How Reliable is Duality Theory in Empirical Work?

Francisco Rosas  
*Universidad ORT Uruguay & Centro de Investigaciones Económicas (CINVE)*

Sergio H. Lence  
*Iowa State University, shlence@iastate.edu*

Follow this and additional works at: https://lib.dr.iastate.edu/econ_las_pubs

Part of the Agricultural and Resource Economics Commons, Behavioral Economics Commons, Econometrics Commons, and the Economic Theory Commons

The complete bibliographic information for this item can be found at https://lib.dr.iastate.edu/econ_las_pubs/686. For information on how to cite this item, please visit http://lib.dr.iastate.edu/howtocite.html.

This Article is brought to you for free and open access by the Economics at Iowa State University Digital Repository. It has been accepted for inclusion in Economics Publications by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
How Reliable is Duality Theory in Empirical Work?

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 parametric form of the production function. 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. Production parameters are not precisely recovered when performing econometric estimation based on the duality approach, and the elasticity estimates are inaccurate. Deviations of own- and cross-price elasticities from initial median values, given our parameter calibration, range 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

Disciplines
Agricultural and Resource Economics | Behavioral Economics | Econometrics | Economic Theory

Comments
This is a manuscript of an article published as Rosas, Francisco, and Sergio H. Lence. "How reliable is duality theory in empirical work?." *American Journal of Agricultural Economics* 101, no. 3 (2019): 825-848. doi: 10.1093/ajae/aay071. Posted with permission.
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 parametric form of the production function. 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. Production parameters are not precisely recovered when performing econometric estimation based on the duality approach, and the elasticity estimates are inaccurate. Deviations of own- and cross-price elasticities from initial median values, given our parameter calibration, range 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 obtain 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 (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 (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.

A typical application of the dual problem begins by approximating the profit (cost) function with a parametric functional form, and applying Hotelling’s (Shephard’s) Lemma to obtain a parametric form for the optimal input demands and output supplies. Then, parameters are econometrically estimated using market data (prices and quantities), and finally used 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, and (e) it is more accurate and tractable for multi-output technologies. However, he argues that curvature properties should be pre-tested, and collinearity of prices and allocatable inputs induces estimation inefficiency.

The reliance of the approach on some restrictive 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 aforementioned 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 need for high-quality data for empirical estimation purposes.

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), because the FFF used drives the results. 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 an 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 assume perfect competition, profit maximization, certainty, and lack of measurement errors, deviations from duality theory only come from the FFF choice. As a result, the performance of duality theory in empirical applications cannot be judged, 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 dual 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 initial technology parameters, Monte Carlo simulations are used to compute the necessary price and quantity variables. While calibrated to be consistent with typical datasets encountered in practice, the levels of noise embedded in the simulated variables affect the data used in estimation, preventing duality theory from holding exactly. Hence, the initial production parameters may not be recovered with enough precision, and the estimated elasticities measurements may be less accurate 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 (i.e., the basic duality assumptions are met), we employ it to confirm that the dual approach is able to recover the production parameters with sufficient accuracy.

Second, we add noise to the generated panel of price and quantity variables to replicate the aforementioned features that characterize the real-world 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

---

1 The ability of the dual theorem to recover the parameters of the dual function (the primal–dual direction) is an analysis as important as the one performed in this study, but it is not addressed here due to space limitations and because the dual-primal direction is the preferred one in empirical applications. Furthermore, the assessment of the primal approach to recover the underlying production parameters is another important analysis, but for similar reasons, is left for future research.
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 yield 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 known production parameters, especially for those that are unobservable.2

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 known production parameters. Comparisons are performed using Lau’s (1976) Hessian identities between production and restricted profit functions.3

2. The Model of Individual Firms

Each firm underlying the simulated dataset is assumed to consist of a producer who chooses netputs so as to maximize the expected utility of uncertain terminal wealth ($\bar{W}_t$):4

\[
\begin{align*}
max_{\{y, y_0\}} E[U(\bar{W}_t)] &= max_{\{y, y_0\}} E[U(W_0 + \bar{\pi})], \\
&= max_{\{y, y_0\}} E[U(W_0 + \bar{\pi}^T \bar{y} + \bar{y}_0)].
\end{align*}
\]

---

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

3 This issue can also be interpreted in the framework of the existence of a representative technology arising from the aggregation of heterogeneous firms. It can be argued that this is not a problem exclusive of duality theory, and that the primal problem also bears it. We decided to work on the dual problem because it is an application extensively used to recover production parameters and estimate elasticities, but the other option is appropriate as well.

