AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
http://ageconsearch.umn.edu
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

Does Duality Theory Hold in Practice?

A Monte Carlo Analysis for U.S. Agriculture

Francisco Rosas
Department of Economics
Iowa State University
Ames, IA 50010
Tel.: 515-294-6357
frosas@iastate.edu

Sergio H. Lence
Department of Economics
Iowa State University
Ames, IA 50010
Tel.: 515-294-6644
shlence@iastate.edu

Selected Paper prepared for presentation at the Agricultural & Applied Economics Association's 2011 AAEA & NAREA Joint Annual Meeting, Pittsburgh, Pennsylvania, July 24-26, 2011.

Copyright 2011 by Francisco Rosas and Sergio H. Lence. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided this copyright notice appears on all such copies.

Abstract

The Neoclassical theory of production establishes a dual relationship between the profit value function of a competitive firm and its underlying production technology. This relationship, usually referred to as the duality theory, has been widely used in empirical work to estimate production parameters without the requirement of explicitly specifying the technology. We analyze the ability of this approach to recover the underlying production parameters and its effects on estimated elasticities and scale economies measurements, when data available for estimation features typical realistic problems. We design alternative scenarios and compute the data generating process by Monte Carlo simulations, so as to know the true technology parameters as well as to calibrate the dataset to yield realistic magnitudes of noise. This noise introduced in the estimation by construction prevents duality theory from holding exactly. Hence, the true production parameters may not be recovered with enough precision, and the estimated elasticities or scale economies measurements may be more inaccurate than expected. We compare the estimated production parameters with the true (and known) parameters by means of the identities between the Hessians of the production and profit functions.

Keywords: duality theory, firm’s heterogeneity, measurement error, data aggregation, omitted variables, endogeneity, uncertainty, Monte Carlo simulations.

JEL Codes: Q12, D22, D81

Does Duality Theory Hold in Practice? A Monte Carlo Analysis for U.S. Agriculture

The Neoclassical theory of production establishes that a competitive firm's optimization problem is characterized by a dual relationship between the value function (profit, cost or revenue function) and the underlying production function (e.g., Mas-Colell, Winston, and Green, Ch. 5). This implies that a given functional form of the production function determines a specific form of the profit, cost or revenue function. Alternatively, for a given functional form used to approximate the value function of the firm's optimization problem, there exists an underlying production function such that some of the value function parameters will appear in it in a specific way.

This dual relationship has been widely used in empirical work as a tool to estimate production parameters without the requirement of explicitly specifying the technology. Shumway (1995) and Fox and Kivanda (1994) list more than one hundred applications of duality theory in nine agricultural economics journals. Typically, empirical studies consist of

i. Approximating the value function (profit, cost or revenue function) by a parametric functional form.
ii. Deriving a set of input demand and output supply equations by applying Shephard's lemma or Hotelling's lemma.
iii. Using econometric methods to jointly estimate the parameters of the system (in some cases together with the parameters of the approximated value function).
iv. Using the estimated parameters to draw conclusions about substitution elasticities, price elasticities, and/or returns to scale.

Some widely cited papers relying on this approach are Ball (1985), Ball (1988), Baffes and Vasavada (1989), Shumway and Lim (1993), and Chambers and Pope (1994). Interestingly, the dataset employed by all of these studies was the one constructed and maintained by Eldon Ball for US input/output price and quantities (USDA-ERS).

Conclusions from applying the duality approach may be heavily influenced by the choice of specific functional forms. As a result, a large number of studies intend to test the validity of duality theory and focus on investigating the most preferable (flexible) functional forms (FFF) for empirical purposes (Guilkey, Lovell and Sickles, 1983; Dixon, Garcia and Anderson, 1987; Thompson and Langworthy, 1989). Analyses of this type usually consist of the following steps. First, a parametric functional form is selected to approximate the production technology. Several parameter scenarios are chosen, and simulated observations corresponding to the "true" production data generating process (DGP) are obtained for each scenario. Second, a set of input and output prices is computed under the assumption of profit maximization. Third, depending on the objective, the profit or cost function is approximated by an FFF and the system of input demands and output supplies is derived. Fourth, econometric methods are applied to estimate the set of parameters of the approximated system, which are finally compared with the true (and known) production parameters.

The aforementioned studies focusing on FFFs assume that the basic tenets underlying duality theory (i.e., perfect competition, profit maximizing behavior, and certainty) hold. Therefore, they only consider empirical deviations from duality theory stemming from the choice of functional form. However, this type of analysis is restrictive in the sense that the DGP used to recover the production parameters is free from the problems usually encountered in the data available to practitioners. Therefore, these studies provide little guidance regarding how well duality theory applies to the empirical analysis of real world data.

In this paper, we propose to analyze the ability of the duality theory approach to recover the underlying production parameters from data featuring typical realistic problems. Among other realistic properties, the simulated data for alternative scenarios include (i) unobserved heterogeneity across firms; (ii) measurement errors in the observed variables; (iii) output and input data aggregation; (iv) omitted variable netputs; (v) omitted quasifixed netputs; (vi) endogenous output and input prices; (vii) optimization under uncertainty. To make the analysis meaningful, the simulated data are calibrated to yield realistic magnitudes regarding the noise arising from each source. Monte Carlo simulations are used to compute each DGP scenario, which allows us to know the true technology parameters. By construction, the aforementioned elements introduce noise in the estimation and prevent duality theory from holding exactly. Hence, the true production parameters may not be recovered with enough precision, and the estimated elasticities or scale economies measurements may be more inaccurate than expected.

