How Reliable is Duality Theory in Empirical Work?

Francisco Rosas †
Universidad ORT Uruguay
& Centro de Investigaciones Económicas (CINVE)
Montevideo, Uruguay
Tel. (598) 2707 1806 ext. 2233
frosas@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 2016 Agricultural & Applied Economics Association Annual Meeting, Boston, Massachusetts, July 31-August 2

(†) Corresponding author

How Reliable is Duality Theory in Empirical Work?

Francisco Rosas †
Universidad ORT Uruguay
& Centro de Investigaciones Económicas (CINVE)
Montevideo, Uruguay
Tel. (598) 2707 1806 ext. 2233
frosas@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, such as elasticities and returns to scale, without the requirement of explicitly specifying the technology. We generate a pseudo-dataset by Monte Carlo simulations, which starting from known production parameters, yield a dataset with the main characteristics of U.S. agriculture in terms of unobserved firm heterogeneity, decisions under uncertainty, unexpected production and price shocks, endogenous prices, output and input aggregation, measurement error in variables, and omitted variables. Econometric estimation conducted with the mentioned pseudo-data show that the initial production parameters are not precisely recovered and therefore the elasticities are inaccurately estimated. The deviation of the own- and cross-price elasticities from their true values, given our parameter calibration, ranges between 6% and 229%, with an average of 71%. Also, own-price elasticities are as imprecisely recovered as cross-price elasticities. Sensitivity analysis shows that results still hold for different sources and levels of noise, as well as sample size used in estimation.

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

JEL Codes: Q12, D22, D81, C18

(†) Corresponding author

1. Introduction

The Duality theorem applied to the Neoclassical theory of production has provided practitioners with a useful method to provide quantitative answers to important economic questions. Provided certain regularities hold, such as perfect competition, profit maximizing behavior, and certainty, the solution of the primal problem (i.e. the optimal input demands and output supplies arising from the maximization of profits given prices and the production function) are the same as those arising from the dual problem, i.e., the application of Hotelling's Lemma (Shephard's Lemma) to the profit function (cost function) to derive the optimal input demands and output supplies. In other words, the Duality theorem implies an explicit algebraic relationship between the value function (profit or cost function) of the firm's optimization problem and its underlying production function. Therefore, both could be used to empirically estimate price or substitution elasticities, returns to scale, and welfare impacts.

Employing the dual problem begins with approximating the profit (or cost) value function with a parametric functional form, and applying Hotelling's (or Shephard's) Lemma to obtain a parametric form for the optimal input demands and output supplies. Then, system parameters are econometrically estimated using market data (prices and quantities), and finally, using them to recover the technology features of interest (elasticities, return to scale, etc.). According to Shumway (1995), attractive features of the dual approach include the facts that (a) no system of first-order equations has to be solved to obtain input (output) demands (supplies), (b) more functional forms can be used, (c) it is less prone to computational errors, (d) it requires data that are usually easier to obtain, (e) it is more accurate and more tractable for multi-output technologies. On the other hand, highlights that curvature properties are advised to be pre-tested, and collinearity of prices and allocatable inputs induces estimation inefficiency.

The reliance of the approach on a set of assumptions prompted a literature seeking to evaluate its performance in empirical applications. Burgess (1975) and Appelbaum (1978) are among the earliest. These authors failed to identify the source of the discrepancy between conclusions from the primal and dual approaches because they used a functional form that is not self-dual (translog), and used real-world data (for which the mentioned assumptions do not necessarily hold and do not allow to know the true data generating process (DGP)) . As a result, when the primal and dual approaches led to conflicting results, the authors could not establish which approach was preferable, and what portion of the whole divergence in the estimated parameters was attributable to a failure of duality versus to the functional specification.

An exception is the study by Lusk et al. (2002), who analyzed the empirical properties of duality theory by simulating various datasets representing scenarios of price variability, length of time series, and measurement error. They found that small sources of measurement error translate into large errors in estimated parameters, emphasizing the necessity of high-quality data to estimate empirical models.

Considerable effort has been put into testing the most appropriate flexible functional form (FFF) for a given dataset (Guilkey, Lovell and Sickles, 1983; Dixon, Garcia and Anderson, 1987; Thompson and Langworthy, 1989), induced by the fact that results are driven by the specified functional form. 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 observations are simulated corresponding to the “true” production DGP 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 a FFF, and the resulting system of input demands and output supplies is derived. Fourth, econometric methods are applied to estimate the parameters of the resulting system, which are finally compared with the true, known production parameters. However, as these authors assumed perfect competition, profit maximization, certainty, and lack of measurement errors, deviations from duality theory only come from the choice of functional form. As a result, we cannot judge the performance of duality theory in empirical applications, because data used by practitioners are usually not free from at least some of these problems.

In this paper, we propose to analyze the ability of the duality theory approach to recover underlying production parameters from data with commonly observed problems. Among other realistic properties, the simulated data include (i) optimization under uncertainty; (ii) prediction errors in prices and quantities of variable netputs; (iii) omitted variable netputs; (iv) output and input aggregation; (v) measurement errors in the observed variables; (vi) unobserved heterogeneity across firms; and (vii) endogenous output and input prices. For meaningful analysis, we calibrate the simulated data to capture realistic magnitudes of the noise arising from each source. Knowing the true technology parameters, Monte Carlo simulations are used to compute the necessary price and quantity variables. While calibrated to represent typical datasets encountered in practice, the levels of noise embedded in these variables affect the data used in estimation, preventing duality theory from holding exactly. Hence, the true production parameters may not be recovered with enough precision, and the estimated elasticities measurements may be more inaccurate than expected.

We first generate a panel of input and output prices and quantities for successive periods of time and coming from a set of firms with heterogeneous technology. As this DGP does not bear the problems described in the previous paragraph, we employ it to confirm that the dual approach is able to recover the production parameters with sufficient accuracy provided its basic assumptions are met.

Second, we add noise to the generated panel of price and quantity variables to replicate the aforementioned real-world problems found in data used by practitioners. We aim at generating noise comparable to that 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-ARMS database, the U.S. Agricultural Census database (USDA-NASS), and the Chicago Mercantile Exchange (CME) futures 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 two data sources provide useful information to calibrate 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., measurement error, endogeneity of output prices, production and price shocks). Moreover, we adopt the criteria of calibrating parameter values to favor recovery of true production parameters, especially for those that are unobservable.[1]

[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-

We set up the expected profit function and derive the system of input demands and output supplies, to then econometrically estimate its parameters for comparison with the true (and known) production parameters. Comparisons are performed using Lau's (1976) Hessian identities between production and restricted profit functions.

2. The model of a single firm

Consider a producer who maximizes the expected utility of uncertain terminal wealth by choosing the level of netputs.[2] The firm's problem[3] is:

$$\begin{aligned}\max_{[\boldsymbol{y},y_0]}\left\{EU\left(\tilde{W}_1\right)\right\} &= \max_{[\boldsymbol{y},y_0]}\left\{EU\left(W_0+\tilde{\pi}\right)\right\} \\ &= \max_{[\boldsymbol{y},y_0]}\left\{EU\left(W_0+\tilde{\boldsymbol{p}}'\tilde{\boldsymbol{y}}+\tilde{y}_0\right)\right\}\end{aligned} \tag{2.1}$$

where *U* is a strictly increasing and twice-continuously differentiable concave utility function of terminal wealth ($\tilde{W}_1$), defined as initial wealth (W_0) plus uncertain end-of-period profits ($\tilde{\pi}$). The tilde ($\sim$) indicates a random variable. Variable $\tilde{\boldsymbol{y}}$ is a vector of *n* variable netput quantities, $\tilde{\boldsymbol{p}}$ are the corresponding variable netput prices normalized by p_0 which is the price of y_0, the numeraire commodity. The expectation operator *E* integrates over the uncertainty of random variables $\tilde{\boldsymbol{p}}$ and $\tilde{\boldsymbol{y}}$.

Defining $\boldsymbol{K}$ as the vector of *m* quasi-fixed netputs, a production plan consists of the vector $\left[y_0,\boldsymbol{y}',\boldsymbol{K}'\right]$ belonging to the production possibilities set $S \in R^{1+n+m}$.[4] As shown

sectional data is as relevant as the one pursued here. 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 these datasets are not as readily available.

[2] According to netput notation, a positive value is a net output and a negative value is a net input.

[3] The model setup follows closely the one used in Rosas and Lence (2015), which in turn is based on Lau (1976)

[4] 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).

by Jorgenson and Lau (1974), there exists a one-to-one correspondence between the set *S* and a production function *G* (also constrained by the quasi-fixed netputs ***K***), such that:[5]

$$G(\boldsymbol{y}, \boldsymbol{K}) = -\max\left\{y_0 / \left[y_0, \boldsymbol{y}', \boldsymbol{K}'\right] \in S\right\} \tag{2.2}$$

Therefore, we can rewrite problem (6) as follows:

$$\max_{[y]}\left\{EU\left(\tilde{W}_1\right)\right\} = \max_{[y]}\left\{EU\left(W_0 + \tilde{\boldsymbol{p}}'\tilde{\boldsymbol{y}} - G\left(\tilde{\boldsymbol{y}}, \boldsymbol{K}; \boldsymbol{\alpha}\right)\right)\right\} \tag{2.3}$$

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

The duality theory establishes a relationship between the production function $G\left(\tilde{\boldsymbol{y}}, \boldsymbol{K}; \boldsymbol{\alpha}\right)$ and the restricted profit function $\pi_R(\boldsymbol{p}, \boldsymbol{K}; \boldsymbol{\beta})$, which Lau (1976) proved in terms of their Hessian matrices under the assumption of convexity and twice continuously differentiability of both functions. In our analysis, these Hessian relationships are a key result because are used to write the estimated parameters of the restricted profit function ($\boldsymbol{\beta}$) as a function of the parameters of the underlying production function ($\boldsymbol{\alpha}$), and then used to compare the recovered parameters with the true ones. See Rosas and Lence (2015) for more details on how this comparison is performed.