4 The model setup follows closely the one used in Rosas and Lence (2015), which in turn is based on Lau (1976).
In the above expression, $E(\cdot)$ denotes the expectation operator, $U(\cdot)$ is a strictly increasing and twice-continuously differentiable concave utility function, $W_0$ is initial wealth, $\bar{\pi}$ are uncertain end-of-period profits, vector $y \equiv [y_1, \ldots, y_N]^T$ comprises $N$ variable netput quantities, vector $\tilde{p}$ contains the corresponding variable netput prices normalized by $p_0$, which is the price of the numeraire netput $y_0$. The tilde ($\sim$) indicates a random variable.

Defining $K$ as the vector of $M$ quasi-fixed netputs, a production plan consists of the vector $[y_0, y, K]$ belonging to the production possibilities set $S \in \mathbb{R}^{1+n+m}$. As shown by Jorgenson and Lau (1974), there exists a one-to-one correspondence between the set $S$ and a production function $G(\cdot)$ (also constrained by the quasi-fixed netputs $K$), such that:

\[(2.2) \quad G(y, K; \alpha) = \max\{y_0/\{y_0, y, K\} \in S\},\]

where $\alpha$ denotes a set of production function parameters. Hence, problem (2.1) can be rewritten as

\[(2.3) \quad \max_y \{E[U(W_i)]\} = \max_y \{E[U(W_0 + \tilde{p}^T \bar{y} - G(\bar{y}, K; \alpha))]\}.\]

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

---

5 According to netput notation, a positive value is a net output and a negative value is a net input.

6 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).

7 The properties of the production function $G(\cdot)$ are: (i) the domain is a convex set of $\mathbb{R}^{n+m}$ 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 $y_0 = -G(y, K)$. 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 $\emptyset$ is the empty set.
Duality theory establishes a relationship between the production function \( G(y, K; \alpha) \) and the restricted profit function \( \pi_R(p, K; \beta) \), which Lau (1976) proved in terms of their Hessian matrices under the assumption of convexity and twice continuously differentiability of both functions. These Hessian relationships are critical for our analysis, because they allow us to not only express the restricted profit function parameters (\( \beta \)) in terms of the underlying production function parameters (\( \alpha \)), but also to compare the recovered parameters with the simulated ones (Rosas and Lence 2015).

To operationalize problem (2.3), we proceed by assigning functional forms. In particular, farmer \( f \) in period \( t \) is assumed to choose the level of expected output at the end of the growing season so as to maximize the expected value of a constant absolute risk aversion (CARA) utility function \( U(\tilde{W}_{ft,1}) = -\exp(\lambda_{ft} \tilde{W}_{ft,1}) \), with parameter \( \lambda_{ft} \) representing the coefficient of absolute risk aversion.

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 \( G(y_{ft}, K_{ft}; \alpha) \):\(^8\)

\[
 G(y_{ft}, K_{ft}; \alpha) = y_{ft}^T A_{1f} + K_{ft}^T A_{2f} + \frac{1}{2} y_{ft}^T A_{11f} y_{ft} + y_{ft}^T A_{12f} K_{ft} + \frac{1}{2} K_{ft}^T A_{22f} K_{ft} - \psi_{ft},
\]

where \( f \) and \( t \), respectively, index firms and time, \( A_{1f} \) and \( A_{2f} \) are \((N\times1)\) and \((M\times1)\) vectors of \( \alpha_{i,f} \) coefficients, \( A_{11f} \) and \( A_{22f} \) are respectively \((N\timesN)\) and \((M\timesM)\) symmetric and nonsingular matrices, \( A_{12f} \) is an \((N\timesM)\) matrix, and \( \psi_{ft} \) is a mean-zero heteroskedastic production shock given by expression (2.5) below. Submatrices \( A_{11f} \), \( A_{12f} \) and \( A_{22f} \) form a symmetric and positive

---

\(^8\) We restrict our analysis to the differentiable case because it is the standard in empirical applications of duality theory that we intend to evaluate.
A semi-definite \(((N+M)\times(N+M))\) matrix \(A_f\) of \(\alpha_{ij,f}\) coefficients.\(^9\) All \(\alpha_{i,f}\) and \(\alpha_{ij,f}\) coefficients are collectively denoted as \(\alpha_f\). This functional form is self-dual, the production and profit function Hessians are only functions of parameters, and is broadly employed in applications of the duality approach.

The mean-zero heteroskedastic production shock \(\psi_{ft}\) in equation (2.4) is specified as

\[
(2.5) \quad \psi_{ft} = \frac{2}{8} \left( y_{ft}^{1/2} \right)^T D_f v_{ft},
\]

where \(D_f\) is an \((N\times N)\) diagonal matrix, and \(v_{ft}\) is an \((N\times1)\) vector. Entries of \(v_{ft}\) corresponding to variable inputs are zero \((v_{nft} = 0)\), whereas those associated with variable outputs consist of the product of a systematic component \((v_{1nt})\) and an idiosyncratic component \((v_{2nft})\): \(v_{nft} = v_{1nt} \times v_{2nft}\).

