Agosto 2022

WP3-2022-004

N° de serie

DOCUMENTO DE TRABAJO

Cross-sectional productivity estimation:

A practical guide

Autores

Nicolás Garcés

Henry Gómez

COLOMBIA CIENTÍFICA
Conocimiento Global para el Desarrollo

ALIANZA EFI
Economía Formal e Inclusiva

Cross-sectional productivity estimation: A practical guide

Nicolás Garcés Rodríguez*
Henry Gómez Ramírez*

August 18, 2022

Abstract

Productivity is a key factor for the growth of production outcomes and the development of industries. While data and methodologies are demanding, many insightful conclusions can be derived from cross sectional estimations, especially with micro data. This paper does a literature review of the most important methodologies for the estimation of cross-sectional productivity and the proper treatment of granular data in such estimations. Taking advantage of a rich dataset of agricultural production at farm level, the paper shows a practical example of the methodologies exposed and recommended data treatment. Although the results shows common directions on the productivity estimations, we conclude that the assumptions underlying each methodology will determine the context in which the estimations would have a better fit.

JEL clasification: O47, D24, Q12.
Keywords: Productivity, Cross-sectional data, Production unit level analysis.

1 Introduction

Total factor productivity (TFP) is understood as the non input related part of the output in a production process. Since Solow (1957) classic paper, productivity has been considered a main determinant of growth because it captures the technology, ways in which some resources are transformed into a new product, and efficiency, ways in which this process develops, involved in an economic activity. When fixing input levels, any change in the output can only be explained by differences in methods, knowledge and processes used in the transformation process, making the TFP a residual measurement that can be effectively compared between production units.

Although we could think of many determinants of this factor, TFP is the sum of widely different elements, making it difficult to conceptualize or to put in a comprehensive scale. Furthermore, to properly identify it, one must correctly characterize the relationship between the output and the inputs. The econometric and conceptual requirements of this estimation are thus heavily reliant on the data available to approach this problem.

In this paper, we discuss the methodologies that can be implemented to estimate total factor productivity in settings where only cross-sectional data is available. This methodologies are then applied to a rich dataset of 654 coffee producers households in Colombia, that includes information on total agricultural output, type of labor and capital used in production, land usage and a wide variety of farm and household characteristics. A main advantage of this dataset is that the level of disaggregation allows to have a detailed definition of factors relations. Yet, this tools can also be applied to aggregated data, giving this methodology review a wide scope.

The paper is structured as follows. First, the methodologies for TFP estimation are described, starting with frontier estimations, followed by indexes and concluding with production function estimations.

*Universidad del Rosario

Following the methodology part, common data treatment practices and potential methodological problems will be described. Finally, a practical application of each methodology is estimated and discussed.

2 Frontier Estimations

The notion of productivity as the value of output relative to a potential production comes as far as Farrell(1957), under the idea of the existence of an optimal upper bound level of production. This production limit, or frontier, can be used as benchmark to compare the performance of different production units if it is assumed that technology and methods are homogeneous for all the producers. This implies that in frontier methodologies, differences in productivity arise from different practices and ways of using the existing technology (technical efficiency), which are expressed as distances from the frontier.

2.1 DEA

Data envelopment analysis (DEA) is a frontier estimation technique based on the assumption of a local linear production frontier. For a vector n in the neighborhood of the input set X_n, the production frontier takes the form

$$\mu' q_n = \alpha + \upsilon' x_n \tag{1}$$

Where q_n is output, x_n is input, μ' and υ' are non negative values and α captures the returns to scale of the production (O'Donell, 2010). The DEA analysis can be oriented to both input or output, depending where the frontier is projected, which gives flexibility to analyze cost minimization or profit maximization production models. Furthermore, productivity is calculated by comparing radial expansions of factors relative to the frontier, the biggest radial expansion possible.

Based on the frontier definition, radial expansions can be measured by distance functions that take the following form when input oriented

$$D_I^t(x_n, q_n) = \frac{\upsilon' x_n}{\mu' q_n - \alpha} \tag{2}$$

and the following when output oriented

$$D_O^t(x_n, q_n) = \frac{\eta' q_n}{\Phi' x_n + \beta} \tag{3}$$

Where η, Φ and β are the analog of υ, μ and α respectively for the output oriented frontier, and are represented this way to state the difference of parameters depending on the orientation. Linear optimization is used to calculate unknown parameters (α, μ, υ or η, Φ,β) by solving a linear problem (maximization of $\mu' q_n - \alpha$ when input oriented or minimization of $\Phi' x_n + \beta$ when output oriented) subject to $D_I^t(x_n, q_n) > 1$ when input oriented and $D_O^t(x_n, q_n) < 1$ when output oriented (O'Donell, 2010). To simplify the optimization process and identify local solutions, is common the normalization of $\upsilon' x_n = 1$(conversely $\eta' q_n$=1 when output oriented), or the assumption of a theoretical value to α (like $\alpha = 0$ to exhibit constant returns to scale, for example)(O'Donell, 2010).

While not assuming a functional form and not needing prices, as q_i and x_i are in quantities, are big advantages, the methodology has some limitations. DEA is sensitive to parameters estimation, that will oscillate depending on the distribution of inefficiencies in the sample (Simar,1999). In addition, when q_i and/or x_i are matrix (multiple inputs and/or outputs), DEA estimates weights that can distort distance function if the overall composition of inputs or outputs in the data are not representative to the real distribution (Şişman,2017). One way of dealing with the latter problem is through Bootstrapping. This method allows to do estimations through sampling with replacements from empirical distribution of weights, allowing a more robust distance function estimation (Tziogkidis, 2012). Although Bootstrapping might help with weight bias, DEA cannot differentiate inefficiencies from data noise or imperfections (Hu, 2020), making it an important limitation of this methodology.

2.2 Stochastic frontier analysis (SFA)

Stochastic frontier analysis, originally proposed in Aigner, Lovell, and Schmidt (1977), builds over the same frontier notion as DEA but adds an extra element: stochastic shocks. The idea is that, for a given frontier, productions units suffers from inefficiencies that reduces output from it's potential, but there are also many unexpected and transitory factors that, while not being related to efficiency, can affect the outcome of the production process. The SFA model can be stated as follows

$$y_i = m(x_i; \beta) - u_i + v_i \tag{4}$$

Where $m(x_i; \beta)$ is the frontier of firm i, y_i is the output in log terms, u_i is the inefficiency term and v_i is the stochastic shock (Kumbhakar, 2020). Defining the frontier as $m(x_i; \beta)$ implies that this estimation gives a parametric functional form to the relation between output and inputs, which usually take the form of log linear Cobb-Douglas or Translog production function. This formulation clearly states that SFA is a generalization of functional forms, as they all include an error term, v_i in this case, but suppose that $u_i = 0$, leaving to the researcher the possibility of testing the assumption of full efficiency, present in functional form, with SFA (Kumbhakar, 2020), this by simply checking if u_i is statistically different from zero.

Consequently, most of the assumptions of functional forms are extended to the inefficiency term, that is, the independence of u_i and v_i and the orthogonality of v_i and u_i with production inputs. However, as u_i is a negative term for production, $E[v_i - u_i|x] \neq 0$ (Kumbhakar, 2020). OLS could be used to estimate SFA if equation (4) is rewritten by adding and subtracting expected inefficiency, changing the intercept of $m(x_i; \beta)$ and implying that $E[v_i - (u_i - E[u])|x] = 0$, but it would only be useful to capture mean inefficiency adjusted technology, not a central issue in productivity analysis (Kumbhakar, 2020). Therefore, Maximum Likelihood estimation is an useful alternative to estimate SFA, requiring to impose distributions to v_i, normal distribution with mean 0 and variance σ_v^2, and u_i, usually assumed a half normal distribution $N_+(0, \sigma_u^2)$ or exponential with parameter σ_u, and allowing to estimate both inefficiency and factor elasticity (Kumbhakar, 2020).

To estimate productivity in SFA, unknown parameters (β,σ_u and σ_v^2) must be estimated. One possible way of measuring productivity can be trough technical efficiency, that is the relative effectiveness in using inputs. This can be estimated by solving equation (4) for the inefficiency therm u_i with the estimated parameters of MLE as follows

$$TE = E[e^{-u_i}] = E[Y_i / e^{m(x_i; \hat{\beta})} e^{v_i}] \tag{5}$$

with $E[e^{-u_i}] \in [0, 1]$ (Kumbhakar, 2020). The inefficiency therm on its own, u_i, is also an interesting indicator that could be used to compare production units, as inefficiency inversely determines productivity. This parameter, while not fully observable, can be estimated for the given density function and parameter obtained in MLE as

$$E[u_i|\epsilon_i] = \mu_{*i} + \frac{\sigma_* \phi(\frac{\mu_{*i}}{\sigma_*})}{\Phi(\frac{\mu_{*i}}{\sigma_*})}$$

Where $\phi(.)$ is a standard normal probability density function and $\Phi(.)$ is a standard normal cumulative distribution function (Kumbhakar, 2020)[1].