To operationalize this problem, we proceed by assigning functional forms in problem (2.3). We assume a constant absolute risk aversion (CARA) utility function of

[5] 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 {***y***, ***K***}. Convexity is required because of the convention used in Lau (1976) that $y_0 = -G(\boldsymbol{y}, \boldsymbol{K})$. We follow the convention that the value of the production function is positive infinity if a production plan is not feasible, that is, $max\{\emptyset\} = -\infty$, where {Ø} is the empty set.

the form $U(\tilde{\pi}) = -e^{-\lambda \tilde{W}_1}$ with λ representing the coefficient of absolute risk aversion (which determines the degree of concavity of the utility function), defined as $\lambda = U' / U''$ where U' and U'' are the first and second derivatives, respectively, of the utility function with respect to the random terminal wealth. The treatment of risk and uncertainty in the duality theory framework with profit functions includes the work by Pope (1982), Coyle (1992), Coyle (1999), Pope and Just (2002). In the case of cost functions, developments are due to Pope and Chavas (1994), Pope and Just (1996), Pope and Just (1998), Chambers and Quiggin (1998), Moschini (2001), and Chavas (2008), among others. Then, we assume a quadratic FFF for the production function $G\left(\boldsymbol{y}_{ft}, \boldsymbol{K}_{ft}; \boldsymbol{\alpha}_f\right)$:

$$G(\bullet) = \boldsymbol{y}'_{ft} A_{1f} + \boldsymbol{K}'_{ft} \boldsymbol{A}_{2f} + \tfrac{1}{2} \boldsymbol{y}'_{ft} A_{11f} \boldsymbol{y}_{ft} + \boldsymbol{y}'_{ft} A_{12f} \boldsymbol{K}_{ft} + \boldsymbol{K}'_{ft} A_{22f} \boldsymbol{K}_{ft} - \psi_{ft} \qquad (2.4)$$

where *f* and *t*, respectively, index firms and time, A_{1f} and A_{2f} are (n×1) and (m×1) vectors of $\alpha_{i,f}$ coefficients, A_{11f} is a symmetric and nonsingular (n×n) matrix, and A_{12f} and A_{22f} are (n×m) and (m×m) matrices of firm *f*. Submatrices A_{11f}, A_{12f} and A_{22f} form a symmetric and positive semi-definite ((n+m)×(n+m)) matrix A_f of $\alpha_{ij,f}$ coefficients.[6] We collectively denote all $\alpha_{i,f}$ and $\alpha_{ij,f}$ coefficients as $\boldsymbol{\alpha}_f$. This functional form is self-dual, the production and profit function Hessians are only functions of parameters, and is broadly employed by practitioners in applications of the duality approach.

Uncertainty in a farmer's decision process comes from events such as random weather, pests, and selling prices not known with certainty at the time of making

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

allocation decisions, among others. In particular, the farmer optimizes by choosing the quantity of expected output at the end of the growing season. We model production uncertainty by introducing a mean-zero, heteroskedastic production shock denoted by ψ_{ft} for each firm *f* and time *t*. The functional form is as follows:

$$\begin{aligned} \psi_{ft} &= g\left(\boldsymbol{y}_{ft}, \boldsymbol{D}\right)\boldsymbol{v}_{ft} \\ &= \tfrac{2}{8}\left[\boldsymbol{D}\cdot\left(\boldsymbol{y}'_{ft}\right)^{3/2}\right]\boldsymbol{v}_{ft} \end{aligned} \tag{2.5}$$

where $\boldsymbol{D}$ is a $(1\times n)$ row vector of constants, "·" is the dot product, and $\boldsymbol{v}$ is an $(n\times1)$ random vector. The entries of $\boldsymbol{v}$ corresponding to variable outputs are distributed as $v \sim -1+2\times Beta(2,2)$, or an independent and identically distributed (*iid*) symmetric shock with mean zero in the interval [-1,1]. Elements corresponding to variable inputs are zero. While this is consistent with firms facing output quantity uncertainty, the jointly specified technology induces uncertainty in all choice variable netputs. There are at least two reasons for choosing functional form in (2.5). First, it guarantees a heteroskedastic production error with a standard deviation increasing at a decreasing rate, consistent with the assumption that bigger firms are less exposed to uncertain events (weather) because a bad draw is more likely to be offset by a good draw within the same firm. Second, the multiplicative constants, the beta distribution parameters, and the random error jointly induce a production shock ranging from plus or minus 10% to 60% of the average quantity produced.[7] The shock enters the solution of variable netput quantities in its first derivative and premultiplied by $\left(A_{11}\right)^{-1}$. To achieve the desired level of variability in each

[7] For comparison, a pooled panel of farm-specific corn yield over a period of five years shows that the 2.5th and 97.5th percentiles are respectively 60% lower and 40% higher than the average yields in the Corn Belt region, 60% lower and 42% higher in the Lake States region, and 80% lower and 70% higher in the Northern Plains region.

netput quantity, and to reduce variability induced by other netputs (especially in the case of inputs), entries of ***D*** are set equal to the inverse of the main diagonal of $(A_{11})^{-1}$.

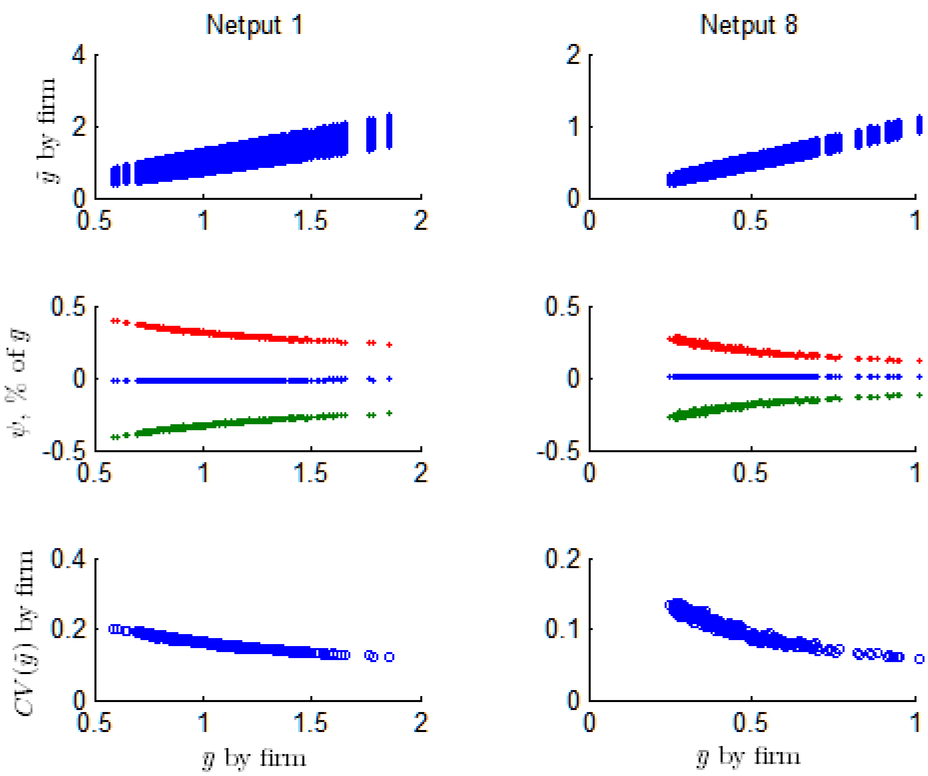

Figure 2.1. Production shock as a function of firm's average variable netput quantity ($\bar{y}_{ft_0}$) at time t = t_0, for selected netputs.

Note: Top panels: distribution of netput quantities ($\tilde{\boldsymbol{y}}_{ft_0}$) faced by each firm. Middle panels: minimum, mean, and maximum shock as percentage of firm's average quantity $\bar{\boldsymbol{y}}_{ft_0}$. Bottom panels: coefficient of variation of the distribution of quantities CV($\tilde{\boldsymbol{y}}_{ft_0}$) by firm.

Figure 2.1 shows selected production shocks computed for netput 1 (output) and netput 8 (input) for all firms f at time $t_0 = 1$. In the top panels, the distribution of the netput quantity faced by each firm $\tilde{y}_{ft_0}$ is plotted against the firm's average netput quantity $\bar{y}_{ft_0}$. The middle panels show the minimum (green), mean (blue), and maximum (red) of the production shock ($\tilde{\psi}$) as a percentage of the firm's average netput quantity.

For firms with higher levels of outputs or inputs, the minimum and maximum shocks represent a lower percentage of the average quantity, ranging between 10% and 50% depending on the netput, yielding a coefficient of variation decreasing in netput quantity, which is consistent with the desired production shock heteroskedasticity (bottom panels).

Firms also face end-of-period output price uncertainty, modeled as a log-normal deviation from the firm-specific price $\boldsymbol{p}_{ft}^{*}$, or:

$$\log\left(\tilde{\boldsymbol{p}}_{ft}\right) = \log\left(\boldsymbol{p}_{ft}^{*}\right) + \boldsymbol{e} \tag{2.6}$$

where $\boldsymbol{e}$ is an $(n \times 1)$ random vector. Entries associated with prices of outputs are *iid* normally distributed shocks with mean zero and standard deviation of 0.2 (Lence 2009). Entries corresponding to inputs are zero, assuming input prices are known at the decision moment.

We induce correlation between the levels of output prices and quantities by the Iman and Conover (1982) method. We assume production shocks have an impact on prices of the opposite sign and set the correlation coefficient to -0.30 based on observed correlations of these variables for the U.S. Further, because commodity prices tend to move together, we impose a strong positive correlation of 0.90 among commodity prices. Similarly, we assume output quantity shocks are positively correlated among them because weather is likely to affect all crops; therefore, we set the correlation coefficient to 0.90.

3. Simulation of panel data

To analyze the empirical properties of duality theory, we generate a noiseless and a noisy dataset. Consistent with Rosas and Lence (2015), the noiseless dataset is used both to illustrate the ability of duality to recover true production parameters when data are free from the aforementioned problems, and to show the implications on parameters recovery when data are aggregated across firms with heterogeneous technology. In this study, we focus on the noisy dataset which allows us to document the effects on production parameter estimation from using duality theory when the dataset features realistic problems.