The systematic shocks are \(v_{1nt} \sim [-1 + 2 \text{ Beta}(2, 2)]\), i.e., symmetric zero-mean shocks independent and identically distributed \((iid)\) over the interval \([-1, 1]\). The idiosyncratic shocks are modeled as \(v_{2nft} \sim \text{Uniform}(0.87, 1.13)\), which allows weather variables to not only affect production quantities over time but also have different local effects in a given year.\(^{10}\) The main diagonal of \(D_f\) is set equal to the inverse of the main diagonal of \((A_{11f})^{-1}\), so as to achieve the desired level of variability in each netput quantity and reduce variability induced by other netputs (especially in the case of inputs).\(^{11}\) Note that while this setup is consistent with firms facing output quantity uncertainty, the jointly specified technology induces uncertainty in all variable netputs.

---

\(^9\) Positive semi-definiteness is required because of the convention used in Lau (1976) that \(y_0 = -G(y, K)\).

\(^{10}\) To calibrate the width of the idiosyncratic shock 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 the idiosyncratic shocks. Details of this estimation are provided in the appendix of Rosas and Lence (2015).

\(^{11}\) Since \(\psi_{ft}\) is a function of \(y_{ft}\), the shock enters the solution of variable netput quantities in its first derivative and premultiplied by \((A_{11f})^{-1}\).
Figure 1, depicting simulated production shocks for netput 1 (output) and netput 8 (input) for all firms at time $t_0 = 1$, helps illustrate the two main reasons for adopting function (2.5). First, shock are heteroskedastic with a standard deviation that increases at a decreasing rate, which is consistent with larger firms being less exposed to uncertainty (e.g., because a bad weather draw is more likely to be offset by a good draw within the same firm). The top panels, plotting the distribution of each firm’s netput quantity $\bar{y}_{t_0}$ against the firm’s average netput quantity $\bar{y}_{t_0}$, show that the dispersion increases as the latter increases. However, as demonstrated by the bottom panels, it does so at a decreasing rate (because the coefficient of variation decreases with the firm’s average quantity produced). Second, consistent with the observed data, firm-specific shocks range from minus 10% to plus 10% (minus 60% to plus 60%) of the average quantity produced in the case of the less (more) disperse distribution. This point is shown in the middle panels, which plot the minimum (green), mean (blue), and maximum (red) production shocks ($\bar{\psi}$) as percentages of firms’ average netput quantities against the respective average quantities produced.

Firms also face end-of-period output price uncertainty. The methods used to incorporate price uncertainty and correlations between prices and quantities are explained in Appendix C.2.

3. Simulation of Panel Data
To analyze the empirical performance of the dual approach, two datasets are generated; namely, one noiseless and the other one incorporating noise. Consistent with, the noiseless data are used to confirm the ability of the dual approach to recover the original production parameters when data are problem-free. In this study, we focus on the noisy dataset, because it allows us to document the effects on the estimated production parameters when the data exhibit more realistic features.

---

12 For comparison, a pooled panel of farm-specific corn yields over a five-year period 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.
Figure 2 sketches the steps involved in the simulations, and the sections where they are explained. The first step is the same for the noiseless and noisy datasets, and consists of generating the starting production parameters $\alpha_f^*$ and quasi-fixed netputs $K_f^*$ by Monte Carlo simulations using the procedures described in Rosas and Lence (2015). To favor parameter identification, production parameters are allowed to vary across firms but not over time.\textsuperscript{13}

In the case of the noiseless simulations, the second step consists of drawing exogenous expected variable netput prices $p_{ft}^{**}$ as described in Appendix A. This is followed by the numerical solution to the firm’s optimization problem (2.3) which, due to the absence of noise and the normalized quadratic production FFF (2.4), collapses to a standard profit maximization with first-order conditions given by $p_{ft}^{**} - A_{1f} - A_{11f} y_{ft}^{**} + A_{12f} K_f^* = 0$. Hence, the optimal variable netput quantities for each farm and time period are computed as

$$y_{ft}^{**} = (A_{11f})^{-1} \left(p_{ft}^{**} - A_{1f} + A_{12f} K_f^* \right).$$

The final step to create the noiseless dataset involves aggregating across heterogeneous firms the simulated individual observations:

$$y_t^{**} = \sum_f y_{ft}^{**},$$

$$K_t^* = \sum_f K_f^*,$$

$$p_{nt}^{**} = \sum_f w_{nft} p_{nft}^{**}, n = 1, \ldots, N,$$

where $w_{nft} \equiv y_{nft}^{**} / y_{nt}^{**}$ is firm $f$’s share of the aggregate $n$th netput quantity at time $t$. That is, netput quantities are aggregated by adding across firms because they are homogeneous.