Early efforts intended to test the validity of duality theory in practice (Burgess, 1975 and Appelbaum, 1978) failed to identify the source of the discrepancy between conclusions from the primal and dual approaches. The authors used real-world data which are expected to suffer from the aforementioned problems, and therefore they did not know the true DGP. As a result, when the primal and dual approaches led to conflicting results, they could not establish which one was preferable. In addition, the authors did not use a self-dual functional form to approximate both the production and the cost function (translog). This prevented them to attribute the whole divergence in the estimated parameters to a failure of duality because there is at least some difference attributable to functional specification. An exception is the study by Lusk et al. (2002) who analyzed the empirical properties of duality theory simulating data bearing realistic problems such as low price variability, length of time series and measurement error.

Since we are not interested in testing different functional forms, for convenience we choose a quadratic production function to generate what we call the true production data (input and output quantities) using Monte Carlo simulations. Key advantages of the quadratic production function for the present purposes are that (i) it is a self-dual FFF, and (ii) its second derivatives depend only on parameters and not on variables, which greatly facilitates the analysis. The set of input and output prices is straightforward to obtain assuming profit maximization.

Before proceeding to set up the normalized quadratic profit function, however, noise is added to the variables to replicate the aforementioned real-world problems found in the data used by practitioners. We aim at generating noise comparable to the noise encountered in Eldon Ball's dataset. This dataset is chosen as a reference because it has been widely used in the literature. Given that not all the factors that produce noise are directly observable from the data, we select levels of noise close to what we find in the dataset (e.g., price variability, length of time series) and solve for the maximum level of noise of the other factors (e.g., measurement error, endogeneity of output prices) such that the production parameters can be recovered with a reasonable precision.

Once the profit function is set up and the system of input demands and output supplies is derived, its parameters are econometrically estimated and compared with the true (and known) production parameters. Comparisons are performed by means of Lau's (1976) Hessian identities between production and restricted profit functions, which is straightforward under the advocated quadratic specification.

Theoretical Model

Consider a producer who chooses the level of netputs[1] so as to maximize profits. The producer's problem can be described as follows:

$$max_{[\boldsymbol{y}, y_0]}\{\boldsymbol{p}'\boldsymbol{y} + y_0\} \tag{1}$$

where $\boldsymbol{y}$ is a vector of n variable netput quantities to be determined, $\boldsymbol{p}$ is a vector of n variable netput prices normalized by p_0 which is the price of the numeraire commodity y_0 (also to be determined). We define the augmented vector $[y_0, \boldsymbol{y}', \boldsymbol{K}']$ as a production plan of the production possibilities set S which is a subset of R^{1+n+m} , with m equal to the number of quasifixed netputs (denoted as $\boldsymbol{K}$) that constrain the production possibilities set[2].

While we acknowledge that agricultural production is subject to uncertainty (due to, e.g., weather, pests, and uncertain output prices) that affects the optimal solution, we choose to define a deterministic maximization problem as the baseline for two reasons. First, it greatly facilitates the estimation and recovery of the production parameters. Second, the deterministic case is commonly used in analyzes employing duality theory.

[1] We use the standard definition of netput, where a positive value represents a net output and a negative value represents a net input.

[2] The properties of the set S are: i) the origin belongs to S, ii) S is closed, iii) S is convex, iv) S is monotonic with respect to y_0, and v) nonproducibility with respect to at least one variable input, which implies that at least one commodity is freely disposable and can only be a net input in the production process (a primary factor of production).

In scenarios vii) and viii) below we redefine the problem assuming a producer optimizing under production uncertainty.

Jorgenson and Lau (1974) showed that there exists a one-to-one correspondence between the set S (with properties i-v described in footnote 2) and a production function G defined as:

$$G(\boldsymbol{y}, \boldsymbol{K}) = -max\ \{y_0\ /\ [y_0, \boldsymbol{y}', \boldsymbol{K}'] \in S\} \tag{2}$$

We follow the convention that $max\{\emptyset\} = -\infty$, where $\{\emptyset\}$ is defined as the empty set, such that the value of the production function is positive infinity if a production plan is not feasible [3]. The set of quasifixed netputs that constrains the set S also constrains the production function G. The maximization problem can be then rewritten as:

$$max_{[\mathbf{y}]}\{\boldsymbol{p}'\boldsymbol{y} - G(\boldsymbol{y}, \boldsymbol{K})\} \tag{3}$$

The solution to this problem is a set of netput demand equations $\boldsymbol{y}^*(\boldsymbol{p}, \boldsymbol{K})$, and a restricted profit function $\pi_R(\boldsymbol{p}, \boldsymbol{K})$, conditional on the vector of normalized netput prices, and the vector of quasifixed netputs.