3 Non parametric indexes

One way of approaching the relation between input and outputs is through indexes. We can think of the TFP as the change in the ratio to which a given amount of inputs can be transform into a given amount of output, as the changes of such value would show the part of output growth that is not explained by input growth (O´Donnell,2010). Productivity indexes are flexible estimations in the sense that they do not assume a functional form for the production process, like DEA. Accordingly to this methodology, TFP of observation n at time t is defined as

$$TFP_{nt} = \frac{Q_{nt}}{X_{nt}} \tag{6}$$

Where Q_{nt} and X_{nt} are non decreasing, non negative and monotonic functions that aggregate multiple inputs and outputs respectively(O´Donnell, 2010). Whenever we cannot account for growth gaps, like in cross-sectional data, a good approach are Non-Parametric indexes.

3.1 Färe-Primont

The Färe-Primont index is a Non-parametric estimation based on aggregated quantities function. This linear and non-decreasing functions allow to define a relation between factors that can delimit a frontier of possible production levels, simply by supposing that technology is given for a moment of time and that all producers have access to such technology (O'Donnell, 2011). Once defined a general frontier, each production unit outcome can be evaluated by its relative distance to the possibilities frontier.

Equation (6) can be rewritten as an index of the TFP of unit i relative to unit h

$$TFP_{h,i} = \frac{D_o(x_o, y_i)}{D_o(x_o, y_h)} \frac{D_I(x_h, y_o)}{D_I(x_i, y_o)} \tag{7}$$

Where $D_O(.)$ $D_I(.)$ are Shephard distance functions (Homogeneous and non decreasing) of output and input respectively (O'Donnell, 2011)[2]. These functions measure the inverse largest radial contraction (expansion) of an input (output) that is technically feasible (Chambers, Chung & Färe, 1998).This index has two key characteristics that makes it the most suitable of indexes for assessing cross sectional productivity. First, it is transitive, which means that observations of production units

[1] And $\epsilon = v_i - u_i$, $\mu_{*i} = \frac{-\epsilon\sigma_u^2}{\sigma^2}$, $\sigma_*^2 = \frac{\sigma_u^2\sigma_v^2}{\sigma^2}$.

[2] Shephard input and output distance functions are defined as

$$D_I(y, x) = sup\{\lambda > 0 : (x/\lambda, y) \in T\} \quad ; \quad D_O(x, y) = sup\{\theta > 0 : (x, y/\theta) \in T\}$$

Where T = {(x,y)} and is closed, convex, (0,0)∈T and assumes no free lunch(Chambers, Chung & Färe, 1998). The term λ ultimately captures the returns to scale of the production.

through time are not needed to compare each other, they all can be compared to the frontier in each moment of time. Second, it is multiplicative complete, which means that it can be decomposed into technical, efficiency and scale changes (O'Donnell, 2011). From equation (7), TFP efficiency (TFPE) can be estimated if unit i is compared to the most productive observation in the dataset (equivalent to the maximum TFP estimated, denoted as TFP^*).

$$TFPE_i = \frac{TFP_i}{TFP^*} \tag{8}$$

Finally, TFPE is decomposed into technical, scale efficiency and scale changes and thus solving for TFP_i would give the following output oriented TFP definition

$$TFP_i = TFP^* * OTE_i * OSE_i * RME_i \tag{9}$$

OTE is the output oriented technical efficiency. OTE captures the difference between TFP_i and TFP^* that it is possible while holding the same varieties of input and output and the input level fixed (Molinos-Senante,Maziotis, & Sala-Garrido, 2016). OTE can be written as follows, where $\bar{Q}_i$ is the maximum aggregate output possible when using input level and mix to produce a scalar multiple of the output level of unit i (Hu, Liung & Peng, 2020)

$$OTE_i = \frac{Q_i}{\bar{Q}_i}$$

OSE is the output oriented scale efficiency. OSE captures the difference between TFP_i at a technically efficient point and TFP^* that is possible while holding the same varieties of input and output (Molinos-Senante,Maziotis, & Sala-Garrido, 2016). OSE can be written as follows, where $\bar{Q}_i$ is the same as in OTE and $\tilde{X}$ and $\tilde{Q}$ are output and input levels at the technically efficient point (O'Donnell, 2010)

$$OSE_i = \frac{\bar{Q}_i/X_i}{\tilde{Q}_i/\tilde{X}_i}$$

RME is the residual mix efficiency. RME captures difference between TFP_i at a point on frontier for given combinations of input and output and TFP^* when input and output combinations and levels can vary (Molinos-Senante,Maziotis, & Sala-Garrido, 2016). RME can be written as follows, where $\tilde{X}$ and $\tilde{Q}$ are the maximum output and input levels for the given combinations and Q_i^*/X_i^* is equal to TFP^*(O'Donnell, 2010)

$$RME_i = \frac{\tilde{Q}_i/\tilde{X}_i}{Q_i^*/X_i^*}$$

As production units are usually heterogeneous in key characteristics, size for example, is reasonable to think that frontiers should be estimated taking into account such differences. Färe-Primont indexes usually estimate frontiers within subsets of more comparable observations in the data sets in order to have estimations consistent with observable characteristics and thus, more precise TFP estimations(Orea & Kumbhakar, 2004). While this enables a more precise results, it also can come as a drawback as estimations would be reliant on the subsets defined by the researcher.

3.2 Cost based shares

As it was explained in section 2.2, functional form methodologies assume no technical inefficiency, and thus productivity estimations only requires factor elasticity in order to solve for the TFP term, commonly referred as A_i. As proposed by Hall (1990), when producers are price-takers in factors markets (perfect competition), firms maximize profits, there is perfect information and constant return to scale are assumed, factor share of input j in a group of m inputs can be estimated as follows

$$\hat{\beta}_j = \frac{P_j X_j}{\sum_{i=1}^{m} P_i X_i}$$

Where $P_j$and P_i is the market price of input j and i respectively and X_j and X_i are the quantities used in production of the same pair. Under those assumptions, the elasticity of output with respect to each input is equal to its share in total cost. Here, no assumption over competition in output market needs to be done, only in inputs markets (Hall, 1990). Once factor shares are estimated, TFP of observation i can be defined as

$$TFP_i = e^{y_i - \hat{\beta} x_i} \tag{10}$$

Where x_i is a matrix with all the j inputs involved in the production process and y_i is the output, both in log form. While this estimation is non parametric, in the sense that factor shares are estimated without structure or theoretical constrain, it does suppose a linear relation between output and input in order to define the TFP. This estimation is a bridge between non parametric estimations and structural functional forms, and, because of the linearity assumption, the result can be compared more accurately to the later group of estimations.

4 Production functions

One of the most popular methodologies for estimating productivity is using structural production functions that theoretically explain how outputs relate to inputs, also known as primal techniques (Jimenez, Abbott & Foster, 2019). Once a given form is assumed, econometric models are used in order to estimate the elasticity of factor substitution, that can be interpreted as cost shares under the same conditions exposed by Hall (1990). Technology in production functions is usually represented as Hicks Neutral (factors are affected equally), following the same notion of technology as everything that is not explained by choices of producers, now making explicit its neutrality to all the possible inputs. Hicks Neutral technology also implies that TFP is estimated by subtracting the theoretical expected output from the observed output, both in log form. Finally, when constant returns to scale are fixed, functional forms assume no technical inefficiencies, implying that differences in productivity only come from differences in the technology frontier.