\textsuperscript{13} The ability of the dual approach to recover parameters associated with technological change is a very important topic, but it is left for future investigation for reasons of space.
commodities, whereas aggregate netput prices are weighted averages of farm-level prices. The resulting time series dataset \([y^*, p^*, K^*]\) is the one used to estimate the production function parameters \((\alpha^*_f)\), as discussed in section 4.

The procedure to construct the noisy dataset is more involved, because it requires the numerical maximization of expected utility (rather than just profits), and the incorporation of various sources of noise. In the second step, explained in Appendix A, *endogenous* expected variable netput prices \(p^*_f\) are drawn conditioning on the values of \(\alpha^*_f\) and \(K^*_f\). An additional step, described in Appendix B, is necessary to obtain calibrated values of initial wealth \(W_{f,0}\) and the risk-preference parameter \(\lambda^*_f\), which are used at the next step to compute the expected netput quantities \(y^*_f\) that maximize the expected utility of end-of-period terminal wealth (see Appendix C.1 for details). Before aggregating across farms, the following sources of noise are incorporated into the simulated data:

- shocks to expected price and quantity variables (outlined in Appendix C.2),
- omission of variables (by eliminating some of the netput series, as explained in section 3.1),
- aggregation across netputs (discussed in Appendix C.3), and
- measurement errors in price and quantity variables (addressed in Appendix C.4).

Finally, the noisy dataset \([y, p, K]\) used to conduct time-series estimation is obtained by aggregating across firms, in a manner analogous to expressions (3.2)-(3.4). The set of production function parameter estimates corresponding to the noisy data is denoted by \(\hat{\alpha}_f\) (see section 4).

The present simulations are parameterized so as to obtain panel data for \(N = 8\) variable netputs and \(M = 1\) quasi-fixed netput over a period of \(T = 50\) years from \(R = 3\) regions, each composed of \(F = 10,000\) heterogeneous firms,\(^{14}\) such that firm heterogeneity is higher across regions than within them. Therefore, conditional on the set of parameters \(\alpha^*_f\), there are \(R \times F \times T = \)

\(^{14}\) This figure roughly represents about one-fifth of the number of farms in a given state of the Corn Belt (IA, IL, IN, MO, and OH), Lake States (MI, MN, and WI), and Northern Plains (KS, ND, NE, and SD) regions. Available U.S. state-level time-series datasets with information on prices and quantities of agricultural outputs and inputs comprise no more than 50 years of observations.
1.5 million observations for each variable in the noisy data panel \([y_t, p_t, K_t; \alpha_f^*]\). Upon aggregation over the 10,000 heterogeneous firms at each time \(t\), we obtain a dataset of 50 observations for each variable per region that we use to estimate netput demands and supplies as shown in system (4.2) below.

### 3.1. Data Used for Estimation

The noiseless data \([y_t^*, p_t^*, K_t^*]\) 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 (see Appendix A). 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 starting parameters for that same region.

In the case of the noisy data \([y_t, p_t, K_t]\), we explore the effects of data omission and aggregation by using only \(N' = 4\) netputs for estimation. The reduced number of netputs results from omitting one input and one output, pooling two variable outputs into one, and pooling two variable inputs into one.\(^{15}\) Panel A in Figure 3 represents the structure of the noisy data for the baseline analysis. To perform the estimation using a sample instead of the entire population (to conform with reality), and to avoid final results to be dependent on a single sample, we proceed as follows. We take the region’s population of \(F = 10,000\) heterogeneous firms and draw 100 samples of 6,000 observations each; for each sample, we aggregate over the heterogeneous firms resulting in a time-series dataset of 50 observations for each variable, and conduct econometric estimation of system (4.2) below.\(^{16}\) For the same reasons stated above, we select region 1 to

\(^{15}\) The producer optimally chooses a set of \(N\) variable netputs to maximize profits, 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 netputs are omitted, or when some inputs are not part of the surveyed set.

\(^{16}\) 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 large compared to real-world datasets used to construct state-level aggregates. For example, the 2004 ARMS dataset consists of samples that average 428 firms per state, ranging between 48 and 1,600 firms depending on the state. For comparison, estimation was also conducted using the entire population in the region and aggregating across all firms, which implies only one time-series dataset to be estimated. Results were very similar to the case of 100 samples from the population.
conduct the estimation. The effects of pursuing estimation with data from more heterogeneous regions (e.g., to capture a broader area and/or increase the sample size, which are common in these applications) are shown as a sensitivity analysis.