Lau's Hessian Identities. Lau (1976) derived the relationships between the Hessian of the production function $G(\boldsymbol{y}, \boldsymbol{K})$ and the Hessian of the restricted profit function $\pi_R(\boldsymbol{p}, \boldsymbol{K})$ under the assumption of convexity and twice continuously differentiability of both functions. Omitting the arguments of each function to simplify notation, the identities are as follows:

$$\begin{bmatrix} \dfrac{\partial^2 \pi_R}{\partial \boldsymbol{p}^2} & \dfrac{\partial^2 \pi_R}{\partial \boldsymbol{p} \partial \boldsymbol{K}} \\ \dfrac{\partial^2 \pi_R}{\partial \boldsymbol{K} \partial \boldsymbol{p}} & \dfrac{\partial^2 \pi_R}{\partial \boldsymbol{K}^2} \end{bmatrix} = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$$

[3] The properties of the production function G are: i) the domain is a convex set of R^{n+m} that contains the origin; ii) the value of G at the origin, say $G(0)$, is non-positive; iii) G is bounded; iv) G is closed; and v) G is convex.

$$B_{11} = \left[\frac{\partial^2 G}{\partial \boldsymbol{y}^2}\right]^{-1}$$

$$B_{12} = (B_{21})' = -B_{11}\left[\frac{\partial^2 G}{\partial \boldsymbol{y}\partial \boldsymbol{K}}\right] \quad (4)$$

$$B_{22} = -\left[\frac{\partial^2 G}{\partial \boldsymbol{K}^2}\right] - \left[\frac{\partial^2 G}{\partial \boldsymbol{K}\partial \boldsymbol{y}}\right] B_{11}\left[\frac{\partial^2 G}{\partial \boldsymbol{y}\partial \boldsymbol{K}}\right]$$

These Hessian relationships allow us to "transform" the estimated parameters of the restricted profit function into the parameters of the underlying production function, and then compare these transformed parameters with the true parameters of the production function. The Hessian of the restricted profit function contains the information necessary to calculate the matrix of input demand and output supply elasticities with respect to (own and cross) prices, and with respect to quantities of quasifixed netputs. Therefore, these Hessian identities ultimately allow us to conclude how precisely we estimate these price and quantity demand and supply elasticities.

Simulation of "true" production data

We generate netput quantities and prices that are consistent with profit maximization by solving the optimization problem in (3). We refer to this known dataset as the "true" production data because it is used to evaluate the accuracy with which production function parameters are recovered in the estimation. To make this problem operational, we assume a quadratic FFF for the production function G.

$$G(\boldsymbol{y}, \boldsymbol{K}; \boldsymbol{\alpha}) = \boldsymbol{y}'A_1 + \boldsymbol{K}'A_2 + 0.5\ (\boldsymbol{y}'A_{11}\boldsymbol{y} + \boldsymbol{y}'A_{12}\boldsymbol{K} + \boldsymbol{K}'A_{22}\boldsymbol{K}) \quad (5)$$

where A_1 and A_2 are (nx1) and (mx1) vectors of α_i coefficients, A_{11} is a symmetric and nonsingular (nxn) matrix, and A_{12} and A_{22} are (nxm) and (mxm) matrices. Submatrices A_{11}, A_{12} and A_{22} form a symmetric and negative-semidefinite

$((n\text{x}m)\text{x}(n\text{x}m))$ matrix A of α_{ij} coefficients. All α_i and α_{ij} coefficients are collectively denoted as $\boldsymbol{\alpha}$. This FFF is selected for three reasons. First, it is self-dual, which implies that the functional form of the (constrained and unconstrained) profit function consistent with this production function is also quadratic. This favors the recovery of the true production parameters because the estimation is conducted free from model specification errors. Second, the Hessian matrices of both the production and profit function are functions of parameters only and not of variables; this proves to be useful when evaluating whether production parameters are recovered with enough precision. Third, the normalized quadratic profit function is widely used in empirical analysis of agricultural production. The first-order conditions (FOCs) of the producer's problem are:

$$\boldsymbol{p} - A_1 - A_{11}\boldsymbol{y} - A_{12}\boldsymbol{K} = 0 \tag{6}$$

This system is jointly solved for the vector of variable netput quantities $\boldsymbol{y}$ as a function of the vector of variable netput prices $\boldsymbol{p}$, the vector of quasifixed netput quantities $\boldsymbol{K}$, and the production parameters $\boldsymbol{\alpha}$. The solution is:

$$\boldsymbol{y}^*(\boldsymbol{p}, \boldsymbol{K}; \boldsymbol{\alpha}) = {A_{11}}^{-1}(\boldsymbol{p} - A_1 - A_{12}\boldsymbol{K}) \tag{7}$$

The data generation process (DGP) consists of generating R = FxT = 7.5 million observations for each variable of the vector $[\boldsymbol{y}^*_{ft}, \boldsymbol{p}^*_{ft}, \boldsymbol{K}^*_{ft}; \boldsymbol{a}^*_f]$ such that, for a given variable, each observation corresponds to a firm (farm) f and a period t.[4] The star is used to represent the true data which are consistent with profit maximization. Random values

[4] This implies F = 150,000 farms over T = 50 years which is close to the 50,000 farms in a given state of the Corn Belt, Lake States and Northern Plains regions in the U.S. (Corn Belt states: IA, IL, IN, MO, OH; Lake States: MI, MN, WI; and Northern Plains states: KS, ND, NE, SD). State-level datasets with information on prices and quantities of agricultural outputs and inputs are available for no more than 50 years in the U.S.

of $\boldsymbol{p}_{ft}^{*}$, $\boldsymbol{K}_{ft}^{*}$ and $\boldsymbol{a}_{f}^{*}$ are drawn from selected distributions and plugged into (7) to obtain the vector of quantities $\boldsymbol{y}_{ft}^{*}$ consistent with profit maximization.