The simulation produces a panel of 8 variable netput prices and quantities over a period of T = 50 years, from R = 3 regions each composed of F = 10,000 heterogeneous firms, such that firm heterogeneity is higher across regions than within it. Therefore, conditional on the set of parameters $\boldsymbol{a}_f^*$, there are $R \times F \times T$ = 1.5 million observations for each variable of the vector $\left[\boldsymbol{y}_{ft}, \boldsymbol{p}_{ft}, \boldsymbol{K}_{ft}; \boldsymbol{a}_f^*\right]$.[8]

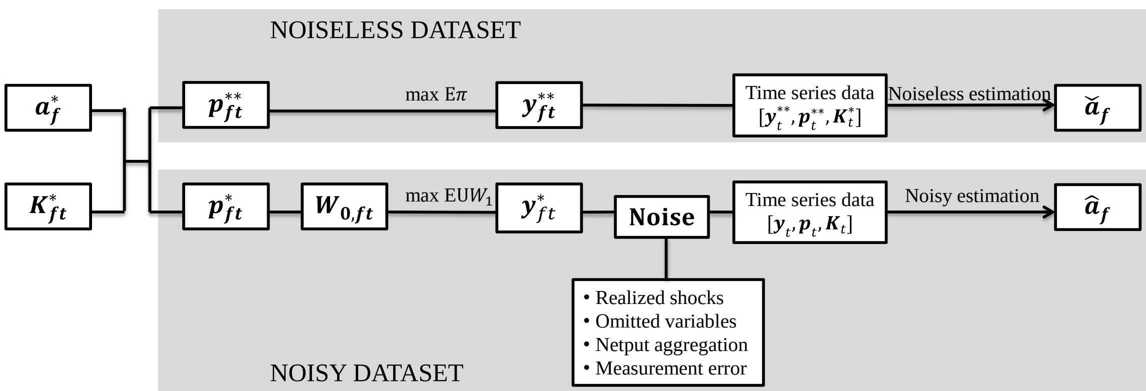

Figure 3.1. DGP of noiseless and noisy datasets used in estimation.

Figure 3.1 sketches the simulation of the data. The first step is to Monte Carlo simulate for each firm and time period the set of starting production parameters $\boldsymbol{a}_f^*$ and the quasi-fixed netputs $\boldsymbol{K}_{ft}^*$ (section 3.1). Second, conditioning on these values, we draw

[8] This figure roughly represents 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.

expected variable netput prices that are exogenous in the case of the noiseless dataset $\boldsymbol{p}_{ft}^{**}$, and endogenous in the noisy data $\boldsymbol{p}_{ft}^{*}$ (section 3.2). For the latter, calibrated values of initial wealth $W_{0,ft}$ (section 3.3) and the coefficient of absolute risk aversion λ_f are provided for the maximization problem's objective function. Third, we solve an expected profit maximization problem to obtain the expected variable netput quantities $\boldsymbol{y}_{ft}^{**}$ (section 3.4). The noiseless panel dataset is then aggregated across heterogeneous firms before proceeding to estimation of the time-series dataset. We finally recover production function parameters, which we denote as $\breve{\boldsymbol{\alpha}}_f$ (section 5).

In the case of the noisy data (section 3.5), we assume a risk-averse individual who chooses optimal expected netput quantities $\boldsymbol{y}_{ft}^{*}$ so as to maximize the expected utility of end-of-period terminal wealth (subsection 3.5.1). Before proceeding to estimation, we disturb the data with the following sources of noise: shocks in production and expected output and price (section 3.5.2), omitted variables (section 3.5.3), aggregation across netputs (section 3.5.4), and measurement errors in price and quantity variables (section 3.5.5). Finally, the variables are aggregated over unobserved heterogeneous firms to conduct time-series estimation (section 3.5.6). The expected netput quantity and prices are denoted as $\boldsymbol{y}_{ft}$ and $\boldsymbol{p}_{ft}$, respectively. Estimation results yield the set of production parameters denoted as $\hat{\boldsymbol{\alpha}}_f$ (section 5).

3.1 Random generation of true production parameters: $\boldsymbol{a}_f^{*}$ and $\boldsymbol{K}_{ft}^{*}$

The production function of each firm $G\left(\boldsymbol{y}_{ft},\boldsymbol{K}_{ft};\boldsymbol{\alpha}_f\right)$ is conditioned by the parameter set $\boldsymbol{a}_f^{*}$ and the set of quasi-fixed netputs $\boldsymbol{K}_{ft}^{*}$**,** which together determine the technology of each firm, and ultimately drive the values of elasticities, returns to scale, and other firm measurements. As Figure 3.1 states, we simulate by Monte Carlo the values for $\boldsymbol{K}_{ft}^{*}$, vectors A_{1f} and A_{2f}, and submatrices A_{11f}, A_{12f} and A_{22f} in equation (3.1), which are used in the random generation of the netputs prices and quantities pseudo-data. We use the methods and procedures described in Rosas and Lence (2015), which for reasons of space are not explained here. It has to be noted that to favor identification of parameters, technology is allowed to change only across firms and not over time. The recovery of technological change parameters can be left as a topic for future research without compromising the conclusions arrived in this study.

3.2 Random generation of expected variable netput prices: $\boldsymbol{p}_{ft}^{**}$ and $\boldsymbol{p}_{ft}^{*}$

Throughout the analysis, we assume producers solve the maximization prices based on expected output prices and current input prices. We generate two sets of firm-specific expected prices for each region.

Prices are exogenous in the first case ($\boldsymbol{p}_{ft}^{**}$), and are used to test duality theory with noiseless data. We start by simulating "national" netput prices calibrated to match the mean, standard deviation, and serial autocorrelation of CME future crop prices and Eldon Ball's input prices, as AR(1) lognormally distributed processes. The details are explained in Rosas and Lence (2015).

In the second case, prices are endogenous with respect to the aggregated netput quantity ($\boldsymbol{p}_{ft}^{*}$), and are used in the evaluation of duality theory properties in empirical work when using more realistic (noisy) data, which is the focus of this study. In a competitive market it is realistic to assume each firm is a price taker, because the netput quantity decisions of any single firm do not affect price levels. This is usually modeled as the firm facing exogenous and fixed netput prices (i.e., a perfectly horizontal demand for outputs and supply for inputs). For an aggregation of firms, this is not necessarily the case. On aggregate, firms face downward sloping demand curves for their outputs and upward sloping supplies for inputs. In this case, changes in netput quantities at the aggregate level result in market-level price changes.

We introduce a system of isoelastic market demands and supplies faced by firms in period *t*, described by $\boldsymbol{Q}_t = \Phi_t \boldsymbol{p}_t^{\eta}$. The *n*-dimensional vector $\boldsymbol{Q}_t$ is the aggregate market demand of output or the aggregate market supply of input n faced by firms; $\boldsymbol{p}_t^{\eta}$ denotes an *n*-dimensional vector of $\boldsymbol{p}_{nt}$ netput market prices, each raised to the power of η_n (the calibrated netput-specific demand or supply own price elasticity); and Φ_t is an $(n \times n)$ diagonal matrix of supply and demand netput-specific shocks ϕ_{nt} coming from the market. All Greek letters represent calibrated parameters.

The objective is to find a vector of netput prices $\boldsymbol{p}_{ft}^*$, where the optimal vector of netput quantities aggregated across firms ($y_t^* = \sum_f y_{ft}^*$) equals the vector of market quantities ($\boldsymbol{Q}_t$). As will become apparent when we set up the firm's maximization problem, we can write the optimal quantity of variable netputs as follows:

$$y_t^* = \sum_f \left(\left(\boldsymbol{X}_f \right) \left(\boldsymbol{p}_{ft} \right) + \boldsymbol{\varphi}_{ft} \right) \tag{3.1}$$

where $\boldsymbol{X}_f$ is a time-invariant matrix of production coefficients summarizing the elements of $\boldsymbol{a}_f$, $\boldsymbol{p}_{ft}$ is the vector of firm-specific prices received (defined as $\boldsymbol{p}_{ft} = \boldsymbol{p}_t \varepsilon_{ft}$ and explained below), and $\boldsymbol{\varphi}_{ft}$ is a vector of production errors, such as optimization mistakes, weather shocks, deviation of prices from expected values, etc. These errors depend on the firm's production parameters due to the claimed heteroskedasticity given by function $g(.)$ in (2.5). By substituting for the firm-specific production coefficients and prices we have:

$$y_t^* = \sum_f \left(\boldsymbol{X}\mu_f\right)\left(\boldsymbol{p}_t\varepsilon_{ft}\right) + g_0\left(\boldsymbol{X}\mu_f\,\boldsymbol{p}_t\varepsilon_{ft}\right)v_{ft} \tag{3.2}$$

where $\boldsymbol{X}$ is the analog of the set of production coefficients $\boldsymbol{a}$, $\boldsymbol{p}_t$ is the vector of "national" prices, and $g_0(.)$ is the analog of function $g(.)$. With a sufficiently large number of farms (F) and by independency of the random variables μ_f, ε_{ft} and v_{ft} (the *iid* shocks in ψ_{ft} and φ_{ft}), y_t^* converges in distribution by the law of large numbers to a Normal random variable whose mean is:

$$\overline{\boldsymbol{y}}_t = F\boldsymbol{X}\boldsymbol{p}_t + F\overline{\boldsymbol{\varphi}}_t \tag{3.3}$$

The expression in (3.3) depends only on the known "average" production parameters and "national" time-t prices $\boldsymbol{p}_t$, which in fact are the same as those on the isoelastic demand or supply function faced by firms.

Therefore, the vector of time-t netput prices is the $\boldsymbol{p}_t^*$ which clears the market ($\boldsymbol{Q}_t = \overline{\boldsymbol{y}}_t$), or in other words, the one which implicitly solves the following system for each t:

$$\boldsymbol{\Phi}_t\boldsymbol{p}_t^{\eta} = F\boldsymbol{X}\boldsymbol{p}_t + F\overline{\boldsymbol{\varphi}}_t \tag{3.4}$$

The system in (3.4) is nonlinear in $\boldsymbol{p}_t$ and is conditional on known values—the set of known production parameters $\boldsymbol{X}$ and time-specific systematic shocks $\boldsymbol{\Phi}_t$. We obtain the desired vector of "national" netput prices $\boldsymbol{p}_t^*$ by numerically solving this system for each time t, given a random market shock $\boldsymbol{\Phi}_t$.

This requires generating values of $\mathbf{\Phi}_t$. We model the systematic shocks coming from the market $\mathbf{\Phi}_t$ as auto-correlated and behaving according to a Log-normal distribution:

$$\log\left(\phi_{nt}\right) = \rho_{n0} + \rho_{n1}\log(\phi_{n,t-1}) + \xi_n \,, \tag{3.5}$$

where $\xi_n \sim Normal(0, \sigma^2_{\xi_n})$. For each netput *n*, the three parameters ρ_{n0}, ρ_{n1}, and ξ_n are calibrated such that the resulting vector of average prices $\boldsymbol{p}^*_t$ have volatility values comparable to those of Eldon Ball's dataset. Appendix I provides calibration details.