4.1 Cobb-Douglas function

Cobb-Douglas function is a classic functional form that has been widely used in economic literature. It generally takes the following form

$$Y_i = A_i L_i^{\alpha} K_i^{\beta} \tag{11}$$

Where Y_i,L_i and K_i are the levels of the output, labor and capital of observation i respectively. Production inputs are perfectly substitutes and essential to production. While these factors are observed, productivity (A_i) is an unknown parameter for the econometrician. Furthermore, this residual could be capturing measurement errors, noise in the data or observable/unobservable shocks that do not necessarily relate to the productivity term. Taking logarithms and solving for A would be as follows

$$y_i = \beta_0 + \alpha l_i + \beta k_i + \varepsilon_i + \eta_i \tag{12}$$

$$ln(A_i) = \beta_0 + \varepsilon_i$$

With β_0 as the mean TFP in the sample and ε_i the deviations of each observation from the mean (Van Beveren, 2010). η_i can be interpreted as any unobserved shocks that deviate the estimation of y_i and are not related to productivity, thus can be assumed i.i.d (Van Beveren, 2010). Although separating A_i in this way allows to clean the effect of noise and/or measurement errors, the deviations captured by ε_i are still problematic. As part of these shocks can be observed to the producer, factor levels can be determined based on such conjectures, arising endogeneity in the OLS estimation of shares (De Loecker, 2007). Furthermore, if there is a static component in ε_i that producers already known, any change in such variables would also cause endogeneity, as producers react to phenomenons they perceive. Whether if it as a shock or a known characteristic, the problem of ε_i comes from the fact that producers reacts to variables that the econometrician might not be aware of. Assuming that ε_i is fixed and known, factor elasticity could be estimated by OLS if we control by the observable that explain ε_i, and thus TFP would be estimated as

$$A_i = e^{y_i - \hat{\alpha} l_i - \hat{\beta} k_i} \tag{13}$$

4.2 Translogaritmic production function

The translogaritmical (translog) function is a flexible production aproximation that can be understood in three possible ways: 1) As a production function of its own, 2) As a second order Taylor series approximation to a general unknown function, or 3) a second order approximation to a CES function (Boisvert,1982).The translog function relax the usual assumption of constant elasticity of substitution and allows non linear relations with the output and between factors. A function with n inputs can be written as follows (when taking natural logarithms in both sides)

$$\log Y = \log A + \sum_{j=1}^{n} \alpha_j \log x_j + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \beta_{ij} \log x_i \log x_j \tag{14}$$

Equation (14) shows that translog function, when assumed as a function of its own, is a generalization of Cobb-Douglas, as the latter suppose $\beta_{ij} = 0$ (assuming that the interactions of factors or its quadratic therms are not related with the output). When the Cobb-Douglas form is rejected ($\beta_{ij} \neq 0$) further assumptions must be tested, convexity and monotonicity more specifically, as translog functions may not be well behaved (Boisvert,1982). Added to these issues, the same limitations from the Cobb-Douglas functions arise as both are estimated by OLS. Like in the previous functional form (and under the same assumptions), A is equal to the total factor productivity, that solving equation (13) would define TFP as

$$\log A = \log Y - \sum_{i=1}^{n} \hat{\alpha_i} \log x_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \hat{\beta_{ij}} \log x_i \log x_j \tag{15}$$

Having multiple elements per factor within the regression makes the translog function more susceptible to having collineality than Cobb-Douglas (Boisvert,1982). Therefore, it is key to check if the defined inputs are colliear. One last common practice with this function is testing for different definitions of inputs, separating or aggregating them based on theoretical concepts, as the flexibility of the function allows to have a broader understanding of the way factors relate to each other. This is specially useful with micro data, that usually have inputs defined in desegregated units.

5 Production functions limitations and solutions

The problems that ε_i brings in functional form estimations are not minor. Biased factor elasticity estimations implies that TFP cannot be estimated as it would not allow the functional form to properly capture the relation between inputs and outputs. Most literature developed to control this issue (Levinsohn and Petrin, 2003) and (Olley and Pakes,1996) is based on panel data. However, for cross-sectional data, there are still some solutions.

5.1 Fixed Effects

As implied in section 4.1, if the endogenous part of TFP, ε_i, is known and measurable, fixed effects can properly control the endogeneity issue in the OLS estimation (Pavcnik, 2002). This could solve the problem when the endogeneity is caused by production unit fixed characteristics that determine the levels of output. Once accounted those therms, it could be assumed that the levels of inputs are independent from the productivity term, implying that observable shocks to the producers are assumed to be equal to zero (Van Beveren, 2010). In this way, TFP would be the output part that is not explained by inputs nor firm level characteristics. While being effective for controlling by observable characteristics that don't change over time, the assumption of no observable shocks is strong (Van Beveren, 2010), making fixed effects a good methodology when there is diversity in observable characteristics and good reasons to assume constant shocks for all the producers.

5.2 IV

One of the most popular approaches is Instrumental Variables. If the researcher manage to find a set of variables that can explain total output only through the endogenous variables (Production factors), and this variables are not correlated to the unobserved shocks, the factor shares of each input could be effectively estimated. IV relaxes the assumption of exogeneity of inputs from productivity (Wooldridge, 2009), allowing to capture production functions where producers choose their factors based on their productivity. However, it also implies that TFP is exogenous to the producers, ruling out the existence of investment in productivity, like R&D (Van Beveren, 2010).

In theory, and when assuming perfect competition in input and output markets, factor's prices are a potentially good instrument as they relate to production only through the endogenous factors and are a key determinant in the optimal choice of such inputs(satisfy exclusion and relevance), (Ackerberg et. al, 2007). However, disaggregated data on prices are hard to find, and differences in prices could also be capturing differences in the quality of the good. Other possible instruments are weather or external shocks, as they imply changes in both supply and/or demand, but this type of data is even harder to find(Van Beveren, 2010).

Although the assumptions are not too demanding, IV has not been widely used in literature precisely because of the difficulties of finding suitable instruments.

5.3 OLS correction

Another way of solving this problem is correcting OLS as proposed by Winsten (1957), that follows the same notion of SFA estimation, just that, instead of estimating with Maximum Likelihood, OLS is cleaned from the stochastic error term. Under the same settings of SFA, the intercept is shifted to ensure that residuals are negative (shifting production in order to make it a frontier to every observation) and technical inefficiency is denoted as the difference between the maximum error to each of the estimated errors. Hence, the corrected OLS frontier would be as follows

$$ln(\hat{y}_i) = \hat{\beta_0}^{cols} + \hat{\beta}_1 x_{i1} + ... + \hat{\beta}_n x_{in} - \hat{\varepsilon_i}^{cols} \tag{16}$$

Where:

$$\hat{\beta_0}^{cols} = \hat{\beta}_0 + \max_i \{\hat{\varepsilon}_i\}, \qquad i = 1, ..., n$$

$$TI = \hat{\epsilon_i}^{cols} = \max_i(\hat{\varepsilon}_i) - \hat{\varepsilon}_i, \qquad i = 1, ..., n$$

This is exactly the same process of SFA estimation with OLS that was exposed back in section 2.2. As also previously noted, this estimates mean inefficiency adjusted technology, cleaned from both external shocks and individual deviations that might cause endogeneity (ϵ_i) (Sickles ,2019). Once technology is adjusted from average inefficiency and endogeneity is not an issue, productivity differences are captured by individual technical inefficiency deviations from the average in the sample, that can be estimated by taking the difference of total output with respect to the frontier ($\hat{y}_i$) as

$$TE_i = 1 - ln(\hat{y}_i) - ln(y_i)$$

Intuitively, the OLS correction and SFA are fairly similar measurements of technical efficiency, and, ultimately, the only relevant difference is the methodology to estimate input elasticity, a factor that may determine which methodology to follow depending on the strengths or limitations of OLS or Maximum Likelihood in the dataset.

6 Summary of methodology review

On previous sections, we have discussed the methodologies available to estimate productivity with cross sectional data. While we have encompassed the main features and theoretical intuitions of each methodology, they have been widely developed in literature and, if the reader is interested in implementing one of these methodologies, we recommend further reading of the cited authors and documents in this paper.

To summarise the methodologies reviewed in the paper and to give the reader a quick guide to check in which situations is auspicious to apply each methodology, this section presents a table containing key information and a recommendation on practical situations in which each methodology should be applied.

Methodology	Assumptions	Limitations	Practical situation
DEA	- Local linear frontier. - Returns to scale (if needed) - Firms follow cost minimization or profit maximization process.	- Sensitive to data noise. - Representative weights of inputs or outputs is needed if dataset is not large enough. - Sensitive to inefficiency distribution in data.	Useful when there is no price data and/or no functional form is known or assumed.
SFA	- Independence of u_i and v_i. - Orthogonality of v_i and u_i with X_i. - Assumptions of the chosen functional form (taken as the general frontier). - Distribution function of v_i and u_i.	- Can only estimate Technical Efficiency from an assumed distribution. - Due to ML estimation, can be biased in small samples and is sensitive to starting points.	Useful to complement functional form analysis, as it allows to test assumptions and give alternative estimations under fairly similar settings.
Färe-Primont	- Homogeneous Technology. - Firms follow cost minimization or profit maximization process.	- Sensitive to data noise. - Representative weights of inputs or outputs is needed if dataset is not large enough. - Sensitive to assumptions over homogeneity of observations.	Useful when no functional form is known or assumed and/or a decomposed measurement of TFP is needed.
Cost based shares	- Perfect competition. -Firms maximize profits. -Perfect information. -Constant return to scale.	- Very strong and unusual assumptions	Useful only when assumptions hold
Cobb-Douglas	- ε_i is fixed and known. -Orthogonality of η_i with production inputs. -No technical inefficiency.	- Requires a lot of data in order to assure that assumptions hold.	Useful in most cases, even when no functional form is known. It is the standard methodology in literature.
Translog	- ε_i is fixed and known. -Orthogonality of η_i with production inputs. - Low collineality between inputs. -No technical inefficiency.	- Requires a lot of data in order to assure that assumptions hold. - Convexity or monotonicity might not hold, even if the rest of assumptions do.	Useful in most situations, specially when input data is abundant and multiple specifications are possible.