The DGP takes into account that there exist three heterogeneous regions composed of heterogeneous firms within each region, such that the heterogeneity of firms across regions is higher than that within each region, and that there exists variability of prices and quantities over time. Note that the vector of parameters $\boldsymbol{a}_{f}^{*}$ does not depend on *t*, which implies the assumption that technology remains unchanged from period one through T. The study of productivity changes over time and productivity measurement, and their effects on the recovery of true production parameters is a topic of future research and is beyond the objective of this paper. This also favors the recovery of the true production parameters because it is free from misspecification arising from the evolution of technology over time.

Random generation of $\boldsymbol{p}_{ft}^{*}$

We generate the true netput prices $\boldsymbol{p}_{ft}^{*}$ by drawing FxT random deviates for each of the n netput prices p_n. This DGP is conducted in two steps to account for the variability over time and the variability across farmers. In the first step we draw T independent random values from a lognormal distribution with mean μ_{p_n} and standard deviation σ_{p_n}. Call each draw as p_{nt}. The value of μ_{p_n} is different for each region and the value of σ_{p_n} is calibrated according to the observed volatility of prices. Price variability is a key element for recovering production parameters, because a high dispersion contributes to the identification of a bigger portion of the production function. We select values of σ_{p_n} that are consistent with observed price variability in Eldon Ball's dataset. In the second step, we shock each period-*t* price draw (p_{nt}) to induce variability across farms. This is done by generating F random deviates from a uniform distribution over the interval $[p_{nt} - e, p_{nt} + e]$. They are denoted as p_{nft}. The value of e is small relative to p_{nt} to acknowledge the contemporaneous low variability of prices that farmers receive and pay. When p_{nt} is an output, we use the ratio between the basis and a futures price to calibrate the value of e. When p_{nt} is an input, transportation costs are used to calibrate de value of e. Also, this value is region-specific to account for the distance between the region and one delivery point.

The independence of the T random draws is made on purpose to avoid autocorrelation in the variables and facilitate the estimation of the system. Imposing autocorrelation in the DGP and taking that specific form into account during estimation to avoid inconsistency in the estimated parameters would yield the same results in terms of recovering production parameters as the claimed serial independency, provided that the sample size is large enough. The correct model specification is done with the objective of favoring the recovery of the true parameters.

Random generation of $\boldsymbol{K}_{ft}^{*}$

We obtain the vector of quasifixed inputs and outputs by drawing R random deviates from a beta distribution. The beta distribution is chosen because it can accommodate different levels of skewness to mimic the observed distribution of these variables at the farm level. Data from the U.S. Census of Agriculture (USDA-NASS) is used to calibrate the parameters of the beta distribution. For example, if farm size is a quasifixed input, we use the 2002 U.S. Census variable “Farms & land in farms, approximate land area” which shows a relative abundance of small-sized farms, and therefore the beta distribution is positively skewed. In this case we use a Beta(0.7,5). If the value of agricultural capital is a quasifixed input, we use the U.S. Census variable “Farms by value of machinery and equipment”, which shows an even more positive level of skewness. In this instance we use a Beta(0.8,4). Similarly, if livestock sales is a quasifixed output, we use the variable “Sales, Cattle & calves sold, farms by number sold”, which is also negative skewed because of the dominance of farms selling a small number of heads. The distribution is Beta(0.8,5).[5] [6]

Random generation of $\boldsymbol{a}_f^*$

The value of $\boldsymbol{a}_f^*$ characterizes the firm’s technology. For each firm, $\boldsymbol{a}_f^*$ consists of the submatrices A_1, A_2, and A (formed in turn by A_{11}, A_{12} and A_{22}). As we mentioned above, there exists firm’s heterogeneity within and across regions, such that firms have a technology that is more similar to that of its peers in the same region than to those in another region. We select first a value of $\boldsymbol{\alpha}$ for each region such that the symmetric ((nxm)x(nxm)) matrix A is negative-semidefinite. Then, to induce heterogeneity across firms within a region, we draw random values for each entry of these matrices and check for negative-semidefiniteness of A; if A is not negative-semidefinite we reject this draw, otherwise we keep it until we reach the required number of $\boldsymbol{a}_f^*$’s in each region.

Random draws from $\boldsymbol{\alpha}$ must imply a distribution of firm’s size that is realistic. We also rely on Census data to accomplish this objective by selecting a positively skewed beta distribution to draw deviates for matrices A_1 and A which heavily influence the size of the variables $\boldsymbol{y}_{ft}^*$ according to equation (7). We use county-level data of the Census variable “Total sales, Value of sales, number of farms”, and by fitting a beta distribution

[5] A common practice is to include land or agricultural capital as a quasifixed input, and livestock sales as a quasifixed output.

[6] Alternatively, the beta parameters can be calibrated by fitting a beta distribution per region using Census’ county-level data for each of the mentioned variables.

we estimate the parameters used to draw the desired distribution of firm's size. Draws for each entry of the matrices have to be correlated between themselves because their size is what determines the ultimate size of the netput quantities $\boldsymbol{y}^*_{ft}$.