4. Estimation

For estimation purposes, the restricted profit function \( \pi_R(p, K; \beta) \) that results from solving problem (2.3) is approximated by the following normalized quadratic FFF:

\[
\pi_R(p, K; \beta) = p^T B_1 + K^T B_2 + \frac{1}{2} p^T B_{11} p + p^T B_{12} K + \frac{1}{2} K^T B_{22} K + p^T \kappa,
\]

where \( B_1 \) and \( B_2 \) are \((N \times 1)\) and \((M \times 1)\) vectors of \( \beta_i \) coefficients, \( B_{11} \) and \( B_{22} \) are symmetric \((N \times N)\) and \((M \times M)\) matrices, respectively, and \( B_{12} \) is an \((N \times M)\) matrix. Matrices \( B_{11}, B_{12}, \) and \( B_{22} \) form a symmetric \(((N+M) \times (N+M))\) matrix \( B \) of \( \beta_{ij} \) coefficients, which in the case of the normalized quadratic profit function is exactly the Hessian matrix with respect to \((p, K)\). All \( \beta_i \) and \( \beta_{ij} \) coefficients collectively form the set \( \beta \). The error structure \((p^T \kappa)\) is consistent with McElroy’s (1987) additive general error model (AGEM) applied to the case of profit functions. The \((N \times 1)\) vector of random variables \( \kappa \) is jointly normally distributed with mean zero and an \((N \times N)\) covariance matrix \( \Sigma \). This covariance matrix induces contemporaneous correlation between equations. Also, the DGPs of netput prices—both exogenous and endogenous—involve AR(1) processes (see Appendix A), implying serial autocorrelation in the independent variables that needs to be addressed in the estimation.

Application of Hotelling’s lemma yields the following set of input demands and output supplies:

\[
y = B_1 + B_{11} p + B_{12} K + \kappa.
\]
System (4.2) is estimated by means of iterated seemingly unrelated regressions (SUR), which converges to maximum likelihood, and is the most common method employed in empirical work based on duality theory. Symmetry cross-equation restrictions ($\beta_{ij} = \beta_{ji}, i \neq j$) in matrix $B_{11}$ are imposed for the estimation.

4.1. Addressing Mean-Independence Violations in Estimation

For estimation purposes, the data are first-differenced, as suggested by an inspection of the autocorrelation and partial autocorrelation functions of the noiseless and noisy time series.

In addition, and only for the case of the noisy dataset, an instrumental variables (IV) approach is employed to address the problems caused by the omission of some variable netput prices from system (4.2). To instrument the omitted prices, we use the price values themselves because they are the best possible instruments. Also only for the noisy dataset, an IV approach is used to control for the endogeneity of the explanatory price variables, which arises because they are correlated with the error terms due to the supply and demand shocks. To instrument for the endogenous prices, we make use of the fact that we know the underlying source of endogeneity, i.e., prices in system (4.2) are correlated with the error term $\mathbf{K}$ as a consequence of systematic market shocks $\phi_{nt}$s (see Appendix A). Since instruments have to be correlated with prices but uncorrelated with the error term, and we know the shocks $\phi_{nt}$ used to construct the price series, we construct instruments by regressing each netput price on its own systematic shock: $p_{nt} = \psi_0 + \psi_1 \phi_{nt} + iv_{nt}$. The residual ($iv_{nt}$) is an ideal instrument for the corresponding price because it is correlated with $p_{nt}$ and orthogonal to the systematic shock ($\phi_{nt}$) by construction; hence, it represents the variation in prices not explained by the systematic shocks. There is one instrument for each netput price, as well as one instrument for each omitted variable.

The parameters comprising 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. Hence, they are the foundation for the estimated profit function Hessian matrix and the elasticities of
netput quantities with respect to own price, cross prices, and quasi-fixed netputs. As depicted in Figure 4, we estimate the Hessian matrices \([ \hat{B} ]\) and \([ \hat{B} ]\) from the noiseless and noisy datasets, respectively. The Hessians are then transformed into the corresponding elasticity matrices \([ \hat{E} ]\) and \([ \hat{E} ]\) in a straightforward manner.

In order to compare estimated elasticities with initial values, we begin from the known firm-specific production function Hessian matrix \([A]_f\) and convert it into the corresponding profit function Hessian \([B]_f\) by resorting to Lau’s Hessian identities. We then transform the Hessian \([B]_f\) into the matrix of own- and cross-price initial elasticities and quasi-fixed initial elasticities of netput quantities \([E]_f\). Finally, as indicated in Figure 4, we compare the initial \([E]_f\) versus the estimated values (\([ \hat{E} ]\) and \([ \hat{E} ]\)) to evaluate how precisely we recover the starting elasticities under duality theory, for both the noiseless and the noisy data. Note that this comparison implies that each initial parameter is represented by a distribution (across farms), whereas the estimation yields a corresponding point estimate and its confidence interval.

5. Results

Estimation results for the noiseless and noisy data are discussed separately next.

