Duality theory econometrics: How reliable is it with real-world data?

Francisco Rosas
Universidad ORT Uruguay
& Centro de Investigaciones Económicas (CINVE)
Montevideo, Uruguay
Tel. (598) 2707 1806 ext. 2233
frosas@uni.ort.edu.uy

Sergio H. Lence
Department of Economics
Iowa State University
Ames IA 50010, United States
Tel. (515) 294-8960
shlence@iastate.edu

Selected Paper prepared for presentation at the 2015 Agricultural & Applied Economics Association and Western Agricultural Economics Association Annual Meeting, San Francisco, CA, July 26-28

Duality theory econometrics: How reliable is it with real-world data?

Francisco Rosas
Universidad ORT Uruguay
& Centro de Investigaciones Económicas (CINVE)
Montevideo, Uruguay
Tel. (598) 2707 1806 ext. 2233
frosas@uni.ort.edu.uy

Sergio H. Lence
Department of Economics
Iowa State University
Ames IA 50010, United States
Tel. (515) 294-8960
shlence@iastate.edu

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, commonly referred to as 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. We compute the data generating process by Monte Carlo simulations such that the true technology parameters are known. Employing widely used datasets, we calibrate the data generating process to yield a dataset featuring important characteristics of U.S. agriculture. 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. We conclude that, when the dataset bears minimum sources of noise, duality theory is able to recover the true parameters with reasonable accuracy. Also, that when it is employed in time series coming from an aggregation of technologically heterogeneous firms, the parameters recovered are close to the firm at the median of the distribution. The proposed calibration sets the basis for analyzing the performance of duality theory approaches when datasets used by practitioners are subject to other observed and unobserved sources of noise.

Keywords: duality theory, firm's heterogeneity, data aggregation, Monte Carlo simulations, elasticities.

I. Introduction

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 1995). 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 firm's value function, there exists an underlying production function wherein the value function parameters appear in a specific way.

This dual relationship has been widely used in empirical work as a tool to estimate production parameters without 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 described in (ii). In some instances, value function parameters are estimated together with those of the input and output supply system.

iv. Using estimated parameters from (iii) to draw conclusions about substitution elasticities, price elasticities, and/or returns to scale.

Conclusions from duality applications may be 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). These studies assume the basic tenets underlying duality theory, including perfect competition, profit maximizing behavior, and certainty. Therefore, these studies only consider empirical deviations from duality theory stemming from the choice of functional form. However, the data generating process (DGP) used to recover the production parameters in this type of analysis is free from problems commonly encountered in data available to practitioners. As a result, these studies provide little guidance regarding how well duality theory applies to empirical analysis of real world data.

Early attempts to test the validity of duality theory in practice include Burgess (1975), Appelbaum (1978), and Lusk et al. (2002). With the exception of Lusk et al. (2002), they fail to identify the source of the discrepancy between conclusions from the primal and dual approaches. Two are the reasons. They use real-world data with unknown production parameters, and furthermore, they use non-dual functional forms.

This study has three main objectives. First, we aim at showing the steps used to generate a dataset by Monte Carlo simulations that replicates important features of U.S. agriculture. The dataset is a panel of price and quantity variables coming from firms with heterogeneous technology, but is free from any other source of noise.

Second, we take the simulated dataset and aggregate over firms to construct a time series to be used in estimation. The majority of 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. The objective is to answer the following question: Whose production parameters are recovered when pooling together production data from several heterogeneous firms? We aim to identify the consequences of the widespread estimation practice which assumes a “representative” firm in lieu of several heterogeneous firms.

Third, we use the generated data to test the ability of the duality theory approach to recover underlying production parameters from a time-series dataset that bears the minimum possible noise. This, by no means, is a trivial exercise, given the degree of complexity of the DGP. But more importantly, it is relevant because it sets a solid base for a whole array of future studies intended to analyze the empirical properties of the dual approach as the datasets used by practitioners are noisier and resemble more those found in empirical research. From the starting point we set up here, various sources of observable and unobservable noise can be added to the data before proceeding to estimation. Some of them are the following: (i) optimization under uncertainty; (ii) prediction errors in prices and quantities of variable netputs; (iii) omitted variable netputs; (iv) output and input data aggregation; (v) measurement errors in the observed variables; and (vi) endogenous output and input prices. Noise is calibrated to represent features of typical datasets encountered in practice. Because the noise prevents duality theory from holding exactly, true production parameters may not be recovered with enough precision, and the estimated elasticities measurements may be more inaccurate than expected.

As stated above, these issues have not been addressed in previous studies. As a result, future work in this direction constitutes relevant contributions to the literature.

In this study, we generate datasets with characteristics comparable to those encountered in widely used datasets, such as the one constructed and maintained by Eldon Ball for U.S. input/output price and quantities (USDA-ERS), the USDA Agricultural Resource Management Survey database (USDA-ARMS), the U.S. Agricultural Census database (USDA-NASS), and the Chicago Mercantile Exchange (CME) future prices database. We chose the first dataset because it is publicly available and it has been used for applications of duality theory in several widely cited papers (Ball 1985; Ball 1988; Baffes and Vasavada 1989; Shumway and Lim 1993; Chambers and Pope 1994). The remaining three datasets are data sources which provide useful information for calibrating cross-sectional parameters. We seek to calibrate parameters and noise levels directly observed (e.g., price variability and length of time series) and also unobserved (e.g., firm heterogeneity). These three datasets provide useful information to calibrate model parameters. We adopt the criteria of calibrating parameter values to favor recovery of true production parameters, especially for those that are unobservable.[1]

Since we are not interested in testing different functional forms, for convenience we use a quadratic production function to generate a "true" production dataset, or input and output quantities, using Monte Carlo simulations. Key advantages of the quadratic production function for present purposes include (i) being a self-dual FFF, and (ii) having second derivatives dependent only on parameters and not variables, which greatly facilitates the analysis. We obtain the set of input and output quantities by assuming profit maximization, conditional on randomly generated prices.

We set up the profit function and derive the system of input demands and output supplies, to then econometrically estimate its parameters and compare them with the true (and known) production parameters. Comparisons are performed using Hessian identities between production

---

[1] In this study, we generate a panel data of observations across firms and over time. We focus here on the properties of duality theory applications using time series data. The analysis of applications with cross-sectional data is as relevant as the one pursued here, but we leave it for future research. The properties of duality theory using panel data can be studied with the data generated, but they are less frequent in the literature because such datasets are not as readily available.

and restricted profit functions (Lau 1976), which are straightforward under the advocated quadratic specification.

II. Model of a Single Firm

Consider a producer who chooses the level of netputs[2] 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 choice vector of n variable netput quantities, $\boldsymbol{p}$ is a vector of n variable netput prices normalized by p_0 or the price of the numeraire commodity y_0. The augmented vector $[y_0, \boldsymbol{y}', \boldsymbol{K}']$ is referred to as the production plan of the production possibilities set S which is a subset of R^{1+n+m}, with m equal to the number of quasi-fixed netputs (denoted as the vector $\boldsymbol{K}$) that constrain the production possibilities set.[3]

Jorgenson and Lau (1974) showed existence of a one-to-one correspondence between the set S (with properties described in footnote 3) 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.[4]

---

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

3 The properties of the set S include: 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) non-producibility with respect to at least one variable input, which implies at least one commodity is freely disposable and can only be a net input in the production process (a primary factor of production).

4 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 in $\{\boldsymbol{y}, \boldsymbol{K}\}$. Convexity is required because of the convention used in Lau (1974) that $y_0 = -G(\boldsymbol{y}, \boldsymbol{K})$.