Finally, with the values of $\boldsymbol{p}^*_{ft}$, $\boldsymbol{K}^*_{ft}$ and $\boldsymbol{a}^*_f$ drawn, we calculate the variable netput quantities for each farm f and time t ($\boldsymbol{y}^*_{ft}$) consistent with profit maximization using equation (7). Next, with the resulting dataset (shown below) we proceed to estimation.

$$[\boldsymbol{y}^*_{ft}, \boldsymbol{p}^*_{ft}, \boldsymbol{K}^*_{ft}; \boldsymbol{a}^*_f] \tag{8}$$

Following the discussion on the previous paragraph the values of $\boldsymbol{a}^*_f$ that imply a large value of $\boldsymbol{y}^*_{ft}$ are associated with the large values of $\boldsymbol{K}^*_{ft}$, because it is expected that big farms in terms of production and input use also hold a higher quantity of quasifixed netputs.

Estimation of Profit Function Parameters

We approximate the restricted profit function $\pi_R(\boldsymbol{p}, \boldsymbol{K})$ solution to the problem in (3) by an FFF, and derive the set of input demands and output supplies by Hotelling's lemma. For the reasons mentioned above, a normalized quadratic (NQ) profit function is the selected FFF. Therefore, the system used for estimation is the following.

$$\pi_R(\boldsymbol{p}, \boldsymbol{K}; \boldsymbol{\beta}) = \boldsymbol{p}' B_1 + \boldsymbol{K}' B_2 + 0.5[\boldsymbol{p}' B_{11} \boldsymbol{p} + \boldsymbol{p}' B_{12} \boldsymbol{K} + \boldsymbol{K}' B_{22} \boldsymbol{K}] + \mu \tag{9}$$

$$\boldsymbol{y}(\boldsymbol{p}, \boldsymbol{K}; \boldsymbol{\beta}) = B_1 + B_{11} \boldsymbol{p} + B_{12} \boldsymbol{K} + \boldsymbol{\epsilon}$$

where B_1 and B_2 are (nx1) and (mx1) vectors of β_i coefficients, B_{11} is a symmetric nxn matrix, and B_{12} and B_{22} are nxm and mxm matrices. Submatrices B_{11}, B_{12} and B_{22} form a symmetric ((nxm)x(nxm)) matrix B of β_{ij} coefficients, which is exactly the Hessian matrix in the case of the NQ profit function. All β_i and β_{ij} coefficients are collectively denoted as $\boldsymbol{\beta}$. The restricted profit function is included in the estimation, contrasting many empirical studies, because only in this case the parameters of the submatrix B_{22} can be estimated.

This system of equations is estimated by iterated SUR which is the most common method employed in empirical works based on duality theory. We assume that the error terms μ and $\boldsymbol{\epsilon}$ are respectively distributed $N(0,\sigma_{\mu}^{2})$ and $N(\mathbf{0},\boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is the covariance matrix that induces contemporaneous correlation among the netput equations. As we mentioned above, the covariance structure does not impose serial autocorrelation.

We compare the estimation of the Hessian matrix B with that of the production function according to Lau's Hessian relationships to evaluate how precisely we recover the true production parameters, and ultimately the elasticities of input demand and output supply with respect to prices and quantities of quasifixed netputs.

Each proposed scenario makes a different use of the generated data. Below we describe in detail each scenario.

(i) Unobserved heterogeneity across firms;

We analyze the effects of aggregating production data across heterogeneous firms and proceed to the estimation as if it came from a single firm. This is a common practice because the majority of the studies applying duality theory use country-, state- or county-level data as if it belonged to a single firm; however, such a firm does not exist. We are interested in answering the following question: Whose production parameters are we recovering when we pool together production data from several firms?

We take the 2.5 million observations of a region (F=50,000 firms in the region times T=50 years). Firm's heterogeneity arises from their different parameter values $\boldsymbol{\alpha}_f$. For each period t, we aggregate the subvector $[\boldsymbol{y}_{ft}^{*},\ \boldsymbol{p}_{ft}^{*},\boldsymbol{K}_{ft}^{*}]$ across firms to obtain a vector of a “single firm” $[\boldsymbol{y}_t,\ \boldsymbol{p}_t, \boldsymbol{K}_t]$ which still changes over time. Aggregation of netput quantities is performed by adding across firms because they are homogeneous commodities. The n^{th} netput price at period t (p_{nt}) is aggregated by means of a weighted average of the quantity that each farmer sells or uses.

$$\boldsymbol{y}_t = \sum_f \boldsymbol{y}^*_{ft}$$

$$\boldsymbol{K}_t = \sum_f \boldsymbol{K}^*_{ft} \tag{10}$$

$$p_{nt} = (y_{nt})^{-1} \sum_f p^*_{nft} y^*_{nft}$$

where the subscript n represents the n^{th} netput. The resulting vector $[\boldsymbol{y}_t, \boldsymbol{p}_t, \boldsymbol{K}_t,]$ of quantities and prices is used to estimate the system of netput demands and supplies together with the profit function; i.e., the system in (9) is estimated with T observations. Note that the coefficients $\boldsymbol{\alpha}_f$ are not part of the data used for estimation because they are the values against which we compare the estimated parameters according to Lau’s Hessian relationships in (7).

We also aggregate the data of the other two regions into single firms, as described in (9), and conduct the estimation of the system in (9) with T=50 observations per region and including regional dummies. The purpose is to analyze how the precision in recovering the parameters worsens as we aggregate firms that are even more heterogeneous. It is common practice to include data from different states in order to increase the degrees of freedom (limited by having only T years of data) even if these states have very different underlying technology. We evaluate this tradeoff in this scenario.