Table 1: Methodology review

7 Data treatment

Once the researcher defines which estimation, and assumptions, suits better, the data must be transformed into the needed variables in order to properly estimate productivity, a process that is specially sensitive in production functions.

Although production functions where initially formulated as physical relation of inputs and outputs (Cobb & Douglas (1928)), this type of data is hard to find and thus functions are usually estimated with monetary value of inputs and outputs. As Felipe et al (2021) expose, estimations with each type of data may incur in different issues. When using physical quantities factors would capture the real technology elasticity and the endogeneity issue would be the main problem to deal with, apart from finding the data (Felipe et al, 2021). On the other hand, estimations with monetary values, when defining output as added value, would capture the accounting identity (added value equals output minus intermediate inputs), implying that the estimated elasticity would be the factor shares of such identity and that TFP would be the weighted average of wage and profit rates rater than the true productivity residual (Felipe et al ,2021).

The criticism to production function estimations comes back to the late 60's, having renowned economist, like Samuelson (1979), on its side, but have generally been overlooked by literature. It is important to understand the relevance of the critiques to production estimations in order to measure TFP. While it all comes down to the availability of data, proper definitions of variables is key to come across this theoretical limitations. When available, quantity data should be used to define inputs and outputs. If only prices value are available, output must be defined as total output value instead of added value, and input prices should be, ideally, firm level prices to avoid omitted price bias (Van Beveren, 2010).

7.1 Defining Output

The final product of the production process is conceived as the output. For simple production process, output is a singular well defined product, but, in practice, many industries produce more than one product. Multi-product production lines must be treated specially to avoid biased TFP estimations. Ideally, if data allows it, the mix of output types and levels are necessary to estimate a complete production function of each production units (Van Beveren, 2010). If data is not available to that level, firms could be aggregated by type of output or parameters could be estimated allowing variation within production units to account differences in the composition of output (Bernard et al, 2009).

Whether is singular or multiple, output have two common ways of being defined when no quantity data is available: As gross output value or as Added value. Although literature tends to use added value extensively, usually in production situations in which intermediate inputs are produced in the same industry, the previous discussion on prices and physical units have made clear the issues with added value, besides having the extra assumption that added value is separable from intermediate production, and thus the former can be obtain by the subtraction of the latter to the total output.

7.2 Defining Inputs

Production inputs are defined as the factors utilized in the creation of a commodity. While there can be significant differences in production inputs across industries, labor and capital tend to be present in almost all of them. It is crucial to have a good understanding of the roles of factors in production to properly define inputs. Separability tests are good practices for understanding factors relation when there is no clarity in the definitions or there are multiple types of inputs in the same category (different types of machines, for example). The test is based on the idea that the marginal rate of substitution would be independent between separable inputs (factors that can be defined independently), thus, a given partition implies separability when factors behave independently, that can be checked proving the following statement

$$f_j f_{ij} - f_i f_{ji} = 0$$

where f_i and f_j are the marginal products of partitions i and j, both part of a total set of possible inputs, and f_{ij} and f_{ji} are the cross partial derivatives of i with respect to j and j with respect to i (Boisvert, 1982).

Inputs can be essential or non essential. While some production units can have output without having some type of inputs, it is unlikely they produce something if they do not have essential inputs. Including non essential inputs can be problematic in production function methodologies as factors are transformed with logarithms and the absence of the factor in one observation wouldn't allow to do estimations. Researchers usually add a constant value to all the observation to avoid this problem, but as Soloaga (2017) shows, different constant values can imply different estimations.

The proper way of including non essential factors, Soloaga argues, is estimating an optimal constant value trough Maximum likelihood. If parameters and the constant values are estimated, the result would be the values that maximizes the probability of observing the data for a given distribution, not just elasticity, but also the given minimum level of each non essential input in the production (Soloaga, 2017).

Factors quality vary across producers. It is a good practice to account for those divergences when possible. A very common input with this type of variance across observations is land, a common factor in agricultural production. As Restuccia et al proposes (2017), one can regress total output over land characteristics in order to construct a land index quality to finally be included in the production functions as total land times the index. Taking into account the depreciation is also key in production inputs like capital. Overall, if data is available, it is desirable to control over observable to assure the comparability between production units and, ultimately, productivity estimations.

7.3 Summary of Data treatment

To summarize the previous discussion and to give the reader a quick guide to check how to define factors for given datasets, table 2 presents information about input and output definitions depending on data availability and a recommendation on methodologies to use for each type of data. The table is focused on functional form definitions as it is the most demanding of all the methodologies and it assumes that previous steps mentioned, like production composition, separability test, and quality and depreciation data, are previously included.

Data available	Input and output definition	Ideal for
Quantities	- Output: Total quantities produced(Kg or units) - Essential inputs: Total quantities used(Kg or units, control for quality and depreciation if data is available) - Non essential inputs: Total quantities used (Kg or units, optimal constant value needed in functional forms, control for quality and depreciation if data is available)	- Functional forms - DEA - Färe-Primont - SFA - Cost based shares
Prices (Ideally firm level)	- Output: Total value of production (Constant prices) - Essential inputs: Total cost of input (Constant prices) -Non essentila inputs: cost of input (constant prices optimal constant value needed in functional forms, control for quality and depreciation if data is available)	- Functional forms - SFA

Table 2: Data treatment

8 Practical application

To apply the previous methodologies, a rich dataset of coffee production is used. The data is a cross-Sectional survey on 654 coffee farms and 2240 individuals in Huila and Tolima, Colombia, realized on August of 2021 with help of Federacion Nacinal de Cafeteros (FNC) and Alianza EFI. At household level, the survey contains information on agricultural production, labor usage (desegregated by household and hired labor), capital usage and types of capital tenure, cultivated land, constructions, fertilizers usage and technical assistance. At individual level, the survey contains basic personal information(age, education, etc), labor (sector, income, months worked) and financial inclusion data. Other external datasets for prices (SIPSA) and weather and soil quality are used to complement the survey data.

Coffee is a permanent and main crop in these regions, making it the principal agriculture output of the interviewed farms, determining labor usage, capital investments and even side crops, cultivated in order to protect coffee plant. The homogeneity in production facilitates the comparability of factors and output across the sample and this orientation towards coffee makes it easier to refine estimations as capital and output measurements and characteristics are specific to this crop.

Output was defined as total agricultural output value. This allows to capture heterogeneity on side crops productions, that was mainly plantain and avocado, and to capture differences in coffee quality, as some farms in the sample produced certified coffee that implies differences in cultivation practices. Multiple definitions of output where used to check robustness, mainly coffee output in physical quantities and coffee output value, and results where fairly similar. For the inputs, labor was defined as total hours employed, land as total hectares cultivated and capital as total value of machines. labor was disaggregated in hours of hired labor, both in permanent and temporal crops, and family labor, in the same categories. Multiple definitions of labor where tested, but separability test showed that proper definition was the aggregation all the units of labor. For capital, there are 10 different types of machinery specialized in coffee production and treatment, initially, each of the ten types of machines where included as individual quantity inputs, taking into account depreciation, but testing for separability showed that elasticity were not statistically different, implying that capital should be defined as a single input. To deal with the heterogeneity in the types of machinery, they where aggregated by firm level price. Land was defined as total hectares cultivated times a land quality index, as proposed by Restuccia et al (2017), that simply account for relative fit of the land for agricultural activity compared to the overall quality of land in the sample. More precisely, the index was constructed with data on coffee production aptitude, nutrient availability, Nutrient retention capacity, Workability, atmospheric pressure, sun radiation and average precipitation. To control over observable characteristics, fixed effects for municipality, farm size and type of coffee (certified or not) where included.