The set of quasi-fixed netputs that constrains the set S also constrains the production function G. The maximization problem can be rewritten as:

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

The solution to problem (3) is a set of netput demand equations $\boldsymbol{y}^*(\boldsymbol{p},\boldsymbol{K})$ and a restricted profit function $\pi_R(\boldsymbol{p},\boldsymbol{K})$ which are dependent on the vector of normalized netput prices and the vector of quasi-fixed netputs.

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 continuous differentiability of both functions. Omitting the arguments of each function to simplify notation, the identities are as follows:

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

$$B_{11} = \left[\frac{\partial^2 G}{\partial\boldsymbol{y}^2}\right]^{-1} \tag{4}$$

$$B_{12} = (B_{21})' = -\left[\frac{\partial^2 G}{\partial\boldsymbol{y}^2}\right]^{-1}\left[\frac{\partial^2 G}{\partial\boldsymbol{y}\partial\boldsymbol{K}}\right]$$

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

By defining, in a similar fashion, the production function Hessian sub-matrices as A_{ij}, the identities can be rewritten in the following more compact form:

$$\begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} = \begin{bmatrix} [A_{11}]^{-1} & -[A_{11}]^{-1}[A_{12}] \\ -[A_{21}][A_{11}]^{-1} & -[A_{22}] - [A_{21}][A_{11}]^{-1}[A_{12}] \end{bmatrix} \tag{5}$$

The Hessian relationships allow us to “transform” the estimated parameters of the restricted profit function into 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 quasi-fixed netputs. Ultimately, the Hessian identities allow us to conclude how precisely we estimate demand and supply elasticities.

To make this problem operational, we assume a quadratic FFF for the production function $G(\boldsymbol{y}_{ft}, \boldsymbol{K}_{ft}; \boldsymbol{\alpha}_f)$:

$$G(.) = \boldsymbol{y}'_{ft}A_{1f} + \boldsymbol{K}'_{ft}\, A_{2f} + \frac{1}{2}\, \boldsymbol{y}'_{ft}A_{11f}\boldsymbol{y}_{ft} + \boldsymbol{y}'_{ft}A_{12f}\boldsymbol{K}_{ft} + \boldsymbol{K}'_{ft}A_{22f}\boldsymbol{K}_{ft} - \psi_{ft} \tag{6}$$

where A_{1f} and A_{2f} are $(n \times 1)$ and $(m \times 1)$ vectors of $\alpha_{i,f}$ coefficients, A_{11f} is a symmetric and nonsingular $(n \times n)$ matrix, and A_{12f} and A_{22f} are $(n \times m)$ and $(m \times m)$ matrices of firm f. Submatrices A_{11f}, A_{12f} and A_{22f} form a symmetric and positive semi-definite $\big((n + m) \times (n + m)\big)$ matrix A_f of $\alpha_{ij,f}$ coefficients.[5] We collectively denote all $\alpha_{i,f}$ and $\alpha_{ij,f}$ coefficients as $\boldsymbol{\alpha}_f$.

The quadratic functional form is selected for three reasons. First, it is self-dual—the functional form of the constrained or unconstrained profit function consistent with this production function is also quadratic. This favors recovery of the true production parameters because the estimation is free from errors arising from functional form specification. Second, the Hessian matrices of both the production and profit functions are only functions of parameters;

---

[5] Positive semi-definiteness is required because of the convention used in Lau (1976) that $y_0 = -G(\boldsymbol{y}, \boldsymbol{K})$.

this proves to be useful because the comparison of the profit and production function Hessians does not depend on the set of model variables at which Hessians are evaluated. Third, the normalized quadratic profit function is widely used in empirical analysis (Schuring, Huffman and Fan 2011; Arnade and Kelch 2007; Lusk et al. 2002; Lim and Shumway 1993; Huffman and Evenson 1989; Thompson and Langworthy 1989).

III. Simulation of panel data

The data generation process (DGP) considers variability of prices and quantities over time within three regions composed of heterogeneous firms. Heterogeneity across regions is assumed to be higher than heterogeneity of firms within each region. The DGP consists of generating a panel of F = 10,000 farms, in R = 3 regions, T = 50 years ($R \times F \times T$ = 1.5 million) for each variable of the vector $[\boldsymbol{y}_{ft}, \boldsymbol{p}_{ft}, \boldsymbol{K}_{ft}; \boldsymbol{a}_f^*]$, where f and t index firms and time periods (years) respectively,[6] conditional on the true (*) value of the production parameters set $\boldsymbol{a}_f^*$.

The vector $\boldsymbol{a}_f^*$ does not depend on time, which implies the assumption that technology remains unchanged from period one through T. This assumption favors the recovery of true production parameters because the estimation is free from misspecification that may arise from the evolution of technology over time. This is equivalent to postulating a specific form of netput technological change and proceed to estimation by exactly specifying its form as if the econometrician knew it with certainty. A different model specification of the mentioned technical change would only add noise in the estimation process. The study of productivity

---

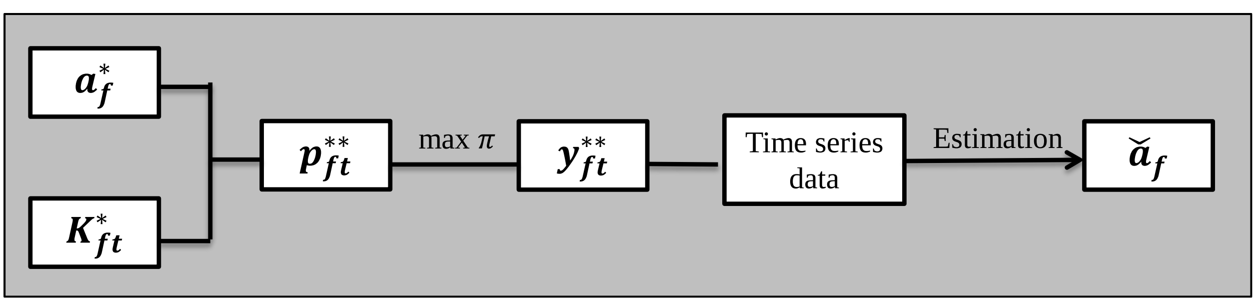
6 The simulated panel corresponds to roughly about one-fifth of the quantity of farms in a given state of the Corn Belt, Lake States and Northern Plains regions in the U.S. (Corn Belt states: IA, IL, IN, MO, OH; Lake States: MI, MN, WI; and Northern Plains states: KS, ND, NE, SD). State-level time-series datasets with information on prices and quantities of agricultural outputs and inputs are available for no more than 50 years in the U.S.

changes over time, their measurement, and their effects on the recovery of true production parameters is a relevant research topic which is beyond the scope of this paper and is left for future research.

To analyze the empirical properties of duality theory, we generate a dataset using Monte Carlo simulations with the objective of illustrating the ability of duality to recover true production parameters when data is free from common problems, and to show the implications on parameters recovery when data is aggregated across firms with heterogeneous technology.

Figure 1 shows the data simulation process. We start by creating the variables conditioning the firm's decisions problem in (3). First, we generate the set of true production parameters $\boldsymbol{a}_f^*$ and the quasi-fixed netputs $\boldsymbol{K}_{ft}^*$. Second, conditioning on these values, we draw variable netput prices $\boldsymbol{p}_{ft}^{**}$. Data generation of $\boldsymbol{a}_f^*$, $\boldsymbol{K}_{ft}^*$, and $\boldsymbol{p}_{ft}^{**}$ are explained in sections III.*A* through III.*C*. Third, we solve a profit maximization problem to obtain the variable netput quantities $\boldsymbol{y}_{ft}^{**}$ (section III.*D*). This study focuses on time-series estimation and therefore we aggregate variables across heterogeneous firms before proceeding to estimation (section III.*E*). The result is a set of estimated production parameters denoted as $\breve{\boldsymbol{a}}_f$.

Figure 1. DGP of dataset used for estimation.