5.1. Estimation with Noiseless Data

Econometric estimates from the data obtained by aggregating across heterogeneous firms but without any other source of noise (i.e., \([y^*, p^*, K^*]\), explained in section 3.1) are omitted to save space, as they can be found in Rosas and Lence (2015). The estimates reveal that the dual approach is able to recover initial production parameters fairly accurately. More specifically, estimated elasticities with respect to prices (quasi-fixed netputs) deviate on average by 12.4% (7.5%) from the median of the starting elasticities according to the computed root mean squared error (RMSE). The RMSE summarizes the average difference between each entry of the

\[ \text{RMSE} = \sqrt{\frac{1}{64 \times 10,000} \sum \sum (\hat{E}_{ij} - \tilde{E}_{ij})^2} \]

where 64 is the number of parameters, 10,000 is the number of draws from the limiting distribution of the SUR

---

17 When compared to the median of the distribution, RMSE is computed as \[ [(64 \times 10,000)^3 \epsilon^{\epsilon} (\tilde{E}_{a} - \tilde{E}_{a})^{1/2} \]
estimated elasticity matrix and the median of the corresponding initial elasticity distribution. The RMSE accounts for two sources of error, namely, one due to the SUR estimation error for each of the 64 parameters, and the other one associated with the difference between the estimated and the initial value of the elasticity across the 64 parameters. Given that the SUR estimation provides only a minor source of error because point estimates are all highly significant, we argue that most of the RMSE can be attributed to the deviations between the estimated and the initial values across elasticities.

5.2. Estimation with Noisy Data

Estimation with noisy data (i.e., \([y_t, p_t, K_t]\), explained in section 3.1) yields 16 own- and cross-price elasticities of variable netput quantities, and 4 elasticities with respect to quasi-fixed netputs. Figure 5 shows the distribution of the firm-specific (initial) price elasticities, and their corresponding SUR point estimates indicated with a red circle (and the bounds of 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 entries (2, 2), (2, 3), (3, 2), and (3, 3) of the own- and cross-price elasticity matrix, the distributions involve more than one initial elasticity due to the aggregation of netputs, as described in section 3.1. In these cases, to be able to compare with the elasticities estimated by means of SUR, we construct “new initial” elasticity distributions as the revenue-weighted averages of the distributions of the corresponding original initial elasticities.

In light of the conclusions from the previous sub-section, we measure the accuracy in recovering initial elasticities by comparing the estimated values to the medians of the respective parameter estimates, and subscript \(s\) indicates the \(s^{th}\) draw of the \(ij^{th}\) parameter. Comparison with the mean can be performed by substituting \(\bar{E}\) by \(\bar{E}\). The RMSE averages over all the 64×10,000 squared differences. A measure of its dispersion is achieved by computing the standard deviation of these 64×10,000 values before averaging over them.
distributions. Visual inspection of Figure 5 suggests that, when comparing where the estimated values fall relative to where the distributions accumulate more mass, the dual approach provides a good approximation of the initial distribution in some instances, but a poor one in others. However, Table 1 shows that the percentage difference between the median of the initial distribution ($\bar{E}_{ij}$) 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 own-price elasticities, reported along the main diagonal, are not recovered with much precision, given that the differences range between 15% and 44%. Importantly, the own-price elasticities for the netputs that are not aggregated with other netputs (i.e., netputs 4 and 5, corresponding to entries (2, 2) and (3, 3)) are more imprecisely estimated than the other main diagonal elements which do arise as aggregated netputs. As expected, the off-diagonal elements (i.e., the cross-price elasticities) are less accurately estimated than the main diagonal entries, as they require more information to be recovered.

As a summary measure of the dispersion in recovering the initial elasticities, we calculate the RMSE of the difference between the median of the initial distribution and the SUR estimated values for all 16 estimated price elasticities. Table 2 shows that the RMSE for the baseline scenario equals 0.22 in elasticity units. The average value of all initial elasticities (calculated as the mean absolute value of all the medians of the initial 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 initial elasticities. These results provide evidence that the dual approach is unable to deliver precise estimates of underlying production parameters when employing data featuring real-world characteristics.

The estimation of variable netput elasticities with respect to quasi-fixed netputs is even less accurate. Results are represented graphically in Figure 6. Each panel titled "$E_{ik}$" is the elasticity of netput $i$ with respect to the quasi-fixed netput. The SUR point estimates of the elasticities are

---

18 Comparisons using the means of the distributions provided less accurate results.
within the support of the respective initial distributions except for $E_{4k}$, in which case the initial support is positive but the estimated elasticity is negative and the 95% confidence interval does not even overlap with the initial support. The baseline scenario in Table 2 shows that the RMSE relative to the median of the initial distribution is 0.67 expressed in elasticity units, and the average value of the elasticities is calculated at 0.54. These results imply that the inaccuracy in recovering the starting elasticities, averaged over the 4 netputs, amounts to 123% of the original elasticities.