The systematic shocks coming from the market, ϕ_{nt}, are generated for each *t* and *n* as follows. Dropping the "*n*" subscript again to ease notation, in the long run, the logarithms of ϕ_t and ϕ_{t-1} converge to $\bar{\phi}$ and therefore we can calculate the long-run expected shocks as $\log\left(\bar{\phi}\right) = \rho_0 / \left(1-\rho_1\right)$. To generate the desired systematic Log-normal shocks, we set $\log\left(\phi_{s=0}\right) = \rho_0 / \left(1-\rho_1\right)$, take a draw from a $Normal(0, \sigma^2_{\xi_n})$ random variable, and use (3.5) to calculate the systematic shock for the first iteration, i.e. $\log(\phi_{s=0})$. This procedure is repeated 10,000 times.

Next, we plug the vector ϕ_s for each iteration into (3.4) and use the MATLAB function *fsolve* to solve this system 10,000 times for the vector of *n* "national" prices that simultaneously clears the *n* markets in each *t*.[9] We keep the last 50 solutions which constitute the vector of endogenous "national" netput prices $\boldsymbol{p}^*_t$.

[9] 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. Random draws from $Normal(0, \sigma_{\xi_n})$ are independent from each other, and therefore systematic shock are as well; however, when plugged into system (3.4) correlation between national prices is induced through matrix $\boldsymbol{X}$. This DGP ultimately generates national netput prices with higher temporal variance and with lower correlation

Firm-specific netput prices, endogenous ($\boldsymbol{p}^{*}_{ft}$) and exogenous ($\boldsymbol{p}^{**}_{ft}$). These are both generated as deviations from "national" price. With the endogenous case as an example, first, a regional average is 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 + 0.10*Beta*(2,2)], and independent from d_r, to allow non-constant deviation from national average prices over time. Then, firm-specific random prices are in turn generated as deviations from the regional average. These deviations are small relative to variability of $\boldsymbol{p}^{*}_{ft}$ (and $\boldsymbol{p}^{**}_{ft}$) to acknowledge for the contemporaneous low variability of prices firms receive and pay. That is, $\boldsymbol{p}^{*}_{ft} = \boldsymbol{p}^{*}_{r} \varepsilon_{ft}$, where ε_{ft} is a symmetric mean one shock distributed as $\varepsilon_{ft} \sim$ [0.80 + 0.40*Beta*(2,2)]. Shocks ε_{rt}, ε_{ft}, and d_r are independent. The calibration of shock ε_{ft} implies firm-specific prices with a coefficient of variation of 0.08, which doubles that of firm prices in USDA-ARMS dataset.

In order to favor identification of parameters in estimation, the following assumptions in calibration were made. Netputs prices were assumed to be independent from firm size. Also, a continuum of firm's prices were generated, while the observed frequency of firm's prices in USDA-ARMS dataset is concentrated in about four values.

3.3 Random generation of initial wealth: $W_{0,ft}$

In the noisy dataset, each firm f at time t is assumed to be an expected utility maximizer. The argument of the utility function is end-of-period terminal wealth calculated as initial

---

between netput prices than Eldon Ball's dataset. These two aspects favor identification in estimation when prices are explanatory variables, as it is our case.

[10] The same procedure and shocks are used for p^{**}_{ft} .

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

wealth ($W_{0,ft}$) plus random profits. Based on the strong correlation observed between total net assets (TNA_{ft}) and value of production (VP_{ft}) in the USDA-ARMS database, we model initial wealth as a function of each firm's value of production.[12] The following model is used for each firm *f* at time *t*:

$$TNA = \gamma_0 + \gamma_1 VP + \gamma_2 VP^2 + \tau \tag{3.6}$$

where τ is an heteroskedastic error term distributed $Normal(0, \sigma_\tau^2)$ which accounts for the non-constant variation observed in total net assets as a function of value of production. We seek to estimate the γ parameters as well as the form of the heteroskedasticity. Following Wooldridge (2003), we model heteroskedasticity as follows $\hat{u}^2 = \exp(\delta_0 + \delta_1 VP + \delta_2 VP^2)\kappa$, where $\hat{u}^2$ is the estimated variance of τ and κ is a mean one multiplicative error term. This implies $\log(\hat{u}^2) = \delta_0 + \delta_1 VP + \delta_2 VP^2 + e$ or a linear regression for which e ~ $Normal(0, \sigma_e^2)$.

We estimate model parameters with USDA-ARMS data. Details are explained in Appendix II and table II.1 shows estimation results. Using these parameter estimates and the value of production coming from our model, we generate the initial wealth for each firm *f* in time *t*.

3.4 Simulation of noiseless dataset

The noiseless dataset is formed by variable netput quantities and prices, and quasi-fixed netputs: $[\boldsymbol{y}_{ft}^{**}, \boldsymbol{p}_{ft}^{**}, \boldsymbol{K}_{ft}^{*}]$. We first solve the problem in (2.3) assuming all farmers are risk

[12] Total net assets are calculated as "value of total farm financial assets" minus "total farm financial debt." Value of production is calculated as "all crops – value of production" plus "all livestock – value of production." These two variables from USDA-ARMS database constitute the dataset used to estimate the model in equation (3.6).

neutral (or expected profit maximizers) and prices received or paid are exogenous ($\boldsymbol{p}_{ft}^{**}$).

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 $\boldsymbol{G}\left(\boldsymbol{y}_{ft}^{**}, \boldsymbol{K}_{ft}^{*}; \boldsymbol{a}_f\right)$ in (2.4), the FOCs are:

$$\boldsymbol{p}_{ft}^{**} - A_{1f} - A_{11}\boldsymbol{y}_{ft}^{**} - A_{12f}\boldsymbol{K}_{ft}^{*} = 0 \tag{3.7}$$

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:

$$\boldsymbol{y}_{ft}^{**}\left(\boldsymbol{p}_{ft}^{**}, \boldsymbol{K}_{ft}^{*}; \boldsymbol{a}_f^{*}\right) = A_{11f}^{-1}\left(\boldsymbol{p}_{ft}^{**} - A_{1f} - A_{12f}\boldsymbol{K}_{ft}^{*}\right) \tag{3.8}$$

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

$$[\boldsymbol{y}_{ft}^{**}, \boldsymbol{p}_{ft}^{**}, \boldsymbol{K}_{ft}^{*}] \tag{3.9}$$

3.5 Simulation of noisy dataset

In this section, we explain how we generate data to mimic the features faced by practitioners when working with real-world data. It contains variable netput quantities and prices, and quasi-fixed netputs: $[\boldsymbol{y}_{ft}, \boldsymbol{p}_{ft}, \boldsymbol{K}_{ft}^{*}]$. We assume risk-averse firms that maximize expected utility, and the data are subject to omitted variables, aggregation across netputs, measurement error, and aggregation across heterogeneous firms.

3.5.1 Maximization of expected utility

We solve the problem in (2.4) for the vector of expected variable netput quantities $\boldsymbol{y}^{*}_{ft}(\boldsymbol{p}^{*}_{ft}, \boldsymbol{K}^{*}_{ft}, \lambda_f, W_{0,ft}; \boldsymbol{a}^{*}_{f})$ conditional on expected netput prices, quasi-fixed netput quantities, the level of absolute risk aversion (λ_f), initial wealth, and the true production parameters. Values of λ_f are consistent with a relative risk aversion coefficient uniformly distributed in the interval [2.0, 4.0] (Pennacchi, 2008 pp. 16). This constitutes a source of noise because duality theorem assumes a deterministic problem whose solution is generally different from the expected utility case.

We solve the problem using numerical methods and employing Gaussian quadrature. Using four nodes for each output price and quantity random variable, we are guaranteed to exactly approximate the problem's objective function up to the seventh moment. For this application, the numerical integration of the objective function has to take into account its multi-dimensions, that nodes behave according to nonstandard distributions, and that they are correlated with each other. Given these problem requirements, we create a routine to calculate nodes and weights used in the objective function approximation. First, based on the MATLAB functions *qnwnorm* (Miranda and Fackler, 2011), that calculates Standard Normal nodes and weights, we generate four independent Log-normally distributed nodes and weights for each of the three output price random variables. Similarly, based on the function *qnwbeta*, that calculates standard Beta nodes and weights, we generate four independent nodes and weights distributed Beta in the interval of interest for the three output quantity random variables. Second, using the Iman and Conover (1982) method, we impose directly to the nodes, negative correlation between output prices and quantities (correlation coefficient equal to -0.30) and positive correlation within them (coefficient of 0.90). These transformations do not affect the weights. Third, we use the MATLAB function *fmincon* to optimize the approximated objective function. We pass the FOCs and SOCs to the optimization routine, respectively, as equality and inequality constraints. Based on the normalized quadratic production function in (2.4), the FOCs are:

$$E\left[U'\left(\tilde{W}_1\right)\left(\tilde{\boldsymbol{p}}^* - A_1 - A_{11}\tilde{\boldsymbol{y}} - A_{12}\boldsymbol{K} + g'\left(\boldsymbol{y}\right)\cdot\boldsymbol{v}\right)\right] = 0 \qquad (3.10)$$

with $U'\left(\tilde{W}_1\right) = \lambda e^{-\lambda\tilde{W}_1}$. The SOCs are:

$$E\left[U''\left(\tilde{W}_1\right)\left(\tilde{\boldsymbol{p}}^* - A_1 - A_{11}\tilde{\boldsymbol{y}} - A_{12}\boldsymbol{K} + g'\left(\boldsymbol{y}\right)\cdot\boldsymbol{v}\right)^2 + U'\left(\tilde{W}_1\right)\left(-A_{11} + g''\left(\boldsymbol{y}\right)\cdot\boldsymbol{v}\right)\right] \le 0$$

(3.11)

where $U''\left(\tilde{W}_1\right) = -\lambda^2 e^{-\lambda\tilde{W}_1}$.

The optimal solution is the vector of expected netput quantities for each farm and time that we donate as $\boldsymbol{y}_{ft}^*$.

3.5.2 Realized shocks of production and prices

Farmers solve the maximization problem given a set of output prices, which reflect their expectations of harvest prices. It is commonly accepted that prediction errors make this difference relevant. Even in the case of locking in the production with instruments such as forward contracts, it might be the case that not all of the production is sold under this type of arrangement. In the case of input prices, some prices might not be known at the beginning of the production period, especially for inputs purchased during the growing season. In either case, deviations from the true expected price induce bias and inconsistency in the parameter estimates of a model when prices are regressors, and consequently in the elasticities of interest.