A. Random generation of true production parameters: $\boldsymbol{a}_f^*$

The value of $\boldsymbol{a}_f^*$ characterizes the firm's technology and is unobserved, making its simulation more challenging. From (6), $\boldsymbol{a}_f^*$ consists of the submatrices A_{1f}, A_{2f}, and A_f (formed in turn by A_{11f}, A_{12f} and A_{22f}). As we mentioned above, firm heterogeneity exists both within and across regions, such that technology is more similar between firms in the same region than across regions. Hence, we select values of the elements of $\boldsymbol{\alpha}$ for a "generic" firm such that the symmetric $\left((n+m)\times(n+m)\right)$ matrix A is positive-semidefinite. To induce variation across regions we obtain "regional" $\boldsymbol{\alpha_r}$ sets as deviations from $\boldsymbol{\alpha}$. Then, firm heterogeneity within a region comes from generating parameters in the firm-specific set $\boldsymbol{\alpha_f}$ as deviations from their corresponding regional $\boldsymbol{\alpha_r}$. To assure the matrix A_f and its inverse are positive-semidefinite we draw the entries of the upper triangular matrix C_f, the Cholesky decomposition of matrix $\left(A_f\right)^{-1}$, such that the latter is formed as the matrix product $C_f'C_f$.

The size, dispersion, and skewness of the elements in $\boldsymbol{\alpha}_f$ determine the size, dispersion, and skewness of the netput quantity variables, $\boldsymbol{y}_{ft}^*$, according to the first-order conditions (FOCs) of the firm's optimization problem. Therefore, these elements must be calibrated so as to yield a realistic distribution of quantities produced and used. We rely on the 2002 U.S. Agricultural Census (USDA-NASS), the USDA Agricultural Resource Management Survey databases (USDA-ARMS), and weather data from PRISM at Oregon State University to accomplish this objective (see Appendix for further details).

We calibrate the skewness of the firm-specific deviations from "regional" $\boldsymbol{\alpha_r}$ by fitting a standard beta distribution to the county-level data of the Census variable "Total sales, Value of sales, number of farms" which serves as a proxy for firm size. The shape parameters are estimated by maximum likelihood, yielding a positive skewed distribution. This is consistent with the higher proportion of small firms observed in each region.

The size of the elements in $\boldsymbol{\alpha_f}$ is tackled by inducing positive rank correlation among the beta random shocks, such that a firm producing high levels of output is more likely to use greater amounts of inputs.

Finally, to calibrate the unobserved dispersion of $\boldsymbol{\alpha_f}$ from $\boldsymbol{\alpha_r}$, we assume that observed yield dispersion in a region is a function of unobserved technology heterogeneity and observed random weather shocks. If all firms used the same technology, the observed yield variability would come only from weather shocks. At the other extreme, where all firms differ but no weather shocks occur, all yield dispersion comes from heterogeneity across firms. Most likely the reality is somewhere between the two extremes. We intend to calculate the portion of yield variation attributable to heterogeneity across firms. To this end, we use a panel of firm-specific crop yields from USDA-ARMS database and county-specific weather data (growing season precipitation and temperature) from PRISM over five years, and estimate a fixed-effects model.

Yields are specified as a function of a county-specific constant (the fixed effect) representing the average county's technology and cumulative precipitation and average temperature over the growing season, assuming the constant is correlated with the weather variables. The objective is to isolate the between effects, or the variation in yields across counties not attributable to weather, from the within effects or the variation in yields within a county over time. Firm-level yields are specified as follows:

$$y_{ft} = b_{0c} + b_1 W_{1ct} + b_2 W_{2ct} + b_3 D_{1t} + \cdots + b_6 D_{4t} + \epsilon_{1ft} \tag{7}$$

where c, f and t index counties, firms, and time respectively. Variables W_1 and W_2 are precipitation and temperature, respectively, for the county, and D_1 through D_4 are year dummy variables (2001 through 2004 respectively, with year 2000 as the base). The parameter b_{0c} represents county-level technology and is the focus of our interest. Because we presume it to be correlated with weather variables, we estimate a fixed-effects model where parameters b_1 through b_6 are estimated by demeaning the data (means taken for each county and over time), resulting in the following model (Greene 2003):

$$\ddot{y}_f = b_1\ddot{W}_{1c} + b_2\ddot{W}_{2c} + b_3\ddot{D}_1 + \cdots + b_6\ddot{D}_4 + \epsilon_{2f} \tag{8}$$

with "¨" indicating demeaned variables, estimated by OLS. The county-specific parameter b_{0c} is then recovered by calculating the following equation:

$$\check{b}_{0c} = \bar{y}_c - \hat{b}_1\overline{W}_{1c} - \hat{b}_2\overline{W}_{2c} - \hat{b}_3\overline{D}_1 - \cdots - \hat{b}_6\overline{D}_4 \tag{9}$$

where the " ¯ " indicates means over time (used in demeaning the model) and the " ^ " indicates the point estimate of the parameters. Table 1 provides estimation results.

Table 1. Parameter estimates of fixed effects model, equation (8), to calibrate production function parameter variation, and realized weather shocks on netput quantities.

Dependent variable: $\ddot{y}_f$	**Region 1**	**Region 2**	**Region 3**
Explanatory variables	**Parameter estimates:** $\boldsymbol{b_i, i = 1, \ldots, 6}$		
Precipitation: $\ddot{W}_{1c}$	-0.0002	0.0040	0.0019
	(0.0008)	(0.0007)	(0.0014)
Temperature: $\ddot{W}_{2c}$	-0.3202	-0.0408	0.0028
	(0.0322)	(0.0229)	(0.0359)
Year 2001: $\ddot{D}_1$	5.7192	-12.5194	4.7602
	(7.2434)	(2.0513)	(3.0744)
Year 2002: $\ddot{D}_2$	-17.5728	0.2315	-7.8958
	(5.1586)	(1.6542)	(3.2074)

Year 2003: $\ddot{D}_3$	-17.5792 (4.1648)	2.7325 (1.8828)	-2.5457 (2.7899)
Year 2004: $\ddot{D}_4$	8.7423 (4.1253)	1.6293 (1.8068)	27.9887 (4.1110)
Firm heterogeneity contribution to yield variation: CV($\breve{b}_{0c}$)	0.0578	0.1702	0.4276
Weather variables contribution to yield variation (CV)	0.0726	0.1263	0.4040

Variable $\ddot{y}_f$ denotes demeaned farm-specific crop yields. Accent character " ¨ " represent a demeaned variable. Standard errors in parenthesis.

Finally, the coefficient of variation of $\breve{b}_{0c}$, representing variation across counties, serves to calibrate the unobserved dispersion of the production parameters $\boldsymbol{a_f}$ around the regional mean $\boldsymbol{a_r}$ that are not attributable to weather changes.[7] Note that this coefficient of variation does not represent the estimation standard error of the parameter but the variation across counties of the fitted production coefficients.

B. Random generation of quasi-fixed netput quantities: $\boldsymbol{K}_{ft}^{*}$

We obtain the vector $\boldsymbol{K}_f^{*}$ of quasi-fixed netputs by drawing $R \times F$ beta distributed random deviates. The beta distribution is chosen because it can mimic the different levels of skewness observed in the distribution of these variables at the firm level. Because we choose to represent farm size as the quasi-fixed netput, we use the 2002 U.S. Agricultural Census variable "Farms &

[7] We calibrate the production parameter variation equal to variation between counties as opposed to between firms. Firstly, we do not have firm-specific weather data to calculate the between firms effects. Secondly, the county (and more aggregated) data is likely to have a smaller variation than at the firm level in a given region, favoring parameter recovery.

land in farms, approximate land area" to calibrate the parameters of the beta distribution for each region.[8] This shows a relative abundance of small-sized farms, implying a positively skewed standard beta distribution. Region-specific distributions include: $\boldsymbol{K}_{f,r=1}^{*}$ ~ Beta(0.5679, 6.9707); $\boldsymbol{K}_{f,r=2}^{*}$ ~ Beta(0.6026,9.0446); and $\boldsymbol{K}_{f,r=3}^{*}$ ~ Beta(0.4929, 2.9624).