5.3. Sensitivity Analysis
We explore the robustness of noisy data estimation results to changes in the sources and levels of noise. Estimation results are not driven by the set of netputs omitted or aggregated. For example, Table 2 shows that estimation with the noisy data structures shown in panel B of Figure 3 yield price elasticity estimates with respect to variable netputs that are off by 57% (case 1) and 69% (case 2) relative to the starting price elasticities. Similarly, the corresponding elasticities with respect to quasi-fixed netputs differ by 76% and 120% from their starting values. Therefore, these values are quantitatively similar to the ones for the baseline scenario.\footnote{We present only a few alternative scenarios due to the computational burden of such analysis.}

We regularly encounter empirical applications of duality theory with time-series data where observations from different regions or states are pooled together for estimation (e.g., Schuring, Huffman, and Fan (2011), and O’Donnell, Shumway and Ball (1999)). By expanding the sample size, pooling has the advantage of increasing the degrees of freedom, which is especially helpful in the presence of several explanatory variables. However, pooling also has the downside of adding observations from states that are likely to have different technology. We explore the consequences of such practice by conducting a sensitivity analysis.

Pooling 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. Hence, for this sensitivity analysis, we exploit the noisy simulated data explained in section 3.1 but for all regions 1, 2, and 3. For 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, and aggregate across its heterogeneous firms to obtain the corresponding state-level time-series. The resulting dataset, consisting of 750 observations (= 50 years × 5 states/region × 3 regions), is then used to estimate system (4.2). Following the aforementioned studies, dummy variables are added for observations corresponding to regions 1 and 2, leaving region 3 as the base. The estimated parameters are transformed into netput elasticities with respect to variable netput prices and with respect to quasi-fixed netputs, and compared with the starting elasticities. As in the previous analysis, the latter are represented by the respective distributions of starting firm-specific elasticities (which now involve firms in the three regions).

The RMSE relative to the median of such distribution and averaged over the 16 price elasticities being calculated equals 53%, which is very similar to the findings above (see last column in Table 2). Such figure is computed by dividing the RMSE relative to the median of the starting distribution (0.18), by the median of the starting distribution of elasticities (0.35). Divergence from initial price elasticities ranges between 11% and 209%, depending on which of the 16 entries of the elasticity matrix we consider. Standard errors of the estimated elasticities are lower than in the previous analysis, which is in part a consequence of the increased number of observations. However, that does not contribute to reducing the bias in the estimated parameters and elasticities relative to the starting values. Similarly, the RMSE for the netput quantity elasticities with respect to quasi-fixed netputs also indicates that production parameters are not recovered accurately.

Therefore, the practice of incorporating data from other regions, characterized by a more heterogeneous technology than within a region, reduces the standard error of the point estimates and enhances statistical significance, but is of little help in reducing the bias relative to the starting elasticity values.
6. 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 parametric form of the production function. However, the duality theorem requires assumptions which are unlikely to hold in practice; 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, elasticity estimates may 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 often 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 an FFF, and estimate its parameters (and the corresponding elasticities) using traditional econometric methods. 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 address them in the 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 instance, 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.

However, the use of noisy data prevents the dual approach from providing parameter estimates that are sufficiently close to their starting values. The root mean squared error, measuring the average deviation of the estimated elasticities from their median initial values, is calculated at 71%, implying that the dual approach estimates elasticities are, on average, 71% away from the initial values. 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.

Future research may complement this analysis by assessing the empirical ability of the primal approach in recovering the underlying production parameters in the context of noisy datasets, with the objective of documenting which approach does a better job under different levels of noise. Other avenues of inquiry may constitute the assessment of this approach’s performance in recovering the production function parameters but when employing cross-sectional data, or the parameters of technical change over time. Also, other sources of noise can be added such as the treatment of some variable netputs as quasi-fixed, an alternative frequently employed by practitioners to address the lack of price data for those netputs.
References


_____. “Agricultural Productivity in the U.S.” Available at: http://www.ers.usda.gov/Data/AgProductivity/


Figure 1. Production shock as a function of firm’s average variable netput quantity (\( \bar{y}_{f_0} \)) at time \( t = t_0 \), for selected netputs.

Note: Top panels: distribution of netput quantities (\( \bar{y}_{f_0} \)) faced by each firm. Middle panels: minimum, mean, and maximum shock as percentage of firm’s average quantity \( \bar{y}_{f_0} \). Bottom panels: coefficient of variation of the distribution of quantities \( CV(\bar{y}_{f_0}) \) by firm.
Figure 2. DGP of noiseless and noisy datasets used for estimation.
Panel A. Baseline Scenario

<table>
<thead>
<tr>
<th>Variable</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
</table>

Panel B. Sensitivity Analysis for Omission and Aggregation of Variable Netputs (Cases 1 and 2)

<table>
<thead>
<tr>
<th>Variable</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
</table>

Figure 3. Structure of the noisy dataset used for estimation.