We obtain the observed price $\boldsymbol{p}_{ft}$ by drawing from the distribution in equation (2.6) and adding that draw to the price used in optimization:

$$\log\left(\boldsymbol{p}_{ft}\right) = \log\left(\boldsymbol{p}^{*}_{ft}\right) + \boldsymbol{e}_{t} \tag{3.12}$$

where $\boldsymbol{e}_t$ is a realization of a $Normal(0, \sigma_e = 0.2)$ random variable for the case of outputs (Lence, 2009) and a $Normal(0, \sigma_e = 0.1)$ for the inputs, implying that the price deviation from decision values is lower for inputs. These shocks are systematic and affect all firms by the same proportion in a given time.

We also claim that data on netput quantities are different from planned quantities or the optimal solution of (2.4) due to uncertain events in agricultural production, such as weather. We model the observed quantities as follows:

$$\boldsymbol{y}_{ft} = \boldsymbol{y}^{*}_{ft} + g\left(\boldsymbol{y}^{*}_{ft}\right)\boldsymbol{v}_{ft} \tag{3.13}$$

where the shock $\boldsymbol{v}_{ft}$ is a realization of the random variable controlling production errors (ψ_{ft}) given by (2.5). We assume this shock has two components; one is systematic given by $\boldsymbol{v}_{ft} \sim -1 + 2Beta(2,2)$, and the other is idiosyncratic modeled as $\boldsymbol{v}_{ft} = \boldsymbol{v}_{t} \times Uniform[0.87, 1.13]$. This allows weather variables to not only affect production quantities over time, but also to have different local effects in a given year. To calibrate the width of the interval, we run a fixed-effects model of farm-level yields at various locations (counties) and time periods (years) on a location-specific effect, weather variables (temperature and cumulative precipitation), and time dummies. After estimation, we measure the contribution of weather variables to yield variation by fitting a “restricted” model with only the weather and time-dummy variables using the estimated parameters. The coefficient of variation of the fitted yields provides the dispersion of weather shocks $\boldsymbol{v}_{ft}$. Details of this estimation are provided in the appendix of Rosas and Lence (2015).

Finally, we introduce contemporaneous negative correlation between quantity and price shocks with a coefficient equal to -0.3 (Rosas, Babcock, Hayes, 2015), and positive correlation with a coefficient of 0.9 within quantities and within prices.

3.5.3 Omitted variable netputs

Production takes place with several netputs, but the econometrician rarely observes them all. 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. In either case, while the producer optimally chooses a set of *n* variable netputs to maximize profits, but the econometrician only observes a subset of them.

3.5.4 Aggregation across netputs

Technology processes employ a variety of inputs to produce several outputs; however, data available to practitioners are usually not as disaggregated. In some cases, even if data can be obtained for several inputs and outputs, they are aggregated because they are not the objective of the study and/or to not excessively penalize the degrees of freedom during estimation. We aggregate netput quantities and prices at the firm and for each time. Dropping the *f* and *t* subscripts, this is done as follows:

$$\begin{aligned} y_{n^i} &= \sum\nolimits_{n^{ij} \in \Omega_i} w_{n^{ij}} y_{n^{ij}} \\ p_{n^i} &= \sum\nolimits_{n^{ij} \in \Omega_i} w_{n^{ij}} p_{n^{ij}} \end{aligned} \tag{3.14}$$

where Ω_i is a subset *i* of netputs, n^{ij} is the j^{th} netput in subset Ω_i, and n^i denotes a new netput formed by the aggregation of those in subset Ω_i. The pooling of variable netput quantities (y_{n^i}) in set Ω_i is performed by a weighted average of the quantities in the set, with weights given by each netput value ($p_{n^{ij}} y_{n^{ij}}$) share on the total value of the set Ω_i;

that is $w_{n^{ij}} = \left(p_{n^{ij}} y_{n^{ij}} \right) \left(\sum_{n^{ij} \in \Omega_i} p_{n^{ij}} y_{n^{ij}} \right)^{-1}$. The procedure to aggregate netput prices is analogous.

3.5.5 Measurement error in prices and quantities

Measurement error is a common problem in datasets available to researchers and induces bias and inconsistency in parameter estimation. Efforts to quantify the level of errors in the data include Morgenstern (1963), who identifies a 10% standard error in the national income data, and reports that the U.S. Department of Commerce in the state-level Food and Kindred Products data have an 8% measurement error in input and output figures. Lusk et al. (2002) study the consequences of applying duality theory using variables measured with error, and Lim and Shumway (1992a, 1992b) analyze violations of maintained hypotheses such as profit maximization, convex technology, and regressive technical change. Based on the mentioned literature, we calibrate an error with a standard deviation of 0.05 around the "true" value for netput prices, 0.08 for variable netput quantities, and 0.10 for quasi-fixed netputs. Calibration values are less than or equal to those reported in the literature, especially in the case of prices. The error is distributed as standard *Beta*(2,2) and we modify its interval to yield the desired standard deviation.

Added noise described in previous subsections implies the following panel for firm *f* and time *t* which we can use to recover production parameters applying either time-series or cross-section analysis:

$$\left[\boldsymbol{y}_{ft}, \boldsymbol{p}_{ft}, \boldsymbol{K}_{ft}; \boldsymbol{a}_f^* \right] \tag{3.15}$$

3.5.6 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 both the noiseless data described in (3.9) and the noisy data in (3.15). 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 $\left[\boldsymbol{y}_{ft}, \boldsymbol{p}_{ft}, \boldsymbol{K}_{ft}\right]$ across firms to obtain observations over T=50 time periods (years) of a "single firm" yielding the dataset $\left[\boldsymbol{y}_t, \boldsymbol{p}_t, \boldsymbol{K}_t\right]$. 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.

$$\begin{aligned} \boldsymbol{y}_t &= \sum\nolimits_f \boldsymbol{y}_{ft} \\ \boldsymbol{K}_t &= \sum\nolimits_f \boldsymbol{K}_{ft} \\ p_{nt} &= \left(y_{nt}\right)^{-1} \sum\nolimits_f p_{nft} y_{nft} \end{aligned} \tag{3.16}$$

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

$$\left[\boldsymbol{y}_t^{**}, \boldsymbol{p}_t^{**}, \boldsymbol{K}_t^{*}\right] \tag{3.17}$$

and the noisy dataset used in estimation is the following:

$$\left[\boldsymbol{y}_t, \boldsymbol{p}_t, \boldsymbol{K}_t\right] \tag{3.18}$$

4. Data for estimation

Noiseless data in (3.17) include 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 500,000 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 (5.2). 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 consequences of pursuing estimation incorporating data from other more heterogeneous regions to capture a broader area and increase the sample size, which are common in these applications, are shown as a sensitivity analysis.

In the case of noisy data in (3.18), n’ = 4 netputs are included due to the omission of one input and one output, the pooling of two variable outputs into one, and the pooling of two variable inputs into one. There is also one quasi-fixed netput because we did not consider the case of omitting the quasi-fixed netput. The figure below represents the structure of the noisy data.

Variable netput	1 2	~~3~~	4	5	6 7	~~8~~

Noisy data are also subject to endogeneity between netput prices and quantities, maximization under expected utility of risk-averse farmers, production and price unexpected shocks, and measurement error. We take the region’s population of 10,000 heterogeneous firms and draw 100 samples of 6,000 observations. We aggregate over the heterogeneous firms resulting in a time-series dataset of 50 observations for each variable and, for each sample, conduct econometric estimation of the system in (5.2). We sample from the population to avoid final results to be dependent on a single sample.[13,14] For the same reasons stated above, we select region 1 to conduct the estimation. The case of pooling observations from heterogeneous regions as a way of increasing the number of observations in estimation, and its effects on parameters recovery, is presented as sensitivity analysis.

5. Estimation

We approximate the restricted profit function $\pi_R(\boldsymbol{p},\boldsymbol{K})$, which solves the problem in (2.3), by the following normalized quadratic flexible functional form:

$$\pi_R\left(\boldsymbol{p},\boldsymbol{K};\boldsymbol{\beta}\right) = \boldsymbol{p}'B_1 + \boldsymbol{K}'B_2 + \tfrac{1}{2}\,\boldsymbol{p}'B_{11}\boldsymbol{p} + \boldsymbol{p}'B_{12}\boldsymbol{K} + \boldsymbol{K}'B_{22}\boldsymbol{K} + \boldsymbol{p}'\boldsymbol{\kappa} \qquad (5.1)$$

where B_1 and B_2 are $(n\times1)$ and $(m\times1)$ 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 $((n+m)\times(n+m))$ 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 McElroy's (1987) additive general error model (AGEM) applied to the case of profit functions. The $(n\times1)$ vector of random variables $\boldsymbol{\kappa}$ is jointly normally distributed with mean zero 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

[13] Given that the population size in each region is relatively large, we do not require too many samples to achieve robust results. Also, the sample size within a region (6,000 observations) is sufficiently high compared to real-world datasets used to construct state-level aggregates. For example, the 2004 ARMS dataset consists of samples that range between 48 and 1600 firms depending on the state, with an average of 428 firms.

[14] For comparison, estimation was also conducted using the entire population in the region and aggregating across all the heterogeneous firms, which implies only one time-series dataset to be estimated. Results were very similar to the case of the 100 samples from the population.

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:

$$y(\boldsymbol{p}, \boldsymbol{K}; \boldsymbol{\beta}) = B_1 + B_{11}\boldsymbol{p} + B_{12}\boldsymbol{K} + \boldsymbol{\kappa} \tag{5.2}$$

We conduct estimation by iterated SUR, which converges to maximum likelihood, and is the most common method employed in empirical works 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.

Treating Mean-Independence Violations in Estimation. First, an inspection of the autocorrelation and partial autocorrelation functions of the noiseless and noisy time series suggests first differentiation of the data for estimation. This is a consequence of the DGP of price data as AR(1) processes.

Second, for the case of the noisy dataset only, we employ an instrumental variables approach to treat the omitted price variables. We use the same omitted variables as instruments which are regarded as the best instrument possible.