Because both $\boldsymbol{K}_{f}^{*}$ and A_f determine size of netput quantities, we generate the vector of quasi-fixed netputs imposing positive correlation with the production function parameters. We use the method in Iman and Conover (1982) to impose rank correlation.

Next, we generate time variation in each firm's quasi-fixed netput quantity by means of a multiplicative and independent shock centered at one and uniformly distributed. That is, $\boldsymbol{K}_{ft}^{*} = \boldsymbol{K}_{f}^{*}\epsilon_{ft}$, where ϵ_{ft} ~ Uniform[0.90, 1.10]. The narrow interval implies low variation in firm size over time, which is meant to represent the observed low dispersion over time of aggregate agricultural area in a region.[9]

C. *Random generation of variable netput prices:* $\boldsymbol{p}_{ft}^{**}$

We generate a set of firm-specific exogenous prices for each region. Exogeneity is with respect to the aggregated netput quantity produced. While we acknowledge the existence of price endogeneity, we generate them exogenously in order to have a dataset with minimal sources of noise.

We begin by simulating "national" netput prices to match the properties (mean, standard deviation, and serial autocorrelation) of those found in a time series of futures crop prices from

---

8 It is common practice to include land as a quasi-fixed output.

9 This creates, for each time period, a distribution of quasi-fixed netput quantities for each firm that is not necessarily the regional Beta (it is Beta with other parameters), but still maintains the required skewed shape due to the lower dispersion of firm size over time.

the CME and of input prices from Eldon Ball's (USDA-ERS) dataset. We assume firms base their production decisions on futures output prices and current input prices.

We model netput prices as lognormally distributed and behaving according to an AR(1) processes:

$$log(p_{nt}) = \theta_{n0} + \theta_{n1}\, log(p_{n,t-1}) + \zeta_n \quad (10)$$

where "n" indexes netputs and ζ_n is an error term distributed $\mathrm{N}(0, \sigma^2_{\zeta_n})$. Parameters θ_n are estimated by OLS regressions. Table 2 shows results for each of the n regressions.

Table 2. Estimation results of the OLS regression model used to generate random exogenous "national" prices from equation (10).

	$n=1$	$n=2$	$n=3$	$n=4$	$n=5$	$n=6$	$n=7$	$n=8$
θ_{n0}	-0.0306	-0.0645	-0.0124	-0.0308	-0.0012	-0.0566	0.0413	0.0008
	(0.0376)	(0.0324)	(0.0304)	(0.0638)	(0.035)	(0.0347)	(0.024)	(0.0366)
θ_{n1}	0.6802	0.3443	0.6711	0.9015	0.8613	0.6029	0.8428	0.9225
	(0.0942)	(0.1437)	(0.1128)	(0.0785)	(0.0799)	(0.1166)	(0.0804)	(0.0536)
$\bar{p}$	0.9087	0.9063	0.9629	0.7317	0.9913	0.8672	1.3007	1.01
$\log(\bar{p})$	-0.0958	-0.0984	-0.0378	-0.3123	-0.0088	-0.1425	0.2629	0.0099
σ^2_ζ	0.068	0.0342	0.0372	0.034	0.0439	0.0392	0.0207	0.0237

Note: Standard errors in parenthesis

Dropping the "n" subscript to ease notation, in the long run, the logarithm p_t and p_{t-1} converge to $\bar{p}$ and therefore we can calculate long run expected prices as $log(\bar{p}) = \theta_0/(1-\theta_1)$. The variance of the error term in (10) can be calibrated from observed price variation of Eldon Ball's datasets: $\sigma^2_{log(p)} = \theta_1^2 \sigma^2_{log(p)} + \sigma^2_\zeta$ which implies that $\sigma^2_\zeta = (1-\theta_1^2)\sigma^2_{log(p)}$. In this case, we calibrate price variation from a combination of data observed variance and regression results, and not exclusively from the latter.

To draw exogenous log-normal netput prices, we fit (10) with the estimated parameters, set $log(p_{s=0}) = \theta_0/(1-\theta_1)$, and take a draw from a $N(0,(1-\theta_1^2)\sigma_{log(p)}^2)$ random variable, yielding a netput price for each n in the first iteration, i.e. $log(p_{s=1})$. We repeat this procedure S=10,000 times; we keep the last 50 iterations for the set of exogenous "national" netput prices and burn the remaining iterations.

Finally, we generate firm-specific netput prices $\boldsymbol{p}_{ft}^{**}$ as deviations from the "national" market price, deviations that are small relative to $\boldsymbol{p}_t^{**}$ to acknowledge for the contemporaneous low variability of prices firms receive and pay. A regional average is first calculated as $\boldsymbol{p}_{rt}^{**} = \boldsymbol{p}_t^{**} d_r \varepsilon_{rt}$,[10] where d_r is a regional indicator with mean one across regions[11] and ε_{rt} is a mean one symmetric shock distributed as $\varepsilon_{rt} \sim [0.95 + 0.1Beta(2{,}2)]$. Random variables d_r and ε_{rt} are symmetric and independently distributed. The indicator implies prices of region r are on average $(d_r - 1)$% away from the national average, and the ε_{rt} allows for non-constant deviations over time.

From the regional prices, we generate F firm-specific random prices per region as deviations from the regional average: $\boldsymbol{p}_{ft}^{**} = \boldsymbol{p}_r^{**} \varepsilon_{ft}$, where ε_{ft} is a symmetric mean one shock distributed as follows: $\varepsilon_{ft} \sim [0.80 + 0.40Beta(2{,}2)]$. Shocks ε_{rt}, ε_{ft}, and d_r are independent. Values for ε_{ft} are calibrated using prices from the USDA-ARMS dataset, such that they yield a coefficient of variation of 0.08, which is twice as large as the one observed in the USDA-ARMS dataset.

---

10 The same procedure and shocks are used for $\boldsymbol{p}_{ft}^{**}$.

11 The values of d_r are 0.90, 1.00, and 1.10 for regions 1 through 3 respectively.

For simulation, netput prices are correlated with quantities at the aggregate level, but independent at the firm level. While actual prices received and paid may arguably be correlated with firm size, we assume independence so as to favor parameter identification. Also, observed prices in USDA-ARMS show the majority of firm-level prices are concentrated in four or fewer different clusters in each region; however, we generate a “continuum” of firm-specific prices to favor identification.