From the total dataset, observations with atypical values in key characteristics, like land and total coffee production, where eliminated. The robustness of the sample allowed us to do the previous exercise by simply defining atypical values as outliers from the sample distribution. Observations without essential inputs (land, labor and capital) or without output (as some producers did not have a harvest in the time covered by the survey) where also eliminated in order to have a set of highly comparable observations, as differences in essential inputs imply severe differences in technology. Finally, when crossing with complementary dataset in order to account for soil characteristics and some fixed effects for functional forms, some observations had missing values and thus where not used in functional form analysis. Hence, observations used in DEA and Färe-Primont methodologies where 540, while functional form estimations used 536 observations.

In order to give the reader further information into the type of data that was used in the practical application, table 3 shows a summary statistic of the main determinant production factor and output of the dataset. Labor is hours worked, land is hectares cultivated, capital is the value in current Colombian pesos, and coffee is on the standard measurement of production, bags, consisting of 125 Kg of the product.

Table 3. Summary statistics

	Mean	SD	Min	Max
Labor (Hrs)	7317	11671.5	24	134400
Land (Hect)	6.0474	7.653582	0.0001	96.9000
Capital (COP)	17412371	18889176	36615	163347787
Coffee (125kg)	28.009	44.52433	0.010	450.000

9 Results

9.1 Technical efficiency (TE)

The first part of the productivity analysis will encompass methodologies that assume no technological difference (or change), that is DEA, SFA, OLS correction (COLS) and Färe-Primont methodologies. As constant returns to scale assumptions are specially strong in agricultural context (all production units operate at optimal scale (Huguenin, 2012)), DEA estimations assumed variable returns to scale and were estimated as output oriented. For methodologies that assume functional forms(SFA and OLS correction), Cobb-Douglas production function was assumed. Table 4 shows the results of technical efficiency estimations.

Table 4. Technical efficiency estimations

	N	Mean	SD	Min	Max
DEA	540	0.1296449	0.1601701	0.0001543	1.0000000
DEA bootstrap	540	0.1089157	0.1291437	0.0001259	0.8584920
OLS correction	536	0.5987	0.1824248	0.3641	1.1132
SFA	536	0.2792563	0.2056839	0.0000054	0.8458131
Färe Primont	540	0.6440326	0.3478945	0.0006377	1.0000000
Färe Primont(multiple frontiers)	540	0.752223	0.3277569	0.000845	1.0000000

Overall, Technical efficiency is dramatically low in the data set. Mean TE in the five methodologies is 0.41, which means that, on average, production units produce 59% less than what is feasible with their inputs. DEA, both simple and bootstrap version, estimate the lowest average technical efficiency. Figure 1 presents the estimated DEA linear frontier.

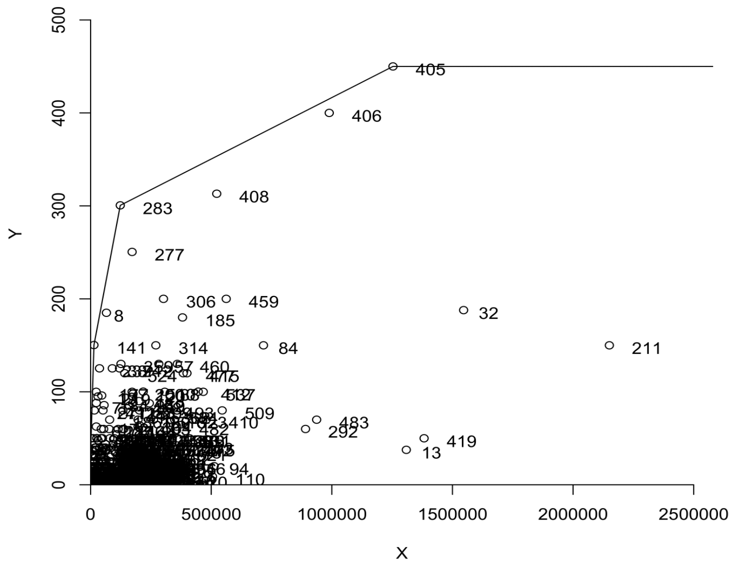


Figure 1: DEA frontier

Production unit labeled as 283 is the most productive in the dataset, as it is in the frontier (is technical efficient) and, compared to other technical efficient units, like 405, has a greater slope, which means that the transformation rate of inputs x into output y is bigger. Apart from the two observations in the frontier and a few others close enough, and, conversely, highly inefficient observations in the down right side of the graph, most of the production units are located in the down left corner, in a zone with low input to output relation and significantly far from the frontier. When estimating DEA with bootstrap, results are fairly similar. Average TE goes from 0.129 to 0.108, implying that original weights of inputs underestimates potential production. Furthermore, maximum technical efficiency in the sample with bootstrap falls to 0.85, this due to the fact that, when allowing to re sampling weights, multiple frontiers are estimated and distance to such frontiers must vary given the composition of inputs of each observation, showing the sensibility of frontier estimation to the relative participation of inputs. Both DEA estimations shows a pervasive low technical efficiency in the sample.

For the functional form based estimations, results are relatively similar. While both estimate higher technical efficiencies, average values are still far from the frontier. COLS estimates a higher level of TE in the sample, with an average of 0.598, and has a significantly higher minimum value than both DEA and SFA. As COLS frontier construction relies on factors and errors distribution, the methodology will be heavily sensitive to extreme values, which, in the presence of such low input to output relationships, will skew production frontier down and estimate relatively higher technical efficiencies. Intuitively, as COLS is a correction of stochastic shocks on the mean, potential frontier estimations will rely on the observed efficiencies rather than the true potential of the overall production structure, overestimating TE in low efficient observation and even allowing observations to be over the frontier, that is why COLS is the only methodology with a maximum value beyond 1.

SFA, in the other hand, shows result more in line with DEA approach. Although average TE is significantly higher (0.27), maximum and minimum values are fairly similar to bootstrap DEA. Due to ML estimation, SFA is able to capture more robust estimations of TE and stochastic shocks compared to COLS. Moreover, the similarity of extreme values of TE with DEA, while clearly coming from the same assumption of linearity in frontier, also show that the assumed distributions appears to be in line with DEA. Under Cobb-Douglas functional form, SFA estimates robust TE values compared to both DEA and COLS, as assumptions are in line with DEA and COLS but doesn't suffer from the problems of the latter, as it properly estimates errors and factor shares. SFA results allow us to conclude that, when properly estimating production frontier and cleaning error therms from stochastic shocks, technical efficiency in the data set is low and, while no production unit is technical efficient, the average one in the sample produces about a third of what is feasible.

Finally, Färe Primont has the highest TE estimations in the sample. This methodology, like COLS, is heavily sensitive to the sample distribution of errors, as OTE is the ratio between maximum output to the observed one with the same input quantities, and thus, if most observations are concentrated in one side of the frontier (as figure 1 shows) the frontier estimations will be downward skewed. While highly productive observations might help OTE index to get closer to the real frontier, they are rare in the dataset and thus will only help observations with the same levels of inputs to get closer to the real TE values, but the majority will be compared to other unproductive observations. When estimating multiple frontiers, aggregating observations by size in order to have more comparable observations, technical efficiency estimations goes even higher. Again, as input quantities are the main determinant, a re distribution of observations based on size can only estimate more precise TE if expected bias come from multiple highly productive observations that should not be compared to structural different production units, like ones with different size. But if low TE is pervasive in the sample, beyond size categorization, this methodology is not efficient.

This six methodologies, while having significant variance on their estimations, show that Technical efficiency is very low in the dataset and, regardless of size, production units are far below their feasible levels of output.

9.2 Index decomposition

In order to fully explore Färe Primont flexibility of decoupling productivity into multiple indicators, in this section we will take a look to the remaining indicators apart from the already analized output oriented technical efficiency (OTE). Table 5 shows the results of scale efficiency, residual mean efficiency and TFP efficiency for estimations with single and multiple frontier.

Table 5. Färe-Primont decompositions

	Single forntier				Multiple frontier			
	Mean	SD	Min	Max	Mean	SD	Min	Max
OSE	0.87444	0.2001869	0.03361	1.00000	0.87083	0.222404	0.03361	1.00000
RME	0.215266	0.1343041	0.002072	1.00000	0.230506	0.1498153	0.003083	1.00000
TFPE	0.120942	0.132622	0.000056	1.00000	0.1421874	0.1394588	0.0000958	1.00000

The output oriented scale efficiency (OSE) is on average 0.87, which means that total factor productivity difference caused by the composition of inputs explains less than 0.13 percentage on the mean. As labor and land is assume identical, the only difference in composition of inputs comes from capital, which have 10 different categories, implying that production units in the dataset are mostly efficient when choosing capital for production. It could also be interpreted that capital affects productivity the same in all of its ten categories, however, the roles of this factors are sharply different, as there are labor augmenting capital inputs, transportation capital and heavy machinery for production within this 10 categories, and it would not be intuitive to think so. If multiple capital categories are assumed to affect productivity the same way, both interpretations could be correct.