Third, we also use instruments to account for the endogeneity of the explanatory variables in the case of the noisy dataset. Endogeneity is present because the independent variables (prices) in the output supplies and input demands system are correlated with the error term $\boldsymbol{\kappa}$ as a consequence of the systematic shocks ϕ_{nt} in the market. Instruments have to be correlated with prices but uncorrelated with the error term. Given that we know the source of the endogeneity (i.e., shocks ϕ_{nt}), we construct instruments by regressing each price on its systematic shock: $p_{nt} = \lambda_0 + \lambda_1\phi_{nt} + iv_{nt}$. The error term ($iv_{nt}$) is, by construction, orthogonal to ϕ_{nt} but correlated with p_{nt} accounting for the variation of prices not explained by the systematic shocks, constituting the ideal instrument. There is one instrument for each netput price, as well as one instrument for each omitted variable.

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 [$\hat{B}$] and the elasticities matrix of netput quantities with respect to own price, cross prices, and quasi-fixed netputs [$\hat{E}$].[15] As described in figure 5.1, we obtain matrices [$\breve{B}$] and [$\hat{B}$] from estimation using noiseless and noisy datasets respectively, that are then transformed into elasticity matrices 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 5.1, we compare the true $[E]_f$ versus the estimated values ([$\breve{E}$] and [$\hat{E}$]) to evaluate how precisely we recover the true price and quasi-fixed netput elasticities under duality theory, both in the case of noiseless and noisy data. Note that this comparison implies that the true values are represented by a distribution of each firm's true

[15] In the case of noiseless data they are denoted as [$\breve{B}$] and [$\breve{E}$], respectively.

parameters, whereas the estimated values consist of a point estimate and its confidence interval.

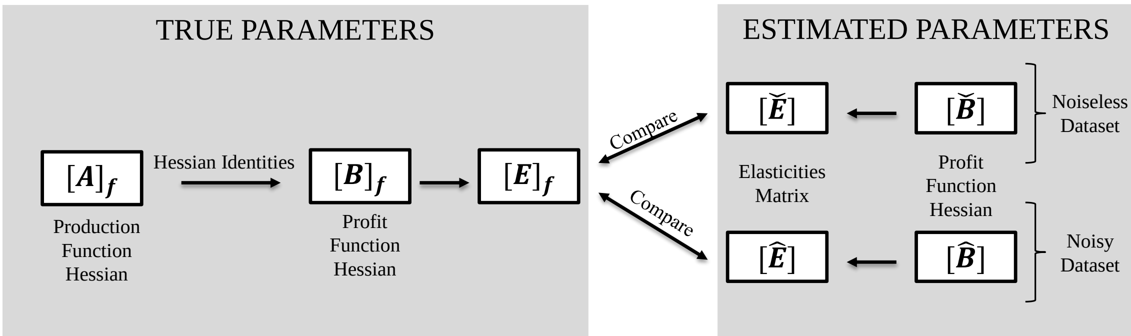

Figure 5.1. Comparison between true and estimated elasticities for noiseless and noisy datasets.

6. Results

Estimation results for the noiseless and noisy data are presented separately.

6.1. Noiseless data estimation

Econometric estimation conducted with the noiseless pseudo-dataset, that is the one arising from the aggregation across heterogeneous firms but without any other source of noise show that the Duality theory is able to recover true production parameters fairly accurately. Precisely, these results, already derived and explained in Rosas and Lence (2015), are that estimated elasticities with respect to prices (quasi-fixed netputs) deviate on average 12.4% (7.5%) from the median of the true elasticities according to the computed root mean squared error (RMSE),[16] which summarizes the average difference

---

[16] When compared to the median of the distribution, the RMSE is $[\frac{1}{64\times S}\sum_i\sum_j\sum_s(\bar{\bar{E}}_{ij,s}-\breve{E}_{ij,s})^2]^{1/2}$ 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. Comparison with the mean can be performed by substituting $\bar{\bar{E}}_{ij}$ by $\bar{E}_{ij}$. The RMSE averages over all the $64\times S$ squared differences. A measure of its dispersion is achieved by computing the standard deviation of these $64\times S$ values before averaging over them.

between each entry of the estimated elasticity matrix and the median of the corresponding true elasticity distribution. The RMSE 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. Given the SUR estimation provides only a minor source of error because the point estimates are all highly significant due to the use of noiseless data, we argue that the majority of the RMSE standard deviation is attributed to the deviations between the estimated and the true value across elasticities.

6.2. Noisy data estimation

Estimation with noisy data consists of 16 own- and cross-price elasticity values of variable netput quantities, and 4 elasticities with respect to quasi-fixed netputs. Figure 6.1 shows the distribution of the true firm-specific price elasticities and its corresponding SUR point estimates indicated with a red circle (and its 95% confidence interval with a red "+" sign). After estimation, we take 10,000 draws from the parameters asymptotic distribution of each of the 100 samples, transform them into elasticities, and calculate their mean, standard deviation, and confidence interval over the 1,000,000 values. Except for the case of entries (2 2), (2 3), (3 2), and (3 3), all other distributions involve more than one true elasticity due to the aggregation of netputs, as described in section 4.6.4. In these cases, and in order to compare with the SUR estimated elasticities, we construct a "new true" elasticity distribution as a revenue weighted average of the original true elasticities.

In light of the conclusions from the previous section, we measure the accuracy in recovering the true elasticity by comparing the estimated value to the median of the true distribution.[17] Figure 6.1 shows that duality theory provides a good approximation to the true distribution in some cases, and a bad one in others, when comparing where the estimated value falls relative to where the true distribution accumulates more mass.

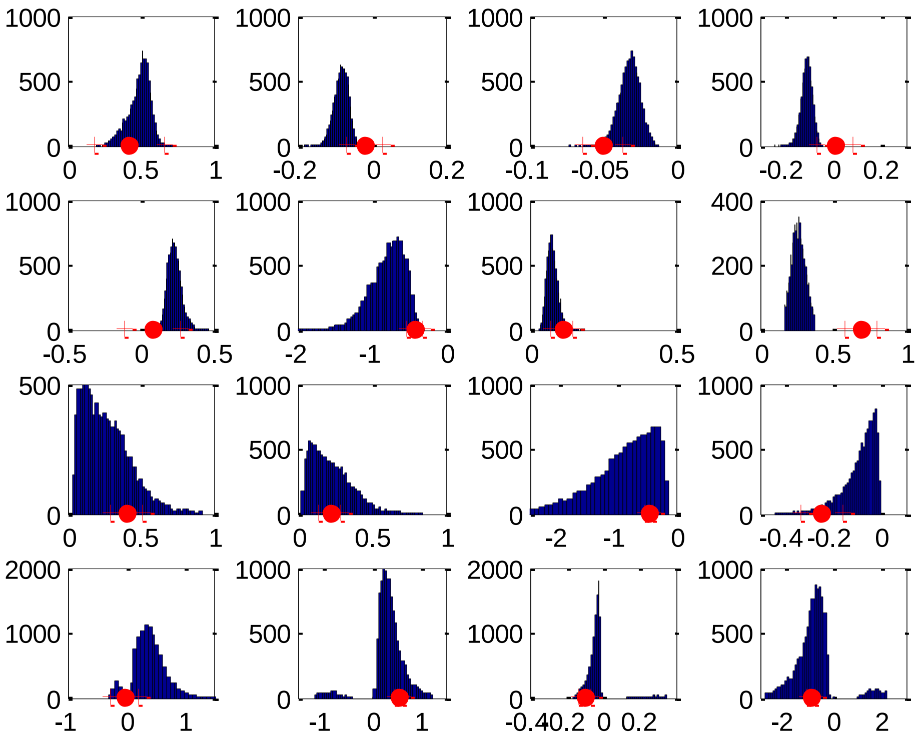

Figure 6.1. Own- and cross-price elasticities of variable netput quantities. True versus estimated values with noisy data.

Note. Each *ij* panel is the *ij* entry of the 4x4 own- and cross-price elasticity matrix *E* in the case of noisy data. The elasticity value is in the horizontal axis and histogram frequency in the vertical. The histograms are the distribution across firms of the true elasticity (E_{ij}). Red dot is the SUR estimated elasticity ($\hat{E}_{ij}$) and red "+" sign is the 95% confidence interval.

However, as table 1 shows, the percentage difference between the median of the true distribution ($\bar{\bar{E}}_{ij}$) and the estimated value ($\hat{E}_{ij}$) is high for the majority of the entries in the elasticity matrix. The difference ranges between 6% and 229%, and is less than 12% in only one entry. The estimation of the own price elasticities (main diagonal) are

[17] A comparison using the mean of the true distribution was conducted and provided less accurate results.

not recovered with sufficient precision given that the differences range between 15% and 44%. Moreover, entries (2 2) and (3 3) which correspond to netputs 4 and 5 (which are not aggregated with other netputs) are more imprecisely estimated than the other main diagonal elements which do arise as aggregated netputs. As expected, the off-diagonal elements (or the cross-price elasticities) are less accurately estimated than the main diagonal entries, because they require more information to be recovered.

As a summary measure of the dispersion in recovering the true elasticities, we calculate the RMSE of the difference between the median of the true distribution and the SUR estimated values for all 16 estimated elasticities. As shown in table 2 under case 1, it yields a value of 0.22 in elasticity units. The average value of all true elasticities (calculated as the mean absolute value of all the medians of the true distributions) is 0.31. Therefore, by comparing both values we conclude that duality theory recovers elasticities which are, on average, off by 71% of the true elasticities. These results provide evidence that the econometric approach of duality is unable to deliver precise estimates of underlying production parameters when employed with data featuring real-world characteristics.