D. *Profit Maximization Problem*

The panel dataset is formed by variable netput quantities and prices, and quasi-fixed netputs: $\left[\boldsymbol{y}_{ft}^{**}, \boldsymbol{p}_{ft}^{**}, \boldsymbol{K}_{ft}^{*}\right]$. We first solve the problem in (3) with exogenous prices received or paid $\boldsymbol{p}_{ft}^{**}$. These results are used to test the accuracy of duality theory in recovering production technology using time-series data whose only source of noise is aggregation across heterogeneous firms. This constitutes the minimum possible noise when interested in applying duality theory with time series. Under the normalized quadratic production function $\boldsymbol{G}\left(\boldsymbol{y}_{ft}^{**}, \boldsymbol{K}_{ft}^{*}; \boldsymbol{\alpha}_{f}\right)$ in (11), the FOCs are:

$$\boldsymbol{p}_{ft}^{**} - A_{1f} - A_{11f}\boldsymbol{y}_{ft}^{**} - A_{12f}\boldsymbol{K}_{ft}^{*} = 0 \tag{11}$$

This system is jointly solved for the vector of optimal variable netput quantities $\boldsymbol{y}_{ft}^{**}$ as a function of the vector of variable netput prices $\boldsymbol{p}_{ft}^{**}$**,** the vector of quasi-fixed netput quantities $\boldsymbol{K}_{ft}^{*}$, and the production parameters $\boldsymbol{a}_{f}^{*}$. The solution is:

$$\boldsymbol{y}_{ft}^{**}\left(\boldsymbol{p}_{ft}^{**}, \boldsymbol{K}_{ft}^{*}; \boldsymbol{a}_{f}^{*}\right) = {A_{11f}}^{-1}(\boldsymbol{p}_{ft}^{**} - A_{1f}A_{12f}\boldsymbol{K}_{ft}^{*}) \tag{12}$$

This produces a panel dataset of $(R \times F)$ firms over T time periods that can be used to recover production parameters using time-series or cross-section. We denote this dataset as follows:

$$[\boldsymbol{y}_{ft}^{**}, \boldsymbol{p}_{ft}^{**}, \boldsymbol{K}_{ft}^{*}] \tag{13}$$

E. *Unobserved Firm Heterogeneity.*

Finally, in agreement with this study's objective of testing duality theory using time-series data, before estimation we proceed to aggregate across the F=10,000 heterogeneous firms as if data came from a single firm. This aggregation is performed on the data described in (13). If the objective were to study empirical properties of duality under a cross-sectional dataset, we would have taken one year of the panel and conducted the analysis without aggregating across firms. This is left for future research.

For each period t, we aggregate the subvector $[\boldsymbol{y}_{ft}, \boldsymbol{p}_{ft}, \boldsymbol{K}_{ft}]$ across firms to obtain observations over T=50 time periods (years) of a "single firm" $[\boldsymbol{y}_t, \boldsymbol{p}_t, \boldsymbol{K}_t]$. For netput quantities, we aggregate by adding across firms since they are homogeneous commodities. The n^{th} netput price at period t (p_{nt}) is a quantity-weighted average of the firm-specific netput prices.

$$\begin{aligned} \boldsymbol{y}_t &= \textstyle\sum_f \boldsymbol{y}_{ft} \\ \boldsymbol{K}_t &= \textstyle\sum_f \boldsymbol{K}_{ft} \\ p_{nt} &= (y_{nt})^{-1} \textstyle\sum_f p_{nft} y_{nft}. \end{aligned} \tag{14}$$

The time-series dataset used in estimation is denoted as follows:

$$[\boldsymbol{y}_t^{**}, \boldsymbol{p}_t^{**}, \boldsymbol{K}_t^{*}] \tag{15}$$

IV. Data for estimation

The dataset in (15) includes all $n = 8$ netput quantities and prices, and $m = 1$ quasi-fixed netput. Variable netput prices are exogenous from quantities, but have serial autocorrelation. The DGP yields 0.5 million observations for each of the three regions (F=10,000 firms in the region over T=50 years). We aggregate the 10,000 heterogeneous firms at each time t, resulting in a dataset of 50 observations for each variable per region that we use to estimate a system of netput demands and supplies in (17). To avoid the addition of another source of noise coming from heterogeneous technology across regions, we select region 1 to conduct the estimation, and compare results with the true parameters of that same region. The estimation incorporating data from other more heterogeneous regions to capture a broader area and increase the sample size, which is common in these applications, is shown as a sensitivity analysis.

V. Estimation

We approximate the restricted profit function $\pi_R(\boldsymbol{p}, \boldsymbol{K})$, which solves problem (3), by the following normalized quadratic flexible functional form:

$$\pi_R(\boldsymbol{p}, \boldsymbol{K}; \boldsymbol{\beta}) = \boldsymbol{p}'B_1 + \boldsymbol{K}'B_2 + 0.5\boldsymbol{p}'B_{11}\boldsymbol{p} + \boldsymbol{p}'B_{12}\boldsymbol{K} + \boldsymbol{K}'B_{22}\boldsymbol{K} + \boldsymbol{p}'\boldsymbol{\kappa} \qquad (16)$$

where B_1 and B_2 are $(n \times 1)$ and $(m \times 1)$ vectors of β_i coefficients, B_{11} is a symmetric $(n \times n)$ matrix, and B_{12} and B_{22} are $n \times m$ and $m \times m$ matrices. Submatrices B_{11}, B_{12}, and B_{22} form a symmetric $\left((n+m) \times (n+m)\right)$ matrix B of β_{ij} coefficients, which in the case of the NQ profit function, is exactly the Hessian matrix with respect to $(\boldsymbol{p}, \boldsymbol{K})$. All β_i and β_{ij} coefficients collectively form the set $\boldsymbol{\beta}$. The error structure $\boldsymbol{p}'\boldsymbol{\kappa}$ is consistent with the McElroy (1987) additive general error model (AGEM) applied to the case of profit functions. The $(n \times 1)$ vector of random variables $\boldsymbol{\kappa}$ is jointly normally distributed with mean equal to a $(n \times 1)$ vector of zeros and an $(n \times n)$ covariance matrix $\boldsymbol{\Sigma}_\kappa$. This covariance matrix induces contemporaneous correlation between the equations. Also, the DGP of netput prices—both exogenous and endogenous—was constructed as an AR(1) process, implying serial autocorrelation in the independent variables that needs to be accounted for in the estimation.

We derive the set of input demands and output supplies by Hotelling's lemma, yielding the system to be estimated:

$$\boldsymbol{y}(\boldsymbol{p}, \boldsymbol{K}; \boldsymbol{\beta}) = B_1 + B_{11}\boldsymbol{p} + B_{12}\boldsymbol{K} + \boldsymbol{\kappa}. \tag{17}$$

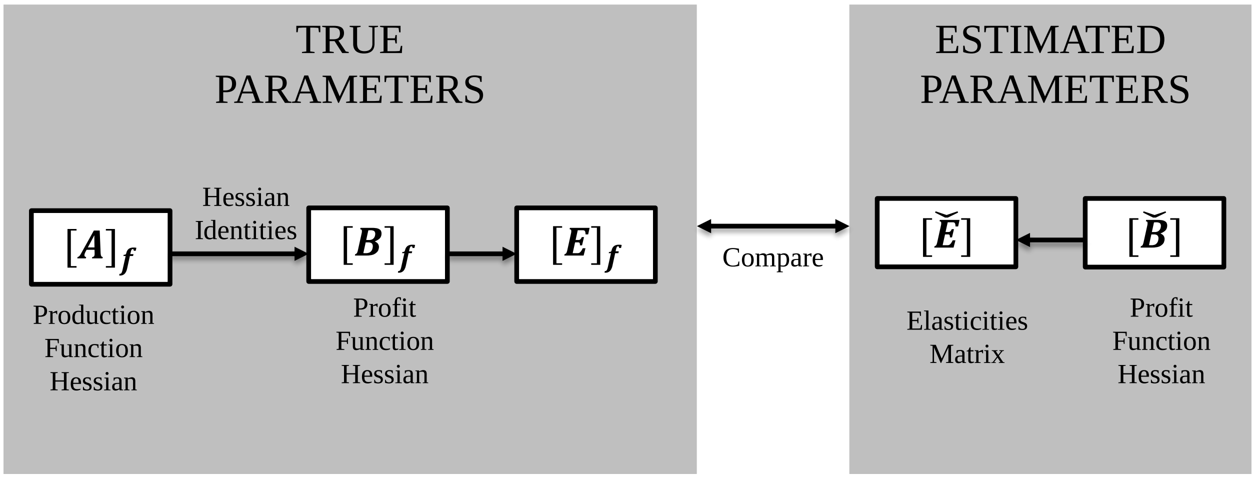
We conduct estimation by iterated SUR, which converges to maximum likelihood, and is the most common method employed in empirical studies based on duality theory. We impose symmetry cross-equation restrictions ($\beta_{ij} = \beta_{ji}, i \neq j$) in matrix B_{11}. We do not estimate the parameters of the profit function because the parameters needed to evaluate the production parameters of interest are present in the demands and supplies.