In what follows, the dataset just described, i.e., with length T and aggregated across farms within a region, is used in the rest of the scenarios. We do not use the farm-level data with yearly observations because this type of information is not usually available in reality. Therefore, when we evaluate the accuracy of parameter recovery in each scenario, we use the results obtained in scenario (i) as the baseline point of comparison. For this reason, we rename the variables resulting in (10) with a star superscript, as well as the estimated values of $\boldsymbol{\alpha}$, to indicate that these are the new points of comparison.

$$[\boldsymbol{y}_t^*, \boldsymbol{p}_t^*, \boldsymbol{K}_t^*; \boldsymbol{\alpha}^*] \tag{11}$$

(ii) Optimization under perturbed prices;

We analyze the effects on recovering the true production parameters when farmers optimize under prices that deviate from the values used by the econometrician for estimation purposes. This is relevant for the following. Farmers solve the maximization problem given a set of output prices that reflects their expectations of the prices at harvest. It is commonly accepted that prediction errors make this difference relevant. Even in the case of locking in the production prices with instruments such as forward contracts, it might be the case that not all the production is sold under this type of agreements. In the case of input prices, some of them are known at the optimization moment because most inputs are used at the beginning of the production period. However, observed prices may differ from true prices as a result of measurement errors in collecting the information. For example, if a farmer uses different types of pesticides but reports an average or just one price, we observe a price that deviates from its true value. In either case, deviations from the true values produce inconsistency in the estimated parameters, and as a result in the elasticities of interest.

The netput price vector $\boldsymbol{p}$ used in the optimization is obtained by adding perturbations to the farm-level true prices.

$$\boldsymbol{p}_{ft} = \boldsymbol{p}^*_{ft} + \boldsymbol{\varepsilon} \quad ; \quad \boldsymbol{\varepsilon} \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{\sigma}^{\boldsymbol{2}}_{\boldsymbol{\varepsilon}}) \tag{12}$$

We aggregate these prices according to equations in (10) to obtain a new set of prices $\boldsymbol{p}_t$ and construct the dataset $[\boldsymbol{y}^*_t, \boldsymbol{p}_t, \boldsymbol{K}^*_t]$ that is used in the estimation of system (9). The estimated parameters $\boldsymbol{\alpha}$ are compared to $\boldsymbol{\alpha}^*$ according to the Hessian identities.

This scenario also captures a different dimension of perturbed prices, which is when the measurement error occurs on the already aggregated price. This is the case when, for example, input or output prices are not readily available for the econometrician and prices from other region or of other similar products have to be used instead.

(iii) Output and input data aggregation

The effects on recovering true production parameters from aggregating different types of outputs and different types of inputs under the same output or input is studied in this section. Production processes employ a variety of inputs to produce several outputs; however, the data usually available to practitioners is not at that level of disaggregation. Also, in some cases, even if data is available for several inputs and outputs, they are aggregated because they are not the objective of the study and/or so as not to excessively penalize the degrees of freedom during estimation.

Aggregation is done as follows:

$$\begin{aligned} y_{n't} &= \textstyle\sum_{n' \in \Omega_1} w_{n't} y^*_{n't} \\ p_{n't} &= (\textstyle\sum_{n' \in \Omega_1} p^*_{n't} y^*_{n't})\,(y_{n't})^{-1} \\ K_{m't} &= \textstyle\sum_{m' \in \Omega_2} K^*_{m't} \end{aligned} \tag{13}$$

where Ω_i is a subset *i* of netputs, and n' and m' are subindices indicating an aggregated netput. The case of $n' \in \Omega_1$ indicates the situations of aggregating variable inputs and output, whereas $m' \in \Omega_2$ indicates the aggregation of quasifixed netputs. The quantity aggregation of the variable netputs ($y_{n't}$) is performed using weights given by the share of the value of each netput ($p^*_{n't}y^*_{n't}$) on the total value of the netputs included in Ω_1; that is $w_{n't} = (p^*_{n't}y^*_{n't})\left(\sum_{n'\in\Omega_1} p^*_{n't}y^*_{n't}\right)^{-1}$. The variable netputs prices ($p_{n't}$) are obtained as an index by taking the ratio between the total revenue of the netputs in Ω_1 and $y_{n't}$. The aggregation of the quasifixed netputs is done by adding quantities across netputs in Ω_2 because prices are not available to construct weights.

We estimate system (9) using the following data $[\boldsymbol{y}_t, \boldsymbol{p}_t, \boldsymbol{K}_t]$, in which some netput quantities and prices (those belonging to the sets Ω_1 and Ω_2) are the result of the described pooling, and others are the ones obtained in (11).

Noting that in this case the production and restricted profit function Hessians have different dimensions, when we evaluate how precise the true production parameters are recovered by the Hessian identities, we have two situations. One situation is with the netputs that were not aggregated, in which case we can still tell which entry of the production function Hessian corresponds to a given entry of the (transformed) restricted profit function Hessian. On the other hand, for the netputs that were aggregated there is more than one entry in the production function Hessian that corresponds to a single entry in the profit function Hessian.

