1548). World Bank Publications.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica: Journal of the Econometric Society*, 1-25.

World Bank (2017). *Gaining momentum in Peruvian agriculture: opportunities to increase productivity and enhance competitiveness.* Washington, D.C.: World Bank Group.

Wouterse, F. (2016). Can human capital variables be technology changing? An empirical test for rural households in Burkina Faso. *Journal of Productivity Analysis*, *45*(2), 157-172.

Zaibet, L. T., & Dunn, E. G. (1998). Land tenure, farm size, and rural market participation in developing countries: the case of the Tunisian olive sector. *Economic Development and Cultural Change*, *46*(4), 831-848.

8. Appendix

Table A1: Estimated parameters for the Regional frontiers

Variables	Costa		Sierra		Selva	
	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
$\ln X_1$	-0.0829	(0.0619)	-0.0175	(0.0422)	-0.1500***	(0.0416)
$\ln X_2$	-0.1874	(0.1548)	-0.2442	(0.1711)	-0.3288	(0.2163)
$\ln X_3$	0.8084***	(0.0873)	0.6084***	(0.0511)	0.5200***	(0.0718)
$\ln X_4$	-0.0186	(0.0721)	0.0085	(0.0719)	0.0303	(0.0758)
$\ln X_5$	0.1028*	(0.0541)	0.0133	(0.0321)	0.0885*	(0.0517)
$0.5\ \ln X_1^2$	0.0364***	(0.0071)	0.0641***	(0.0048)	0.0485***	(0.0057)
$0.5 \ln X_2^2$	0.0437	(0.0289)	0.1021***	(0.0328)	0.0933**	(0.0426)
$0.5 \ln X_3^2$	-0.0167*	(0.0091)	-0.0113**	(0.0048)	-0.0613***	(0.0063)
$0.5 \ln X_4^2$	0.0052	(0.0111)	0.0179	(0.0132)	0.0104	(0.0124)
$0.5 \ln X_5^2$	0.0027	(0.0063)	0.0362***	(0.0038)	0.0284***	(0.0085)
$0.5 \ln X_1 \ln X_2$	0.0077	(0.0110)	-0.0138*	(0.0073)	-0.0037	(0.0065)
$0.5 \ln X_1 \ln X_3$	-0.0276***	(0.0049)	-0.0134***	(0.0028)	-0.0115***	(0.0032)
$0.5 \ln X_1 \ln X_4$	0.0033	(0.0027)	-0.0066**	(0.0033)	-0.0009	(0.0015)
$0.5 \ln X_1 \ln X_5$	-0.0043	(0.0029)	-0.0071***	(0.0015)	-0.0017	(0.0014)
$0.5 \ln X_2 \ln X_3$	-0.0042	(0.0166)	0.0302***	(0.0092)	0.0528***	(0.0139)
$0.5 \ln X_2 \ln X_4$	0.0035	(0.0052)	-0.0067	(0.0079)	-0.006	(0.0068)
$0.5 \ln X_2 \ln X_5$	-0.0026	(0.0089)	-0.0209***	(0.0053)	-0.0219***	(0.0063)
$0.5 \ln X_3 \ln X_4$	0.0069*	(0.0038)	0.0048	(0.0040)	-0.0100***	(0.0033)
$0.5 \ln X_3 \ln X_5$	0.0067	(0.0042)	-0.0065***	(0.0019)	-0.0048	(0.0031)
$0.5 \ln X_4 \ln X_5$	-0.0060***	(0.0018)	-0.0015	(0.0017)	-0.0050***	(0.0014)
d1	0.4476***	(0.1187)	0.2479***	(0.0678)	-0.1923**	(0.0766)
d4	-0.0101	(0.2211)	-0.2597*	(0.1499)	-0.0874	(0.1978)
d5	0.1129	(0.1285)	-0.2207***	(0.0495)	0.0857	(0.1157)
Department Fixed Effect	YES		YES		YES	
Observations	4731		14003		4952	

Source: Author's calculations

∗∗∗99%, ∗∗95%, ∗90% confidence level.

Table A2: Estimated parameters for the metafrontier

Parameters	Coef.	S.E.
$\ln X_1$	-0.1476***	(0.0143)
$\ln X_2$	-0.1890***	(0.0431)
$\ln X_3$	0.7778***	(0.0147)
$\ln X_4$	0.0721***	(0.0183)
$\ln X_5$	0.0385***	(0.0098)
$0.5 \ln X_1^2$	0.0558***	(0.0017)
$0.5 \ln X_2^2$	0.0676***	(0.0083)
$0.5 \ln X_3^2$	-0.0232***	(0.0020)
$0.5 \ln X_4^2$	-0.0057**	(0.0027)
$0.5 \ln X_5^2$	0.0241***	(0.0014)
$0.5 \ln X_1 \ln X_2$	-0.0002	(0.0019)
$0.5 \ln X_1 \ln X_3$	-0.0172***	(0.0008)
$0.5 \ln X_1 \ln X_4$	-0.0031***	(0.0005)
$0.5 \ln X_1 \ln X_5$	-0.0042***	(0.0004)
$0.5 \ln X_2 \ln X_3$	0.0114***	(0.0027)
$0.5 \ln X_2 \ln X_4$	0.0035**	(0.0017)
$0.5 \ln X_2 \ln X_5$	-0.0169***	(0.0015)
$0.5 \ln X_3 \ln X_4$	-0.0018*	(0.001)
$0.5 \ln X_3 \ln X_5$	-0.0073***	(0.0007)
$0.5 \ln X_4 \ln X_5$	-0.0020***	(0.0003)
d1	0.1214***	(0.039)
d4	0.068	(0.0456)
d5	-0.0476**	(0.0199)
Constant	8.8777***	(0.1351)
Environmental Variables		
Soil quality	-0.9852	(0.7461)
Rainfall	0.2349***	(0.0632)
Temperature	-0.1585***	(0.041)
Altitude	0.0007***	(0.000)
Constant	-4.8809***	(0.4640)
Log-Likelihood	-4,380	
Observations	23686	

Source: Author's calculations

$***$99%, $**$95%, $*$90% confidence level.

Table A3. Hypothesis testing for the adopted metafrontier model

Model	All Sector				
	Ln L(H_0)	**λ**	**$\chi^2_{0.95}$**	**d.f.**	**Decision**
Translog Best Specification	-4357				
No Inefficiency effects	-5609	2505	29.55	19	Rejected
Cobb-Douglas	-5987	3260	24.38	15	Rejected

Source: Author's calculations