We treat mean-independence violations in estimation by noting that an inspection of the autocorrelation and partial autocorrelation functions of the time series suggests first differentiation of the data for estimation. This is a consequence of the DGP of price data as AR(1) processes.

The estimated values of matrix B_{11} and vector B_{12} are the focus of our attention; they are, respectively, the marginal effects of prices and quasi-fixed netputs on netput quantities, and therefore they are the base to construct the estimated profit function Hessian matrix $[\check{B}]$ and the elasticities matrix of netput quantities with respect to own price, cross prices, and quasi-fixed netputs $[\check{E}]$. As described in Figure 2, we obtain matrix $[\check{B}]$ from estimation using the described panel data. This matrix is then transformed into an elasticity matrix in a straightforward way.

In order to compare estimated elasticities with true values, we proceed as follows. We begin from the true and known firm-specific production function Hessian matrix $[A]_f$ and convert it into the corresponding profit function Hessian $[B]_f$ using Lau's Hessian identities. We further transform the true profit function Hessian into the true matrix of own- and cross-price elasticities and quasi-fixed elasticities of netput quantities $[E]_f$. Finally, as indicated in Figure 2, we compare the true $[E]_f$ versus the estimated values ($[\check{E}]$) to evaluate how precisely we recover the true price and quasi-fixed netput elasticities under duality theory. Note that this comparison implies that the true values are represented by a distribution of each firm's true parameters, while the estimated values consist of a point estimate and its confidence interval.

Figure 2. Post-estimation comparison between true and estimated elasticities



VI. Results

Figure 3, Figure 4, and Table 3 summarize the results from estimating output supplies and input demands parameters in (17).

In Figure 3 we show how the estimated own- and cross-price elasticity of netput quantities ($\breve{\boldsymbol{E}}_{ij}$) compares with the distribution of true firm-specific elasticities, the mean of the distribution ($\overline{\boldsymbol{E}}_{ij}$) and its median ($\overline{\overline{\boldsymbol{E}}}_{ij}$), for the 64 entries of the 8×8 elasticity matrix. The vertical axis represents the mean of the distribution of true elasticities and the horizontal axis show descriptive statistics of the distribution of true elasticities: the 90% highest probability density interval of the true distribution (the horizontal line), the mean of the distribution (diamond), median of the distribution (filled square), and the SUR estimated elasticity (circle). Therefore, all of the means (diamonds) are along the 45º line. The median (filled square) is to the left or to the right of the mean depending on the skewness of the distribution. The elasticity point estimates (circle) and the 95% confidence intervals (vertical lines) are in all cases within the support of the true distribution. This implies that estimation with a dataset constructed as the aggregation across heterogeneous firms (as if it belonged to a representative firm) is able to recover elasticities that are not only within the relevant range of the distribution but also fairly close to the median and the mean.

A second conclusion arises by noting that the point estimates are closer to the median of the distribution than to the mean. The representative firm is better described by the median of the distribution than the mean. The root mean squared error (RMSE) helps illustrate this conclusion. The RMSE is the average difference between each entry of the estimated elasticity matrix versus its corresponding true elasticity, expressed in elasticity units. We show two alternative values to describe the true elasticity: the median of the true firm-specific elasticity distribution and its mean. When compared to the median of the distribution, the RMSE is:

$$RMSE = \left[\frac{1}{64 \times S} \sum_{i} \sum_{j} \sum_{s} \left(\bar{\bar{E}}_{ij,s} - \breve{E}_{ij,s}\right)^2\right]^{1/2} \tag{18}$$

where S = 10,000 is the number of draws from the limiting distribution of the SUR parameter estimates and the subscript s indicates the s^{th} draw of the ij^{th} parameter. For comparison with the mean we substitute $\bar{\bar{E}}_{ij}$ by $\bar{E}_{ij}$. The RMSE averages over all the $64 \times S$ squared differences. We also provide a measure of its dispersion by calculating the standard deviation of these $64 \times S$ values before averaging over them. The RMSE standard deviation contains two sources of variation or error. One is due to the SUR estimation error within each of the 64 parameters and the other is associated with the variation of the difference between the estimated and the true value of the elasticity across the 64 parameters.

As shown in Table 3, RMSE is 0.048 in the case of the median and more than double (0.111) for the mean. To put these values into perspective, we calculate the percentage deviation of the

RMSE with respect to the descriptive statistics of the true distribution of elasticities. Relative to the median it yields a difference of 12.4% and, as expected, it is higher relative to the mean, 26.3%.

The RMSE standard deviation is 0.078 for the median and a higher value (0.196) for the mean. Given the SUR estimation provides only a minor source of error because the point estimates are all highly significant due to the use of a data with only minor sources of noise,[12] the majority of the RMSE standard deviation is attributed to the deviations between the estimated and the true value across elasticities.

Figure 4 illustrates the estimated results of the eight netput quantity elasticities with respect to the quasi-fixed input. The SUR estimated elasticities (circles) are within the interval of true elasticity distribution for all cases, and similar to the variable netputs case, closer to the median of the distribution than to its mean. As Table 3 indicates, the RMSE is 0.035 in the case of the median, and 0.071 for the mean. The size of the RMSE standard deviation also suggests high variation (of their dispersion relative to the true value) across the 8 elasticities. Our estimated elasticities are 7.5% apart from the median absolute value of the true elasticity and 14.7% from the mean absolute value.

---

[12] Results are available from the authors.

Figure 3. Elasticities of variable netput quantities with respect to prices. True versus estimated values.