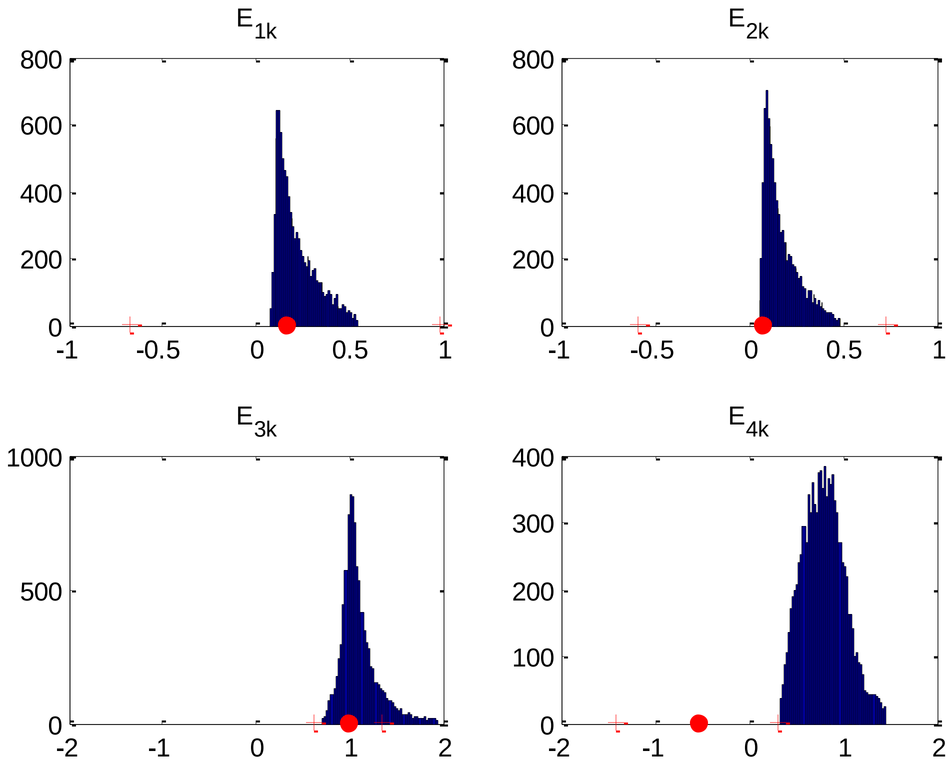

Figure 6.2. Elasticity of variable netput quantities with respect to quasi-fixed netputs. True versus estimated values with noisy data.

Note. Each E_{ik} panel is the elasticity of netput *i* with respect to the quasi-fixed input in the case of noisy data. The elasticity value is on the horizontal axis and histogram frequency on the vertical. The histograms are the distribution across firms of the true elasticity (E_{ik}). The red dot is the SUR estimated elasticity ($\hat{E}_{ik}$) and the red "+" signs denote the 95% confidence interval.

The estimation of variable netput elasticities with respect to quasi-fixed netputs is even more inaccurate. Results are shown in figure 6.2. Each panel titled as E_{ik} is the elasticity of netput *i* with respect to the quasi-fixed netput. The SUR point estimates of the elasticities are within the support of the true distribution except for E_{4k}, in which case the estimated elasticity has the reversed sign. As a similar summary measure, the RMSE relative to the median of the true distribution is 0.67 expressed in elasticity units, and the average value of the elasticities is calculated at 0.54. These results, shown table 2 under case 1, imply that the inaccuracy in recovering the true elasticities, averaged over the 4 netputs, amounts to 123%.

6.3. Sensitivity analysis

We explore the robustness of noisy data estimation results to changes in the sources and levels of noise. Estimations that consider different sets of omitted netputs and different sets of aggregated netputs provide similar results. For example, two estimations with the noisy data structure shown below yields price elasticity estimates with respect to variable netputs that are 57% and 69% deviated from the true price elasticities. Elasticities with respect to quasi-fixed netputs differ 76% and 120% relative to their true values.

Variable netput	~~1~~	2	3	~~4~~	5	6	7	8
Variable netput	1	2	~~3~~	4	5	6	~~7~~	8

These values, shown in table 2 under case 2 and case 3 respectively, are quantitatively similar to our previous conclusions, suggesting that other combinations of omitted netputs and aggregated netputs would provide results that are at least qualitatively similar.[18]

We regularly encounter empirical applications of duality theory with time-series data where observations from different regions or states are pooled together for estimation; examples are Schuring, Huffman and Fan (2011) and O'Donnell, Shumway and Ball (1999). Although pooling increases the sample size favoring the degrees of freedom, which is especially advantageous in the presence of several explanatory variables, it also involves including observations from states that are likely to have different technology. We conduct a sensitivity analysis to study the consequences of such

[18] We present only a few cases due to the computational burden of such analysis

practice. This practice implies seeking to recover production parameters from firms that are more heterogeneous than in the case of a single state usually by adding regional- (or state-) level dummy variables. For this sensitivity analysis we employ our noisy simulated data from regions 1, 2, and 3 in (3.8). From each region and in each of the 50 time-periods, we take five samples of 2,000 observations representing samples of firms from five states within the region. We aggregate each sample across its heterogeneous firms to obtain the time-series data for each state. We stack the observations by state first, and then by region, resulting in a dataset composed by 750 observations. Note that results above are derived with 50 observations. We estimate model in (5.2), and as it is done in the mentioned studies, we add dummy variables for observations in region 1 and 2, leaving region 3 as the base. We transform the estimated parameters into netput elasticities with respect to variable netput prices and with respect to quasi-fixed netputs, and compare them with the true elasticities. These are represented by the distribution of true firm-specific elasticities, as it was in the previous analysis, but in this case, is the distribution over the firms in the three regions.

We find that the RMSE relative to the median of such distribution of true elasticities, and averaged over the 16 elasticities calculated is 53%, which is very similar to the findings above. This is formed as the ratio between the RMSE relative to the median of the true distribution (0.18) and the median of the true distribution of elasticities (0.35). Divergence from true elasticities ranges between 11% and 209% depending on which of the 16 entries of the elasticities matrix we consider. Standard errors of the estimated elasticities are lower than in the previous analysis, which is in part a consequence of in increased number of observations. However, that does not contribute to reduce the bias in the estimated parameters and elasticities relative to the true ones. Results, when RMSE is calculated with respect to the mean of the distribution and results of netput quantity elasticities with respect to quasi-fixed netputs, also indicate inaccuracy in recovering production parameters. Therefore, the practice of incorporating data from other regions, characterized by a more heterogeneous technology than within the region, contributes to reducing the standard error of the point estimates (large statistical significance) but does not help in reducing the bias relative to the true elasticity values.

7. 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 bear 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.

In this paper we analyze the ability of the approach to recover the technology features when the dataset taken to estimation reflects real-world characteristics comparable to those found by practitioners in empirical applications. Based on a model of maximization of expected utility of terminal wealth, we first choose the parametric form of the production function and use Monte Carlo simulations to generate its set of parameters for a number of firms with heterogeneous technology. In particular, from the solution of this problem, we generate a pseudo-dataset of netput prices and quantities for heterogeneous firms, coming from different regions and for successive years, such that their features are comparable to those found in data on U.S. agriculture and typically used by practitioners in empirical applications. In this regard, the DGP incorporates optimization under uncertainty, prediction errors in prices and quantities of variable netputs, endogenous prices, omitted variable netputs, output and input data aggregation, measurement errors in the observed variables, and unobserved heterogeneity across firms. We calibrate model parameters using datasets (both time-series and cross-sectional) widely employed in practice.

We apply the Duality approach to this multi-netput pseudo-dataset, with consists of deriving the system of input demands and output supplies from a profit function approximated by a FFF, and estimate its parameters (and the corresponding elasticities) using traditional econometric methods (more precisely a SUR regression). Because the initial (primal) production parameters are known to us, we can evaluate the ability of this approach to recover these parameters by transforming the estimated parameters from the dual model into the primal parameters, and then comparing them. This transformation is performed by means of the so-called Hessian identities.

Also, because we know the existing sources of noise in the data, we explicitly treat them in estimation. We deal with serial autocorrelation by estimating the model with data in first differences. To tackle omitted variables, we employ an instrumental variables approach in which our instruments are precisely the variables we omit in the first place.

Similarly, we use instruments to consider the presence of endogeneity in aggregate prices. In this case, we also know the source of endogeneity and therefore we can construct the best set of instruments possible.

Results show that the dual approach applied on a time-series dataset bearing the minimum noise possible, i.e., only arising from aggregating firms with heterogeneous technology, is able to recover elasticities within the support of the distribution of initial elasticities, and considerably close to the mean and median of such distributions.

Second, the use of noisy data prevents the dual approach from providing parameter estimates that are sufficiently close to their true values. The root mean squared error, measuring the average deviation of the estimated elasticities from their true values, is calculated at 71%, implying that the dual approach estimates elasticities are, on average, 71% away from the true values. It is the case that, conditional on the dataset, own-price elasticities require less information from the data to be estimated with the same level of precision than cross-price elasticities; however, both own- and cross-price elasticities are inaccurately recovered. The case of netput elasticities with respect to quasi-fixed netputs is even more inaccurate. Results are robust to different calibrations of the data structure, specifically, the omission and aggregation of different sets of netputs, as well as the sample of firms used in estimation. Also, sensitivity analysis shows that the common practice of pooling data from different states and/or regions in order to increase the degrees of freedom in estimation yields a similar bias in the estimated elasticities as in the case of considering a single and more technologically homogeneous state.

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

Chambers R. G. and J. Quiggin. 1998. “Cost Functions and Duality for Stochastic Technologies” *American Journal of Agricultural Economics* 80(2):288-295

Chavas J. –P. 2008. “A Cost Approach to Economic Analysis Under State-Contingent Production Uncertainty” *American Journal of Agricultural Economics* 90(2):435-446

Coyle, B. T. 1992. “Risk Aversion and Price Risk in Duality Models of Production: A Linear Mean-Variance Approach” *American Journal of Agricultural Economics* 74(4):849-859

Coyle, B. T. 1999. “Risk Aversion and Yield Uncertainty in Duality Models of Production: A Mean-Variance Approach” *American Journal of Agricultural Economics* 81(3):553-567

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

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

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

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

Lim, H. and C. R. Shumway. 1992a. “Separability in State-Level Agricultural Technology” *American Journal of Agricultural Economics* 74(1):120-131

Lim, H. and C. R. Shumway. 1992b. “Profit Maximization, Returns to Scale, and Measurement Error” *The Review of Economics and Statistics* 74(3):430-438

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

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

Miranda M. J. and P. L. Fackler. 2002. “Applied Computational Economics and Finance.” The MIT Press. Cambridge, Massachusetts.

Morgenstern, O. 1963. “On the Accuracy of Economic Observations”, Second Edition Princeton University Press

Moschini GC. 2001. “Production Risk and the Estimation of Ex-ante Cost Functions.” *Journal of Econometrics* 100:357-380

O’Donnell, C. J., C. R. Shumway, and V. E. Ball. 1999. “Input Demands and Inefficiency in U.S. Agriculture” American Journal of Agricultural Economics 81(4): 865-880

Pennacchi, G. 2008. “Theory of Asset Pricing.” Pearson Addison Wesley. Boston, Massachusetts.