(iv) Omitted variable netputs

Production takes place with several netputs but it is often the case that the econometrician does not observe all of them. This situation can arise due to a misreporting of data from a surveyed producer in which one or more than one netputs are omitted, or when some inputs are not part of the surveyed set of inputs. In either case, while the producer optimally chooses a set of n variable netputs to maximize profits, the econometrician only observes N' of them with $N' < n$. In our setup, we proceed to estimate the system in (9) with a dataset $[\boldsymbol{y}_t,\ \boldsymbol{p}_t,\ \boldsymbol{K}_t^*]$ defined as follows:

$$y_{n't} = \sum_f y_{n'ft}^* \ , \qquad\qquad n' = 1, 2, \dots, N'$$

$$p_{n't} = (y_{n't})^{-1} \sum_f p_{n'ft}^* y_{n'ft}^* \ , \qquad n' = 1, 2, \dots, N' \qquad (14)$$

where the $y_{n'ft}^*$'s in the first equation are the first N' of the vector $\boldsymbol{y}_{ft}^*$ in (8); and the $y_{n't}$'s form a vector $\boldsymbol{y}_t = \{y_{1t}, y_{2t}, \dots, y_{N't}\}$. The prices of the observed variable netputs $\boldsymbol{p}_t = \{p_{1t}, p_{2t}, \dots, p_{N't}\}$ are constructed, similarly to equation (10), as the weighted average of the quantity that each farmer sells or uses. Finally, $\boldsymbol{K}_t^*$ is the same as in (11) because we assume all the quasifixed netputs are observed.

The estimation of system (9), with the mentioned dataset, produces values of $\boldsymbol{\alpha}$ that are compared with $\boldsymbol{\alpha}^*$ by the Hessian identities, taking into account that, as in the previous scenario, the Hessians of the production and restricted profit functions are not of the same dimension.

(v) Omitted quasifixed netputs

This scenario follows directly from the previous but when only the first M' of the m quasifixed netputs are observed by the econometrician. The dataset in this case is $[\boldsymbol{y}_t^*, \boldsymbol{p}_t^*, \boldsymbol{K}_t]$, where $\boldsymbol{y}_t^*$ and $\boldsymbol{p}_t^*$ are given by equation (11) because all the variable netputs are observed, and $\boldsymbol{K}_t$ is defined as:

$$K_{m't} = \sum_f K_{m'ft}^* \qquad\qquad m' = 1, 2, \dots, M' \qquad (15)$$

where the $K_{m'ft}^*$**'s** are the first M' elements of K_{ft}^* in (8), and the $K_{m't}$'s form a vector $\boldsymbol{K}_t = \{K_{1t}, K_{2t}, \dots, K_{M't}\}$. We use the dataset for estimation of system (9) to obtain the parameters $\boldsymbol{\alpha}$ that we compare with $\boldsymbol{\alpha}^*$ taking into account the different dimension of the Hessian matrices.

(vi) Endogenous netput prices

The underlying assumption in the maximization problem in (3) is that farmers are price takers, and according to this, our DGP in (7) is conducted by drawing random prices that completely determine the levels of netput quantities. However, when we deal with aggregate data at the country-, state- or sometimes county-level it is likely that the quantities have a feedback effect on prices. This effect induces endogeneity between prices and quantities, and inconsistency in the estimated parameters and elasticities, if it is not taken into account in the estimation.

To generate data that induces endogeneity between variable netput prices and quantities we create a new netput price variable from the true data vector $[\boldsymbol{y}_t^*, \boldsymbol{p}_t^*, \boldsymbol{K}_t^*]$ in (11) as follows:

$$p_{nt} = p_{nt}^* + \gamma(y_{nt}^* - \bar{y}_n^*) \tag{16}$$

where the parameter γ is negative if n is an output, accounting for the fact that a higher aggregate output quantity induces a reduction in the output price; and is positive if n is an input because a high aggregate use of inputs drive input prices up. Netput quantities' effect on prices comes from deviations of netput quantity from its mean ($\bar{y}_n^*$).

The data used in estimation of system (9) is $[\boldsymbol{y}_t^*, \boldsymbol{p}_t, \boldsymbol{K}_t^*]$ where $\boldsymbol{p}_t$ is given by (16) and the remaining values are those from (11). The estimated value of the production parameters are compared to $\boldsymbol{\alpha}^*$ in (11).

(vii) Optimization under uncertainty;

In this scenario we analyze the effects on the precision of recovering production parameters of ignoring the uncertain nature of the farmer's problem and assuming that it is deterministic. Uncertainty in farmer's decision process comes from events such as random weather, the presence of pests, and unobserved selling prices, among others. In particular, the farmer optimizes by choosing the quantity of output that expects to harvest at the end of the growing season, given the expected prices at which this production will be sold. At the decision moment, the farmer does not observe the true value of these variables.

We generate the data by transforming the problem in (3) into a problem of maximization of expected utility of a risk averse farmer.

$$max_{[\mathbf{y}]}\, \{EU(\tilde{\pi})\} = max_{[\mathbf{y}]}\, \left\{EU\left(\tilde{\boldsymbol{p}}'\tilde{\boldsymbol{y}} - F(\tilde{\boldsymbol{y}}, \boldsymbol{K})\right)\right\} \tag{17}$$

where U is a strictly increasing and twice-continuously differentiable concave utility function whose argument is the uncertain end-of-period profits $\tilde{\pi}$. The tilde ($\sim$) indicates that it is a random variable and E is the expectation operator that integrates over the uncertainty of both random variables ($\tilde{\boldsymbol{p}}$ and $\tilde{\boldsymbol{y}}$). The concavity of the utility function determines the degree of risk aversion. We assume a constant absolute risk aversion (CARA) utility function of the form $U(\tilde{\pi}) = -e^{-\lambda\tilde{\pi}}$ for which the parameter λ is the coefficient of absolute risk aversion, defined as $\lambda = \frac{U''}{U'}$ where U' and U'' are respectively the first and second derivatives of the utility function with respect to the random profits. The FOCs of this problem are:

$$E\left(U'(\tilde{\pi})(\tilde{\boldsymbol{p}} - A_1 - A_{11}\tilde{\boldsymbol{y}} - A_{12}\boldsymbol{K})\right) = \ 0 \tag{18}$$

where $U'(\tilde{\pi}) = \lambda e^{-\lambda\tilde{\pi}}$. This system is solved simultaneously for the optimal values of the expected variable netput quantities using numerical methods. The data generated in this scenario is used to estimate the system in (9) which ignores the uncertainty embedded in the DGP.