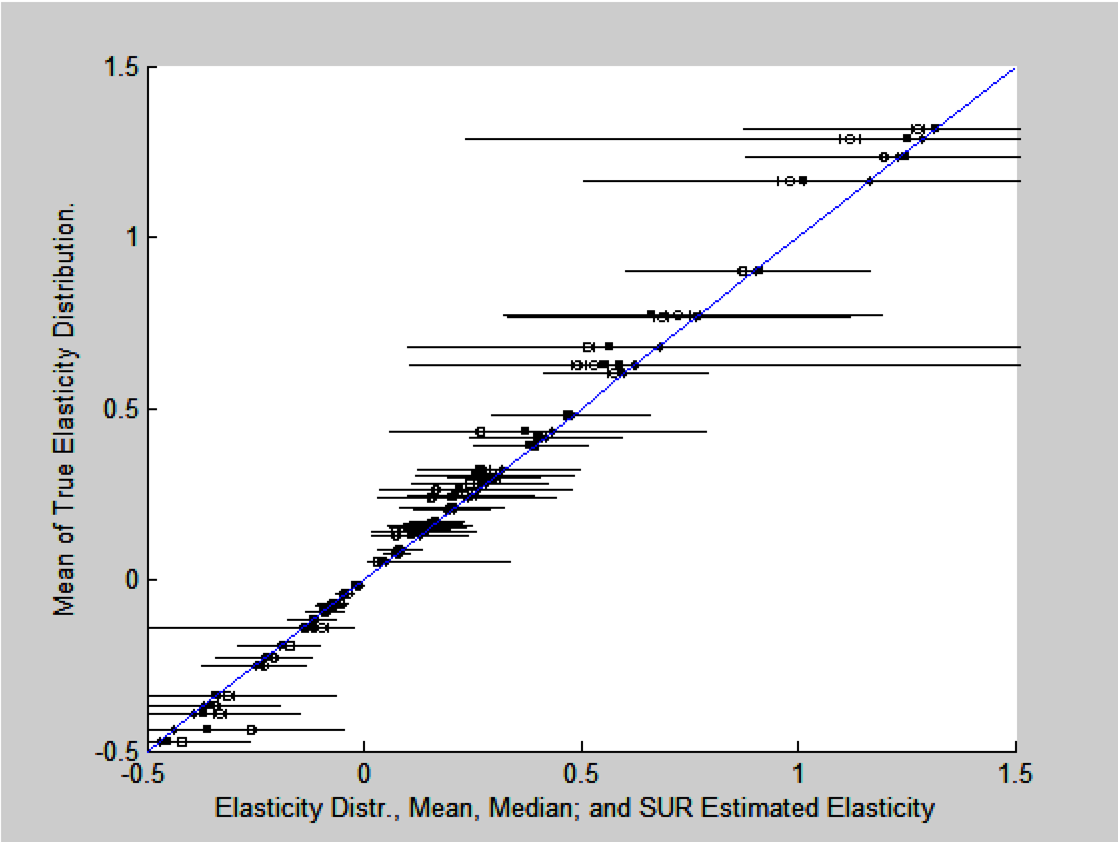
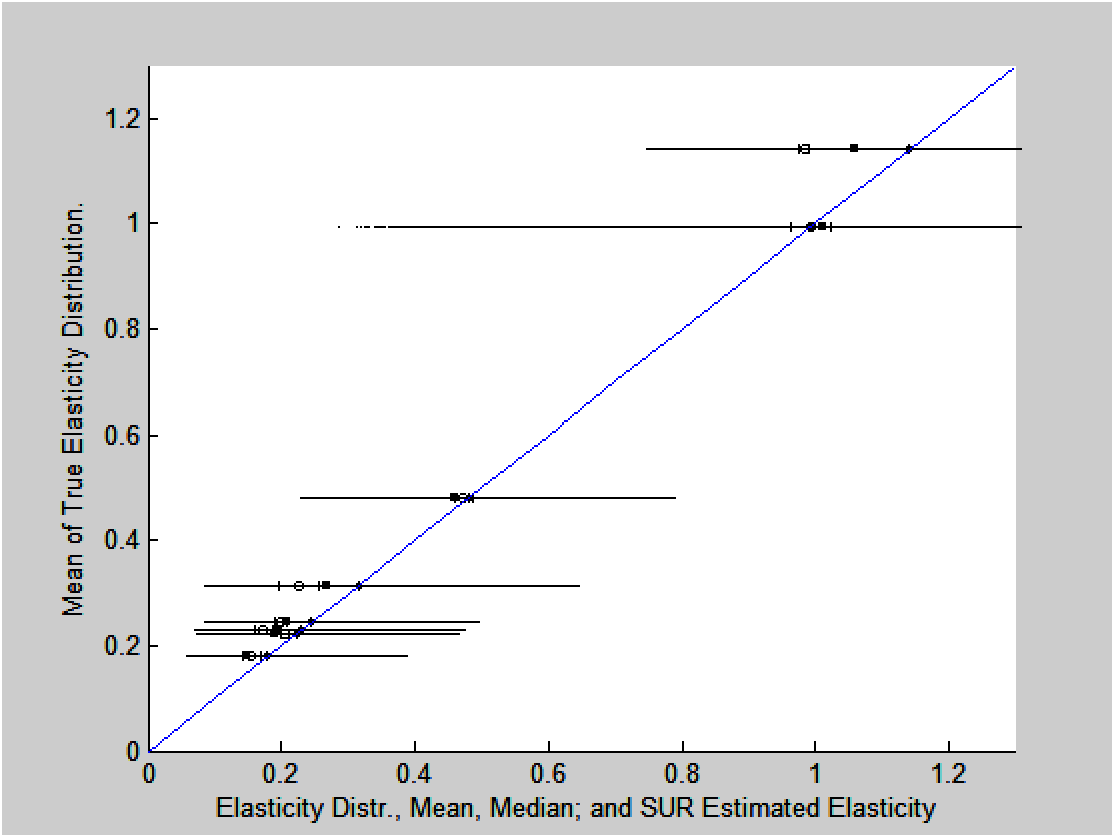


Figure 4. Elasticities of variable netput quantities with respect to quasi-fixed netputs. True versus estimated values.



VII. Conclusions

The dual relationship between the production function and the profit or cost function established by the Neoclassical theory of the firm has been widely applied in empirical work with the objective of obtaining price elasticities, substitution elasticities, and return to scale estimates. This empirical method, usually referred to as "duality theory approach" has the advantage of providing the mentioned features of the production function using market data on input and output prices and quantities, without the requirement of explicitly specifying the technology relationships. However, the duality theorem requires a set of assumptions, which we claim fail to hold in practice; or in other words, market data typically employed in this type of studies bears levels of noise that prevent the theorem from holding exactly. If this is the case, the elasticity estimates will be biased with respect to their true values.

Table 3. Comparison of estimated elasticities versus moments of the distribution of true elasticities.

Elasticities with respect to		**Moment of True Distribution**	
		Median	Mean
Variable Netputs Prices	RMSE	0.048	0.111
	Std. error	0.078	0.196
	% deviation	12.4	26.3
Quasi-fixed Netputs Quantities	RMSE	0.035	0.071
	Std. error	0.044	0.089
	% deviation	7.5	14.7

In this paper we analyze the ability of the approach to recover the technology features using simulated data. We start by selecting a parametric form of the production technology and choosing its set of parameter values. Using Monte Carlo simulations, we generate observations of netput prices and quantities such that they are comparable to those found in data on U.S. agriculture. More precisely, we generate a panel of production and price data for successive periods of time, coming from a population of technologically heterogeneous firms that belong to different regions. We calibrate model parameters using datasets (both time-series and cross-sectional) widely employed in empirical applications.

Estimated parameters (and resulting elasticities) come from applying econometric methods to a system of input demands and output supplies with the simulated data. Because the true parameters are known from the outset, we can judge the degree with which the dual approach is able to recover these parameters. Comparison between true and recovered parameters relies on the use of Hessian identities.

Three are the main objectives of this study. The first one is to describe in detail the procedures used to Monte Carlo simulate a panel of observations featuring important characteristics of U.S. agriculture. The second is to document the consequences of pooling together data from firms with heterogeneous technology and proceed to estimation as if it belonged to a representative firm. Finally, to test the ability of the dual approach to recover production parameters when the only source of noise is firm's heterogeneity.

Together, these objectives allow us to set a solid base to develop studies to further analyze the empirical properties of duality theory, especially when the datasets used by practitioners in empirical work realistically bear more sources of noise, such as uncertainty, prediction errors, omitted variables, netput aggregation, and endogeneity.

Results show that the dual approach applied to a time-series dataset bearing the minimum noise possible, i.e., only the aggregation across technologically heterogeneous firms, is able to recover elasticities that not only are within the support of the distribution of true elasticities, but also considerably close to the mean and median of such distribution. We also conclude that this approach applied to aggregated (county-, state-, or country-level) data as if it belonged to a representative firm optimizing for the entire region, will recover technology features that are close to the firm in the median of the distribution, provided that the data do not contain other sources of noise.

VIII. References

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

Arnade, C. and D. Kelch. 2007. Estimation of area elasticities from a standard profit function. *American Journal of Agricultural Economics*, 89(3), 727-737.

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

Greene, W. H. 2003. Econometric Analysis. 5th Edition. Prentice Hall. Upper Saddle River, N.J.