Residual mean efficiency (RME) complements OSE and OTE results by showing that the composition of inputs are not determinants of productivity and thus, within the same groups of input classes, production units produce 0.79 times less than what is feasible. It is no surprise that this number is very close to the TE estimations of SFA and DEA methodologies, as the latter assume fixed classes of inputs and compare output result to such fixed definitions. This point is key to understand that OTE differences analysed in the previous section are also explained by the fact that this measurement do allow more flexible definitions of inputs and that, once accounted the differences within homogeneous input classes, the remaining differences of output throughout the compositions is not that determinant of the overall productivity and mainly shows differences in production structures, if this categories can be assumed as such.

Lastly, total factor productivity efficiency (TFPE) shows the skewed distribution of productivity in the sample as it captures the ratio between most productive unit and the observed one. With mean of 0.12, TFPE shows that average production units output is 0.88 less than the most productive observation. Furthermore, standard deviation shows that most observation are still far from most efficient observation, in line with DEA implications. This is the only case where multiple frontier might be helpful as both OSE and RME would not intuitively change within size groups, but TFPE might not be fully comparable between production units with different scales. The results of TFPE with two frontiers, where one group includes observations with land cultivated below or equal to the median and the other the upper side of such distribution, shows that, effectively, a single frontier appears to underestimates efficiency, but the difference in estimations is not statistically significant.

To finish the Färe Primont section, the TFP estimations will be analyzed, that is, as it was exposed in section 3.1, the multiplications of previous factors with no weighing. Table 6 shows the result for single and multiple frontier estimations

Table 6.Färe Primont TFP estimations

	Mean	SD	Min	Max
Färe Primont(single frontier)	0.0692268	0.07591214	0.0000321	0.5723947
Färe Primont(multiple frontiers)	0.1200852	0.1177633	0.0000813	0.8486428

Average values for both methodologies are, once again, strikingly low for both methodologies. While estimating with multiple frontiers shows significantly higher results, both estimations imply extremely low levels of productivity in the sample, with low maximum and average values. Recalling the composition of this measurement, low TFP values imply that the interaction of such components, when multiplied by the highest TFP, TFP^*, is quite low. While OTE and OSE are relatively high, Residual Mean Efficiency (RME) appears to be the determinant component of the low levels of TFP.

9.3 Factor elasticity

This section will analyse the results of functional forms factor elasticity estimated for land, labor and capital. Table 7 shows the estimated elasticity of inputs for the principal functional forms, Cobb-Douglas and Translog function. Both methodologies outcomes are the OLS results with fixed effects (controlling by municipality, farm size and type of coffee).This methodology was chosen to fix the OLS estimations due to the abundance of data available on observable characteristics and the fact that all of the observations where geographically on the same region of the country, allowing the assumption of constant shocks to be feasible. IV correction was tested as well, but we where unable to find proper instruments, mainly because the proposed ones didn't satisfy relevance in the first stage.

While multiple definitions where tested, for both labor and capital, separability test showed that the proper partition of such factors where the simple aggregation of the multiple categories into one. This was unexpected, specially in the case of capital, as multiple categories had quite different roles in production. As the discussion in section 9.2 hinted, the results of OSE are compatible with the separability test results and we can assume that the relation between output and capital categories is the same.

Table 7. Functional forms factor elasticity

	Labor		Capital		Land	
Method	β_w	SE	β_k	SE	β_l	SE
Cobb-Douglas	0.133947*	[0.058680]	0.306084*	[0.126803]	0.563561***	[0.123998]
Translog	0.1261383*	[0.05402999]	0.3103539***	[0.0647618]	0.5437222***	[0.05704137]

Both Cobb-Douglas and Translog function show very similar factor elasticity estimation for the three inputs, the only difference is the statistical significance of capital, that is higher in Translog function due to lower levels of standard errors. The previous result is explained by the fact that translog function do not reject the Cobb-Douglas form (β_{ij} are not statistically different from zero) and thus rejects any non linear relation between inputs and output. Further analysis on Translog functional form validates the assumptions of Cobb-Douglas. Test for monotoniciy and convexity hold, meaning that translog function is well behaved. When analysing the relation between factors, partial elasticity of substitution show that capital, land, and labor are substitutes between themselves, once more, validating the Cobb-Douglas functional form as the appropriated one for this production settings. Consequently, most of the following analysis will be carried out with Cobb-Douglas results.

Land is the most determinant factor in production, explaining more than a half (0.56%) of total input elasticity, followed by capital with 30% and labor with 13%. As figure 2 shows, returns to scale are barley above 1, implying that constant returns to scale might be assumed in functional form estimations. While DEA and Färe Primont methodologies assumed variable returns to scale to allow more flexible estimations, factor elasticity estimated in functional forms precisely shows the assumption of this methodology, that is, production units operate at optimal scale and thus augmenting all inputs in the same scale will have an effect of the same scale in output, ruling out the existence of technical efficiency.

This might sound counter intuitive to the reader, as previous section showed high levels of technical inefficiency. Recall that TE estimated with SFA or COLS take the Cobb-Douglas form as the general upper bound of production and inefficiency as the non stochastic side of production that retains inputs to be transform efficiently, while TFP estimations with functional form assume Cobb-Douglas as an input and output relation that is affected by the non stochastic side of production, the TFP term A_i. Ultimately, both methodologies try to capture the production side that is not explained by inputs nor the error therm with different concepts but with the same functional form. While the notion of an upper bound for production is an attractive way to understand productivity, it cannot capture elements of know how and expertise that allow the same scale of inputs to be transformed into a greater level of output, something that TFP do. While in some cases a frontier can be a precise way of comparing production units, like in highly automatized production lines, TFP is more flexible for capturing human heterogeneity involved in production, a key element in agriculture.

One last remark is that the clear importance of land in the set of inputs shows the need for DEA and Färe Primont, and methodologies estimated with distance functions in general, to estimate robust aggregation weights in order to capture this relative heterogeneity within inputs. While the results in both methodologies are different, the use of distance function require proper weightings of factors to account for the relative importance of production, and thus both estimations will be biased if no proper weight definition is done.

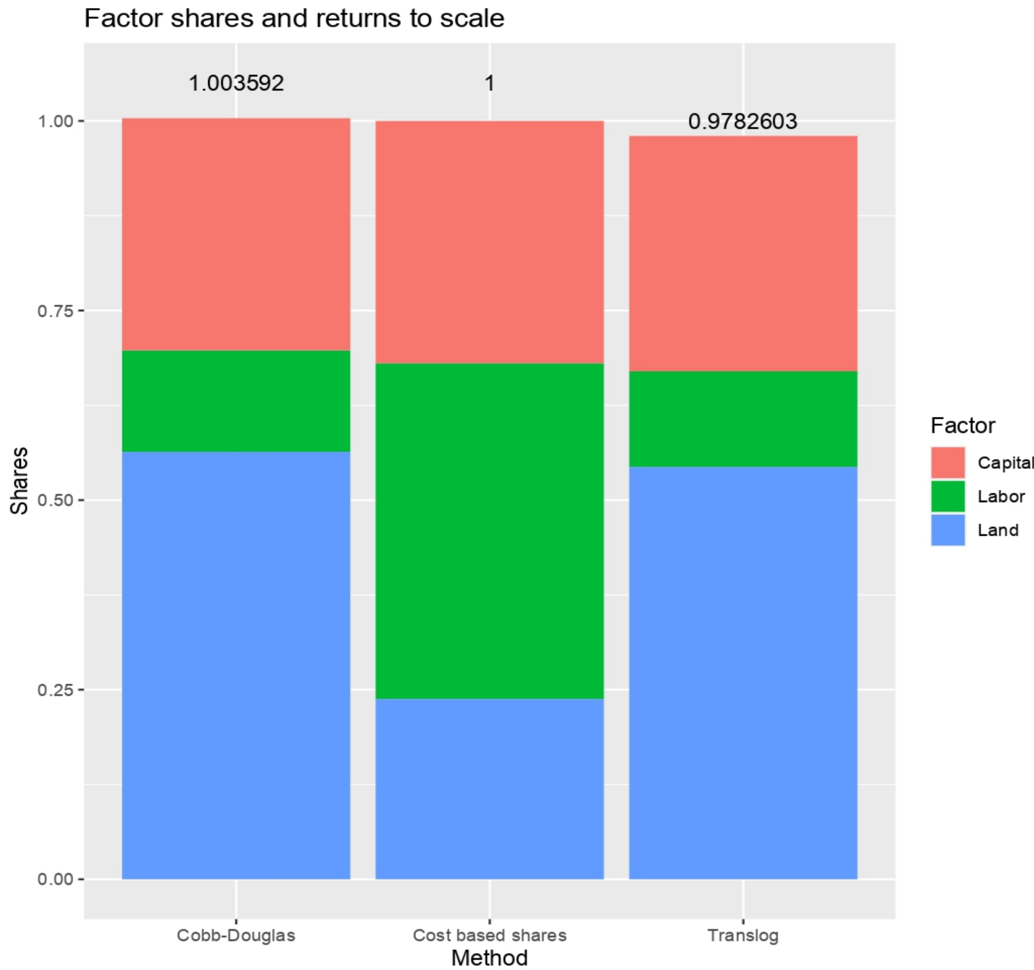


Figure 2: Factor shares

To close this section, cost based shares results are reported. Factor elasticity are somehow similar to the ones in Cobb-Douglas and, by construction, constant returns to scale are attained. While it was clear from section 3.2 that assumptions for this methodology are extremely demanding and do not apply in the context of agricultural production, specially the perfect information assumption, another problem was found. Namely, producers report significantly lower values of cost of usage of land than the ones that could be expected in this settings. Adding this problem of undervaluation of land costs makes the results broadly inconclusive, showing the sensibility of this methodology to data quality and assumptions.