In order to isolate the effects on the precision of the estimated parameters from ignoring the mentioned uncertainty, we conduct the estimation of system (9) with farm-level data from of the regions. The DGP of the netput prices, quasifixed netput quantities and the true production parameters was already described in page 9 through 12 and denoted as $[\boldsymbol{p}_{ft}^{*}, \boldsymbol{K}_{ft}^{*}; \boldsymbol{a}_{f}^{*}]$. We use these data to numerically solve for the optimal expected quantities of variable netputs, according to the system in (18), with the parameter λ calibrated so that it is consistent with a relative risk aversion coefficient equal to 2. We denote the solution vector as $\overline{\boldsymbol{y}}_{ft}^{*}(\overline{\boldsymbol{p}}_{ft}^{*}, \boldsymbol{K}_{ft}^{*}; \boldsymbol{a}_{f}^{*})$ whose entries are functions of expected prices ($\overline{\boldsymbol{p}}_{ft}^{*} = E(\boldsymbol{p}_{ft}^{*})$), quasifixed netputs and the true production parameters. Note that in this case we assume no prediction error in prices, such that farmer's price expectations are realized with no deviation from their end-of-season values.

To estimate the parameters $\boldsymbol{\alpha}$ with system (9) we use the data $[\overline{\boldsymbol{y}}_{ft}^{*}, \boldsymbol{p}_{ft}^{*}, \boldsymbol{K}_{ft}^{*}]$ and then proceed to the comparison of the estimated parameters with the true parameters $\boldsymbol{\alpha}^{*}$ with the Hessian relationships.

References

Appelbaum, E. 1978. “Testing Neoclassical Production Theory” *Journal of Econometrics* 7(1):87-102

Baffes, J. and U. Vasavada. 1989. “On the Choice of Functional Forms in Agricultural Production Analysis” *Applied Economics* 21:1051-1061

Ball, V.E. 1985. "Output, Input, and Productivity Measurement in U.S. Agriculture, 1948-79." *American Journal of Agricultural Economics* 67(3):475-486

Ball, V.E. 1988. "Modeling Supply Response in a Multiproduct Framework." *American Journal of Agricultural Economics* 70:813-25

Burgess, D. F. 1975. “Duality Theory and Pitfalls in the Specification of Technologies” *Journal of Econometrics* 3(2):105-121

Chambers R. G. and R. D. Pope. 1994. “A Virtually Ideal Production System: Specifying and Estimating the VIPS Model” *American Journal of Agricultural Economics* 76(1):105-113

Dixon, B. L., P. Garcia and M. Anderson. 1987. “Usefulness of Pretests for Estimating Underlying Technologies Using Dual Profit Functions” *International Economic Review* 28(3):623-633

Fox, G. and L. Kivanda. 1994. “Popper or Production?” *Canadian Journal of Agricultural Economics* 42(1):1-13

Guilkey, D. K., C. A. K. Lovell and R. C. Sickles. 1983. “A Comparison of the Performance of Three Flexible Functional Forms” *International Economic Review* 24(3):591-616

Jorgenson, D. W. and L. J. Lau. 1974. "The Duality of Technology and Economic Behaviour" *The Review of Economic Studies* 41(2):181-200

Lau, L. J. 1976. "A Characterization of the Normalized Restricted Profit Function" *Journal of Economic Theory* 12(1):131-163

Lence, S. H. 2009. "Joint Estimation of Risk Preferences and Technology: Flexible Utility or Frutility?"*American Journal of Agricultural Economics* 91(3):581-598

Lusk, J. L., A. M. Featherstone, T. L. Marsh and A. O. Abdulkadri. 2002. "Empirical Properties of Duality Theory" *The Australian Journal of Agricultural Economics* 46(1):45-68

Mas-Colell, A., M. D. Winston, and J. R. Green. "Microeconomic Theory" *Oxford University Press*. 1995

Shumway, R. C. and H. Lim. 1993. "Functional Form and U.S. Agricultural Production Elasticities" *Journal of Agricultural Resource Economics* 18(2) 266-276

Shumway, R. C. 1995. "Recent Duality Contributions in Production Economics" *Journal of Agricultural and Resource Economics* 20(1):178-194

Thompson, G. D. and M. Langworthy. 1989. "Profit Function Approximations and Duality Applications to Agriculture" *American Journal of Agricultural Economics* 71(3):791-798

U.S. Department of Agriculture, Economic Research Service (USDA-ERS). "Agricultural Productivity in the U.S.". Available at: http://www.ers.usda.gov/Data/AgProductivity/

U.S. Department of Agriculture, National Agricultural Statistics Service (USDA-NASS). "2002 U.S. Agricultural Census". Available at: http://www.agcensus.usda.gov/Publications/2002/index.asp (accessed April 2011).