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

Huffman, W. E. and R. E. Evenson. 1989. "Supply and Demand Functions for Multiproduct U.S. Cash Grain Farms: Biases Caused by Research and Other Policies" *American Journal of Agricultural Economics* 71(3):761-773

Iman, R. L. and W. J. Conover. 1982. "A Distribution-free Approach to Inducing Rank Correlation among Input Variables" *Communications in Statistics - Simulation and Computation* 11(3): 311-334

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

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

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

McElroy, M. B. 1987. "General Error Models for Production, Cost, and Derived Demand or Share Systems" *The Journal of Political Economy* 95(4):737-757

PRISM Climate Group. 2011. Oregon State University. Available at: http://www.prism.oregonstate.edu/

Schuring, J., W. E. Huffman, and X. Fan. 2011. "The Impact of Public and Private R&D on Farmers' Production Decisions: Econometric Evidence for Midwestern States, 1960-2004." *Department of Economics Working Paper Series*. WP #10021. Iowa State University, Ames IA, U.S.

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, Economic Research Service (USDA-ARMS). "Agricultural Resource Management Survey" Available at: http://www.ers.usda.gov/data-products/arms-farm-financial-and-crop-production-practices.aspx

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.

IX. Appendix

Random generation of the firm-specific sets $\boldsymbol{a}_f^*$

We start from the "generic" set $\boldsymbol{a}$ which is composed by the vectors A_1, A_2, A_{11}, A_{12}, A_{21}, A_{22}, and matrix A_{11}, where

$$A_1 = [0.007 \quad 0.007 \quad 0.004 \quad -0.059 \quad -0.314 \quad -0.028 \quad -0.017 \quad -0.081]'$$

$$A_2 = -70.620$$

$$A_{11} = \begin{bmatrix} 1.569 & 1.062 & 0.027 & 2.245 & 1.986 & 1.378 & 0.742 & 3.120 \\ 1.062 & 1.899 & 0.051 & 3.025 & 2.840 & 1.985 & 1.143 & 4.435 \\ 0.027 & 0.051 & 3.154 & 4.059 & 1.030 & 1.530 & 3.356 & 1.646 \\ 2.245 & 3.025 & 4.059 & 16.789 & 7.045 & 5.359 & 7.293 & 10.778 \\ 1.986 & 2.840 & 1.030 & 7.045 & 17.592 & 4.210 & 2.169 & 9.662 \\ 1.378 & 1.985 & 1.530 & 5.359 & 4.210 & 5.624 & 3.481 & 6.311 \\ 0.742 & 1.143 & 3.356 & 7.293 & 2.169 & 3.481 & 7.016 & 3.539 \\ 3.120 & 4.435 & 1.646 & 10.778 & 9.662 & 6.311 & 3.539 & 19.629 \end{bmatrix}$$

$$A_{12} = (A_{21})' = [5.493 \quad 8.967 \quad 8.221 \quad 35.195 \quad 36.758 \quad 19.172 \quad 20.705 \quad 38.203]$$

$$A_{22} = 150.640$$

These values are based on profit function estimated parameters $\boldsymbol{B}_{ij}$ found in literature using Eldon Ball's dataset (Schuring, Huffman, and Fan 2011). We transform the original estimates to meet the desired convexity properties and convert them to production function parameters using Hessian identities. This provides us with a first approximation of the parameters size.

Vectors $A_{1,r}$, $A_{2,r}$, $A_{1,f}$ *and* $A_{2,f}$: For r = {1,2,3}, we obtain the regional vectors $A_{1,r}$ and $A_{2,r}$, by respectively affecting each entry of A_1 and A_2 by independent multiplicative shocks $\nu_a \sim$ Uniform [0.60, 1.40]. Then the farm-specific vectors within each region $A_{1,f}$ and $A_{2,f}$, are obtained from each regional value also by inducing variation with correlated and multiplicative shocks μ_a distributed beta. These shocks determine the size, dispersion and skewness of the netput quantities produced, so they need to be calibrated accordingly. To control for the skewness, we use the county-level variable "Total sales, Value of sales, number of farms" of the 2002 U.S. Agricultural Census as a proxy of firm's size, to fit a standard beta distribution for each region; the results are: a Beta(0.3062, 2.5654) for region 1; a Beta(0.2810, 2.4012) for region 2; and a Beta(0.3315, 2.1364) for region 3. To obtain the desired variability of the firm-specific parameters we modify the beta distributions interval widths to [0.90, 1.40] [0.90, 2.00], and [0.90, 4.20] for regions 1 through 3 respectively, so they match the coefficient of variation of the technology parameters estimated by the fixed-effects regression using USDA-ARMS and PRISM datasets and described in the text. Because parameters determine firm size, we impose a positive correlation of 0.9 between the shocks, so that firms producing high output quantities also use more inputs. In all cases, correlation is imposed by the method in Iman and Conover (1982).

Matrices $A_{11,r}$ and $A_{11,f}$: We generate the inverse of the regional and firm-specific matrices $A_{11,r}$ and $A_{11,f}$, because the latter is the one entering the FOCs of the firm's problem. First, we perturb each entry of an upper triangular matrix C representing the Cholesky factorization of the "generic" positive-semidefinite matrix $(A_{11})^{-1}$, such that $\left(A_{11,r}\right)^{-1} = C_r' C_r$. This guarantees the matrices of interest are positive-semidefinite in each iteration. The regional deviations come from using an independent and multiplicative shock denoted as ν_b and distributed Uniform [0.70, 1.30]. Then, to obtain the firm-specific submatrices $\left(A_{11,f}\right)^{-1}$ in each region we induce variation on the Cholesky factors of $\left(A_{11,r}\right)^{-1}$ with correlated and multiplicative beta shocks μ_b with shape parameters mentioned in the previous paragraph, but over the intervals [0.90, 1.20], [0.90, 1.60] and [0.80, 2.60] for regions 1, 2 and 3 respectively. Again, we set the interval width so that the coefficient of variation of the parameters matches that from the fixed effects regression for each region. Also, we impose positive correlation among the parameters of the matrix to control for firm size.

Vectors $A_{12,f}$ and $A_{22,f}$: Similar to the case of matrices $A_{11,r}$ and $A_{11,f}$, we construct these vectors, as well as the "generic" vectors A_{12} and A_{22}, starting from the "generic" profit function parameters B_{12} and B_{22}, and using Hessian identities in (4). This is done not only to guarantee theoretically consistent values of the vectors of interest, but also because profit function parameters are readily available in the literature. We respectively shock each entry of B_{ij} by independent multiplicative deviations $\nu_c \sim$ Uniform [0.95, 1.05], obtaining regional $B_{ij,r}$. The corresponding firm-specific values ($B_{ij,f}$) within a region come from deviations of the regional $B_{11,r}$, $B_{12,r}$ and $B_{22,r}$ by means of multiplicative and correlated shocks beta μ_c, and then transformed into $A_{12,f}$ and $A_{22,f}$ using the Hessian identities in (4). Note that in this process we do not directly generate regional vectors $A_{12,r}$ and $A_{22,r}$. The Beta distribution shape parameters are the ones stated above and the intervals for $B_{12,f}$ are set at [0.90, 2.00], [0.90, 2.20], and [0.80, 3.60] for each region, and at [0.90, 1.10] for all regions in the case of $B_{22,f}$. The narrow interval in the latter case is due to the fact that enough variation is already induced on $A_{22,f}$ by $B_{11,f}$, $B_{12,f}$ and $B_{22,f}$ through the Hessian relationship. Finally, we impose positive correlation between the entries of $A_{12,f}$ and $A_{22,f}$ to take care of firm size.

We calibrate the width of the beta intervals enumerated above by trial and error such that they yield a set of firm-specific production parameters $\boldsymbol{a}_f^*$ in each region whose coefficient of variation is consistent with $\check{b}_{0c}$ estimated with the fixed-effects model. These are 0.06, 0.17, and 0.43 for regions 1 through 3 respectively.