9.4 Total factor productivity in functional forms

The TFP estimations of functional form are displayed in table 8. Again, due to the no rejection of the Cobb-Douglas function, TFP estimation of both methodologies are fairly similar. Mean TFP estimated with Cobb-Douglas is 10.14279, this could be interpreted as: average producers transform their given set of inputs into 10 time the output it would without any skill, know how, expertise and any other differential technology that producer brings into production. Extreme values are also very illustrative of the notion of productivity in TFP. The minimum value in the sample (0.00017) closely resembles the notion of technical inefficiency, as this ”Productivity” negatively affects output and the production unit would be better off by not applying this practices, or inefficiencies. Once again, TFP is a more flexible notion of productivity that can also capture inefficiencies. Maximum values, as in previous methodologies, show the overall low productivity in the sample as some observations are able to produce up to 205 times for their level of inputs with their technology while the average produce around 10.

In this methodology the notion of relative productivity might be fuzzier than in previous definitions. For example, two observation with the same TFP level can have completely different methods, knowledge and technology and, yet, the effect of this elements in production is the same. Conversely, there can be a constant element that explains TFP for all the observations with values greater than 1, and the rest of difference might be due to divergences over the same idea or technology. This examples illustrate the weakness of TFP as a productivity measurement, the flexibility turns into a downfall for the interpretation and applicability of the term.

Table 8. TFP estimations in functional forms

Method	N	Mean	SD	Min	Max
Cobb - Douglas	536	10.14279	24.64682	0.00017	205.24523
Translog	536	14.95783	36.9897	0.00026	302.50141

When analyzing the dispersion in productivity, very similar results are found in functional forms. Throughout farm sizes, most observations are concentrated around low values of TFP, while some smaller group observations have significantly high levels of productivity. A key determinant in this behaviour is the type of coffee produced, as the premium price pay for certified coffee multiplies outcomes. This certifications usually imply special treatment in the production process that do not need different types of labor, capital or land, just need different practices with the product. Most of the outliers in figure 3 are precisely farms with such practices.

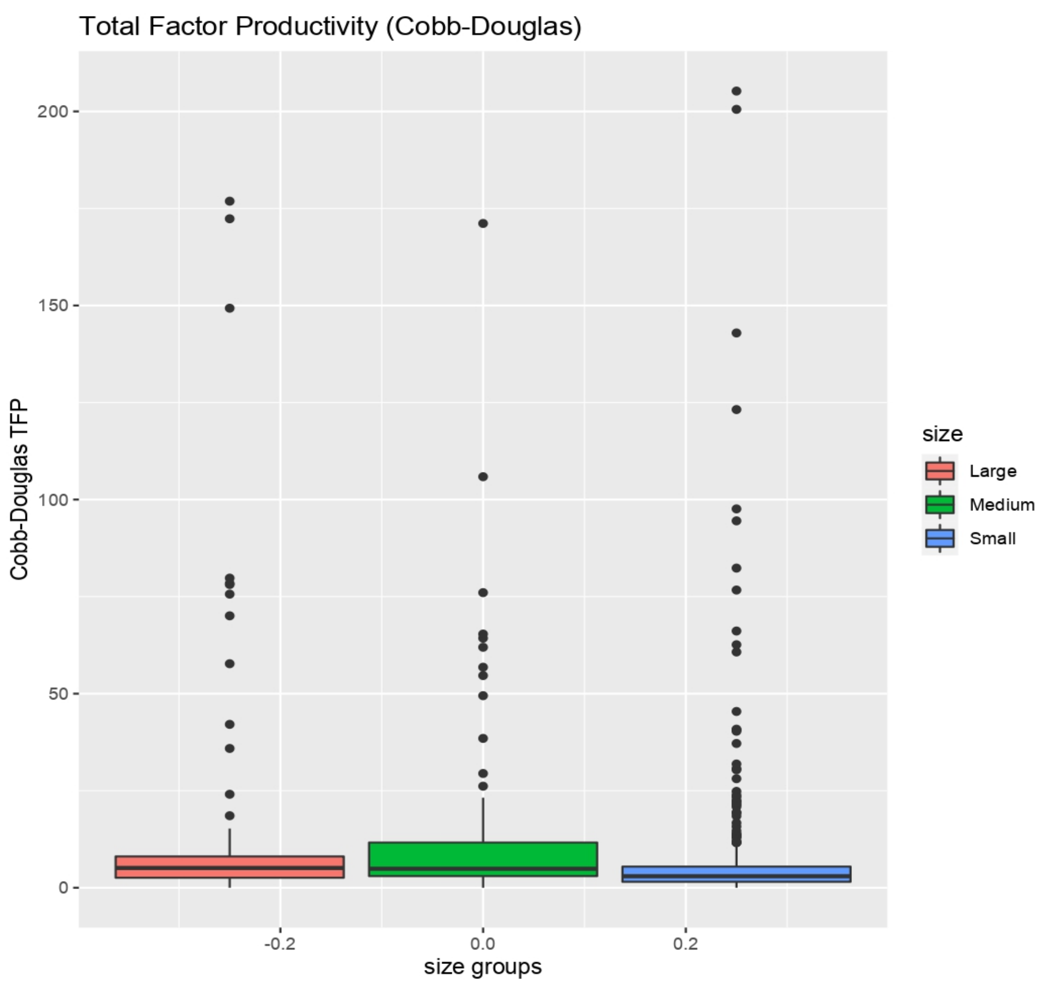
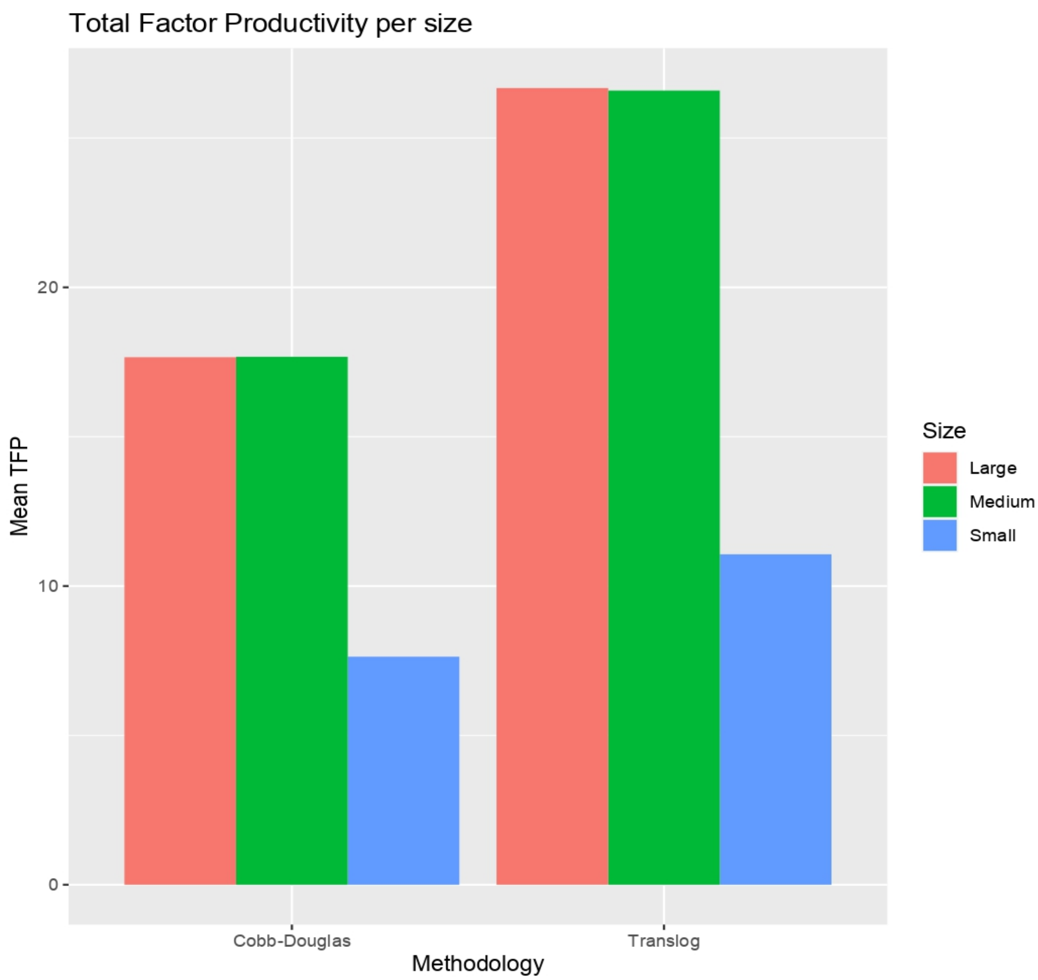