Pope, R. D. 1982. “Expected Profits, Price Change, and Risk Aversion” *American Journal of Agricultural Economics* 64(3):581-584

Pope, R. D. and J. -P. Chavas. 1994. “Cost Functions Under Production Uncertainty” *American Journal of Agricultural Economics* 76(2):196-204

Pope, R. D. and R. E. Just. 1996. “Empirical Implementation of ex ante Cost Functions” *Journal of Econometrics* 72:231-249

Pope, R. D. and R. E. Just. 1998. “Cost Function Estimation Under Risk Aversion” *American Journal of Agricultural Economics* 80(2):296-302

Pope, R. D. and R. E. Just. 2002. "Random Profits and Duality" *American Journal of Agricultural Economics* 84(1):1-7

Rosas, J.F. and S.H. Lence. 2015. Duality theory econometrics: How reliable is it with real-world data?. In *2015 AAEA & WAEA Joint Annual Meeting, July 26-28, San Francisco, California* (No. 205294). Agricultural and Applied Economics Association & Western Agricultural Economics Association.

Rosas, F., B. A. Babcock, D. J. Hayes. 2015. "Nitrous Oxide Emission Reductions from Cutting Excessive Nitrogen Fertilizer Applications." *Climatic Change* 132 (2): 353-367.

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-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, 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).

Wooldridge, J. M. "Introductory Econometrics. A Modern Approach." Second Edition. Thomson South-Western. 2003

Table 1. Comparison of estimated elasticities ($\hat{E}_{ij}$) versus median of true elasticities distribution ($\bar{\bar{E}}_{ij}$) in the case of noisy data.

		1	2	3	4
1	$\bar{\bar{E}}_{ij}$	0.505	-0.081	-0.031	-0.106
	$\hat{E}_{ij}$	0.424	-0.019	-0.050	0.009
	Std. Dev.	(0.125)	(0.024)	(0.007)	(0.038)
	Interval	0.178 0.670	-0.066 0.027	-0.064 -0.036	-0.066 0.084
	% diff.	16%	76%	60%	108%
2	$\bar{\bar{E}}_{ij}$	0.230	-0.757	-0.076	0.263
	$\hat{E}_{ij}$	0.078	-0.423	0.112	0.688
	Std. Dev.	(0.097)	(0.056)	(0.019)	(0.055)
	Interval	-0.111 0.268	-0.532 -0.312	0.074 0.150	0.580 0.796
	% diff.	66%	44%	247%	162%
3	$\bar{\bar{E}}_{ij}$	0.244	0.206	-0.810	-0.075
	$\hat{E}_{ij}$	0.399	0.217	-0.457	-0.246
	Std. Dev.	(0.057)	(0.038)	(0.035)	(0.045)
	Interval	0.288 0.510	0.144 0.291	-0.525 -0.389	-0.334 -0.158
	% diff.	64%	6%	44%	229%
4	$\bar{\bar{E}}_{ij}$	0.397	0.332	-0.035	-0.782
	$\hat{E}_{ij}$	-0.028	0.547	-0.101	-0.895
	Std. Dev.	(0.124)	(0.044)	(0.018)	(0.069)
	Interval	-0.271 0.215	0.461 0.633	-0.136 -0.065	-1.031 -0.759
	% diff.	107%	65%	190%	15%

Note: Interval is the 95% confidence interval of the point estimate $\hat{E}_{ij}$

Table 2. Sensitivity analysis. Comparison of estimated elasticities ($\hat{E}_{ij}$) versus median of true elasticities distribution ($\bar{\bar{E}}_{ij}$) in the case of noisy data, and different sources of noise.

Elasticities with respect to		RMSE relative to Median		
		case 1	case 2	case 3
Variable Netput Prices	RMSE	0.22	0.26	0.20
	Median	0.31	0.46	0.29
	% deviation	71%	57%	69%
Quasi-fixed Netput Quantity	RMSE	0.67	0.44	0.42
	Median	0.54	0.57	0.35
	% deviation	123%	76%	120%

Note: Each case consists of a different set of omitted netputs and a different set of netputs aggregated together. In case 1, netputs 3 and 8 are omitted, and netputs 1 and 2, and 6 and 7 are aggregated. In case 2, netputs 1 and 4 are omitted, and netputs 2 and 3, and 7 and 8 are aggregated. In case 3, netputs 3 and 7 are omitted, and netputs 1 and 2, and 5 and 6 are aggregated.

APPENDIX I

Random generation of market shocks for each netput n, used to endogenize netput prices:

The vector of market elasticities of the isoelastic demand or supply ($\boldsymbol{Q}_t = \boldsymbol{\Phi}_t \boldsymbol{p}_t^{\eta}$) faced by firms is the following (FAPRI Elasticities Database and other sources): $\eta =$ [-0.25 -0.21 -0.75 0.90 0.87 0.85 0.83 0.80]. The calibrated parameter values of market shocks (ϕ_{nt}) in equation (3.5) used to generate random endogenous "national" prices are in the table below.

Table I.1. Calibrated parameter values of market shocks (ϕ_{nt}) in equation (3.5)

	***n* = 1**	***n* = 2**	***n* = 3**	***n* = 4**	***n* = 5**	***n* = 6**	***n* = 7**	***n* = 8**
ρ_{n1}	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000
$\log(\bar{\phi})$	10.2193	10.5644	10.3588	9.528	8.5391	8.9875	8.9013	9.2518
$\sigma^2_{\log(\phi)}$	0.2372	0.104	0.1874	0.4071	0.4807	0.6218	0.9402	0.3587
ρ_{n0}	5.1096	5.2822	5.1794	4.764	4.2696	4.4937	4.4506	4.6259
$\sigma^2_{\xi n}$	0.1779	0.078	0.1406	0.3053	0.3605	0.4664	0.7051	0.269

We calculate these values as follows. We generate data for variables ϕ_{nt} in equation (3.4). These shocks are not observed directly so we approach the problem as follows. What we do observe are time series of netput prices (as in Eldon Ball's dataset), which are related to the market shock by equation (3.4). So we first plug values of randomly generated netput prices (10,000 time periods described in section 3.2) into the system, and solve for starting values of ϕ_{nt} with MATLAB function *fsolve*. This yields a time series of ϕ_{nt} that allows us to "learn" about its moments conditional on the AR(1) log-normally distributed prices, "average" production parameters, and elasticities.

From equation (3.5), the long run mean of market shocks is $\log(\bar{\phi}) = \rho_0 / (1-\rho_1)$, and the long run variance is $\sigma^2_{\log(\phi)} = \rho_1^2 \sigma^2_{\log(\phi)} + \sigma^2_\xi$, or $\sigma^2_\xi = (1-\rho_1^2)\sigma^2_{\log(\phi)}$ if we solve for the variance of the error term. Both the mean and the variance are directly obtained from the time series of starting values of ϕ_{nt}. Therefore we have two equations and three parameters (ρ_0, ρ_1, and σ^2_ξ) implying that we need to fix one. We fix ρ_1 = 0.5, and from the mean and variance of the mentioned time series (log($\bar{\phi}$) and $\sigma^2_{\log(\phi)}$ respectively) we calculate ρ_0 and σ^2_ξ for each netput *n* (as reported in table II.1).

Then we set $\log(\phi_{s=0}) = \rho_0 / (1-\rho_1)$, draw a random deviate from a *Normal*$(0, \sigma^2_\xi$), and using equation (3.5) and ρ_0 and ρ_1, we calculate market shock for the initial period, $\log(\phi_{s=0})$. We iterate over this procedure *S*=10,000 times and for each netput *n*, obtaining a time series of market shocks. These are finally plugged in system (3.4) to solve for the vector of "national" netput prices using MATLAB function *fsolve*, as explained in the text.

APPENDIX II

Estimation results of model (3.7) are presented in table II.1

Table II.1. Parameter estimates of initial wealth and the form of its heteroskedasticity, equation (3.7)

Dependent variable: *TNA*	**Region 1**	**Region 2**	**Region 3**
Explanatory variables	**Parameter estimates: γ_i , i = 1,2,3**		
Constant	0.724 (0.040)	0.8429 (0.058)	0.8917 (0.058)
VP	1.279 (0.064)	1.1375 (0.062)	0.5743 (0.042)
VP^2	-0.066 (0.008)	-0.019 (0.002)	-0.0096 (0.001)
R^2	0.16	0.1947	0.0922
Dependent variable: $\log(\hat{u}^2)$	**Region 1**	**Region 2**	**Region 3**
Explanatory variables	**Parameter estimates: δ_i , i = 1,2,3**		
Constant	-2.062 (0.045)	-1.925 (0.058)	-1.5069 (0.048)
VP	1.544 (0.071)	0.9639 (0.063)	0.416 (0.035)
VP^2	-0.105 (0.008)	-0.0264 (0.002)	-0.0061 (0.001)
R^2	0.16	0.1131	0.0805
$\hat{\sigma}_e$	2.084	2.170	2.0104

Note: *TNA*: Total Net Assets. *VP*: Value of Production. Standard errors in parenthesis.

Random generation of initial wealth $W_{0,ft}$: Based on parameter estimates in (3.7), the firm- and time-specific initial wealth ($W_{0,ft}$) is generated as follows:

STEP 1: Obtain the value of production of firm *f* and time *t*, calculated as: $VP_{ft} = \boldsymbol{y}_{ft}^{***'}\boldsymbol{p}_{ft}^{*}$.

Endogenous prices $\boldsymbol{p}_{ft}^{*}$ are those from section 3.2. We approximate firm's netput quantities $\boldsymbol{y}_{ft}^{***}$ as the solution of problem (2.3) under risk-neutrality. We do not assume risk-aversion at this stage because solving the expected utility problem requires conditioning on initial wealth which is what we are trying to calculate. The value of $\boldsymbol{y}_{ft}^{***}$ allows us to calculate a proxy for firm's value of production and initial wealth and is used exclusively in this step and nowhere else in the analysis.

STEP 2: Take a draw from $\kappa \sim logn(-\tfrac{1}{2}\hat{\sigma}_e^2, \hat{\sigma}_e^2)$. Together with $\hat{\delta}$ and VP_{ft} find the estimated variance of τ using the equation for $\hat{u}^2$. Because e is normal then κ is *log-normal*; since the mean of κ is one, the mean of *e* is $-\tfrac{1}{2}\hat{\sigma}_e^2$. The standard notation is that parameters in the log-normal are mean and variance of the underlying normal distribution.

STEP 3: Take a draw from a $Normal(0, \hat{u}^2)$ for the error term $\hat{\tau}$.

STEP 4: Calculate the initial wealth as: $W_{0,ft} = \hat{\gamma}_0 + \hat{\gamma}_1 VP_{ft} + \hat{\gamma}_2 VP_{ft}^2 + \hat{\tau}$.