Figure 3: Cobb-Douglas TFP distribution

To compare scales differences present in technical efficiencies, TFP results were tested within size groups of production units. Figure 3 shows that medium and large size farms are significantly more productive than the small ones, with an average TFP value of around 18 while average small farm is barely above 6. While medium and large farm size TFP results are not statistically different, it appears that size and productivity are positively correlated.

When comparing this findings with TE, size do play an important role for TFP. This implies that production units with different size produce with different technology, but as Färe-Primont decomposition showed, scales appears to be irrelevant for technical efficiency and TFP. By the way TE and TFP are estimated in Färe-Primont, results do not compare relative productivity between scale size but relative productivity within the same scale, as size defined the frontier. This can be the main reason why functional forms do captures differences in scale as opposed to Färe-Primont.



10 Conclusions

Although the basic assumptions of each methodology are quite different, TFP estimations points to fairly similar conclusions. Table 9 shows the correlation of productivity estimation of each methodology. Intuitively, methodologies with similar assumptions are more correlated, like Fare Primont,DEA and SFA, or Cobb-Douglas and Translog functions. SFA appears as the bridge between both types of methodologies due to its structure and has relatively high correlation with most of the other methodologies.

Table 9. Correlation in TFP estimations

Method	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) Färe Primont	1.00						
(2) DEA	0.77	1.00					
(3) Cobb - Douglas (Fixed effects)	0.34	0.37	1.00				
(4) OLS correction (Cobb - Douglas)	0.21	0.32	0.18	1.00			
(5) SFA (Cobb - Douglas)	0.54	0.43	0.63	0.15	1.00		
(6) Translog (Fixed effects)	0.34	0.37	1.00	0.18	0.62	1.00	
(7) Cost based shares elasticity	0.21	0.27	0.92	0.14	0.49	0.92	1.00

Ultimately, different outcomes come from different assumptions. We cannot immediately say that DEA estimations are flawed if there is prove of constant returns to scale in functional forms and they where estimated with VRS. The same can be argued with technical efficiency, as functional forms show low levels technical inefficiency(negative productivity). Methodologies, ratter than competing forces, are different lens that allow the researcher to understand production results under different perspectives. The results showed how very different methodologies allowed to clarify results from each other and how cross analysis where extremely useful to understand agricultural production.

Due to the discussion in the results section, we argue that Cobb-Douglas TFP results are the most appropriate indicator of productivity in the dataset. Both data and context are key to define which of this tools is the most useful, but a deep analysis requires to understand the phenomena form different lens. While in this case Cobb-Douglas and SFA estimations showed the most robust results, other methodologies, or combinations of methodologies, can become more appropriate in different context.

In conclusion, cross sectional methodologies for estimating productivity are complementary tools that, when properly applied, can give very useful insight as a group. While assumptions and data context determines which methodology can have more robust results, multiple estimations are highly recommended and desirable to have a more holistic analysis of production performance.

References

Ackerberg, D., Benkard, C. L., Berry, S., Pakes, A. (2007). Econometric tools for analyzing market outcomes. Handbook of econometrics, 6, 4171-4276.

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

Bernard, A. B., Redding, S. J., Schott, P. K. (2009). Products and Productivity*. Scandinavian Journal of Economics, 111(4), 681–709.

Boisvert, R. N. (1982). The translog production function: its properties, its several interpretations and estimation problems.

Chambers, R. G., Chung, Y., Färe, R. (1998). Profit, directional distance functions, and Nerlovian efficiency. Journal of optimization theory and applications, 98(2), 351-364.

Cobb, C. W., Douglas, P. H. (1928). A theory of production.

De Loecker, J. (2011). Product differentiation, multiproduct firms, and estimating the impact of trade liberalization on productivity. Econometrica, 79(5), 1407-1451.

Farrell, M. J. (1957). The measurement of productive efficiency. Journal of the Royal Statistical Society: Series A (General), 120(3), 253-281.

Felipe, J., McCombie, J., Mehta, A., Bajaro, D. F. (2021). Production function estimation: Biased coefficients and endogenous regressors, or a case of collective amnesia?

Hall, R. E. (1989). Invariance properties of Solow's productivity residual.

Hu, Y., Liu, C., Peng, J. (2021). Financial inclusion and agricultural total factor productivity growth in China. Economic Modelling, 96, 68–82.

Huguenin, J. M. (2012). Data envelopment analysis (DEA). A pedagogical guide for decision makers in the public sector, Swiss Graduate School of Public Administration, Lausanne.

Jiménez, M. I., Abbott, P., Foster, K. (2018). Measurement and analysis of agricultural productivity in Colombia. Ecos de Economía, 22(47), 4-37.

Kumbhakar, S. C., Parmeter, C. F., Zelenyuk, V. (2020). Stochastic frontier analysis: Foundations and advances I. Handbook of production economics, 1-40.

Levinsohn, J., Petrin, A. (2003). Estimating production functions using inputs to control for unobservables. The review of economic studies, 70(2), 317-341.

Molinos-Senante, M., Maziotis, A., Sala-Garrido, R. (2017). Assessment of the total factor productivity change in the English and Welsh water industry: A Färe-primont productivity index approach. Water Resources Management, 31(8), 2389-2405.

O'Donnell, C. J. (2010). Measuring and decomposing agricultural productivity and profitability change*. Australian Journal of Agricultural and Resource Economics, 54(4), 527–560.

O'Donnell, C. J. (2011). The sources of productivity change in the manufacturing sectors of the US economy.

Olley, S., Pakes, A. (1992). The dynamics of productivity in the telecommunications equipment industry.

Orea, L., Kumbhakar, S. C. (2004). Efficiency measurement using a latent class stochastic frontier model. Empirical economics, 29(1), 169-183.

Pavcnik, N. (2002). Trade liberalization, exit, and productivity improvements: Evidence from Chilean plants. The Review of economic studies, 69(1), 245-276.

Sickles, R. C., Song, W., Zelenyuk, V. (2020). Econometric analysis of productivity: Theory and implementation in R. In Handbook of Statistics (Vol. 42, pp. 267-297). Elsevier

Soloaga, I. (1999). The treatment of non-essential inputs in a Cobb-Douglas technology: an application to Mexican rural household-level data. Available at SSRN 632569.

Solow, R. M. (1957). Technical change and the aggregate production function. The review of Economics and Statistics, 312-320.

Şişman, Z. (2017). Using data envelopment analysis and malmquist total factor productivity index for agriculture of Turkish regions (Master's thesis, Fen Bilimleri Enstitüsü).

Tziogkidis, P. (2012). Bootstrap DEA and hypothesis testing (No. E2012/18). Cardiff Economics Working Papers.

Van Beveren, I. (2010). TOTAL FACTOR PRODUCTIVITY ESTIMATION: A PRACTICAL REVIEW. Journal of Economic Surveys, 26(1), 98–128.

Winsten, C. B. (1957), 'Discussion on Mr. Farrell's paper', Journal of the Royal Statistical Society Series A, General 120(3), 282–284.

Wooldridge, J. M. (2009). On estimating firm-level production functions using proxy variables to control for unobservables. Economics Letters, 104(3), 112–114.

Zellner, A., Kmenta, J., Dreze, J. (1966). Specification and estimation of Cobb-Douglas production function models. Econometrica: Journal of the Econometric Society, 784-795.