Figure 4. Comparison of initial and estimated elasticities for noiseless and noisy datasets.
Figure 5. Own- and cross-price elasticities of variable netput quantities: Initial farm-level distributions versus estimated values obtained from noisy data.

Note. Each $ij$ panel is the $ij$ entry of the $4 \times 4$ own- and cross-price elasticity matrix $E$ estimated from noisy data. The elasticity value is in the horizontal axis and histogram frequency in the vertical. Each histogram depicts the distribution across firms of the initial elasticity ($E_{ij}$). The red dot is the SUR estimated elasticity ($\hat{E}_{ij}$) and the red “+” signs denote the bounds of the 95% confidence interval.
Figure 6. Elasticities of variable netput quantities with respect to quasi-fixed netputs:
Initial farm-level distributions versus estimated values obtained from 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. Each histogram depicts the distribution across firms of the initial elasticity ($E_{ik}$). The red dot is the SUR estimated elasticity ($\hat{E}_{ik}$) and the red “+” signs denote the bounds of the 95% confidence interval.
Tables

Table 1. Comparison of price elasticities estimated from noisy data ($\hat{E}_q$) with medians of distributions of initial price elasticities ($\bar{E}_i$).

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

Note: Interval is the 95% confidence interval of the point estimate $\hat{E}_i$. 
Table 2. Comparison of elasticities estimated from noisy data ($\hat{E}_{ij}$) with medians of distributions of initial elasticities ($\bar{E}_{ij}$), and different sources of noise.

<table>
<thead>
<tr>
<th>Elasticities with Respect to</th>
<th>Baseline Scenario</th>
<th>Sensitivity Analysis</th>
<th>Regional Pooling</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMSE</td>
<td>Omission and Aggregation of Variable Netputs</td>
<td>Case 1</td>
</tr>
<tr>
<td>Variable Netput Prices</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average of Absolute Medians</td>
<td>0.22</td>
<td>0.26</td>
<td>0.20</td>
</tr>
<tr>
<td>% deviation</td>
<td>0.31</td>
<td>0.46</td>
<td>0.29</td>
</tr>
<tr>
<td>% deviation</td>
<td>71%</td>
<td>57%</td>
<td>69%</td>
</tr>
<tr>
<td>Quasi-Fixed Netput Quantity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMSE</td>
<td>0.67</td>
<td>0.44</td>
<td>0.42</td>
</tr>
<tr>
<td>Average of Absolute Medians</td>
<td>0.54</td>
<td>0.57</td>
<td>0.35</td>
</tr>
<tr>
<td>% deviation</td>
<td>123%</td>
<td>76%</td>
<td>120%</td>
</tr>
</tbody>
</table>

Note: Each scenario consists of a different set of omitted netputs and a different set of netputs aggregated together.

In the baseline scenario, netputs 3 and 8 are omitted, and netputs 1 and 2, and 6 and 7 are aggregated. In case 1 of the sensitivity analysis, netputs 1 and 4 are omitted, and netputs 2 and 3, and 7 and 8 are aggregated. In case 2, netputs 3 and 7 are omitted, and netputs 1 and 2, and 5 and 6 are aggregated. The regional pooling consists of data coming from three regions, each with 5 states, and each state with 2000 firms, over 50 years. After aggregation across heterogeneous firms, results in 750 observations (= 50 years × 5 states/region × 3 regions), instead of only 50. Regional dummies are put in regions 1 and 2, leaving region 3 as the base. Average of Absolute Medians is the mean absolute value of all the medians of the initial distributions.
Appendix A. Random Generation of Expected Variable Netput Prices ($p_{ft}^{**}$ and $p_{ft}^*$)

Firm-specific endogenous ($p_{ft}^*$) and exogenous ($p_{ft}^{**}$) netput prices are generated as deviations from the respective “national” prices ($p_{US,t}^*$ and $p_{US,t}^{**}$) discussed in the next two subsections.

With the endogenous case as an example, first, regional prices for the $n$th netput are calculated as:

$$p_{nrt}^* = p_{n,US,t}^* d_r \varepsilon_{nrt},$$

where $d_r$ is a regional indicator with mean one across regions ($[d_1 d_2 d_3] = [0.90 \ 1.00 \ 1.10]$), and $\varepsilon_{nrt} \sim [0.95 + 0.10 \ \text{Beta}(2, 2)]$ is a mean-one symmetric shock independent from $d_r$ representing random deviations from national prices. Then, firm-specific random prices are generated as deviations from the respective regional average,

$$p_{nt}^* = p_{nft}^* \varepsilon_{nft},$$

where $\varepsilon_{nft} \sim [0.80 + 0.40 \ \text{Beta}(2, 2)]$ is a symmetric mean-one shock independent of $\varepsilon_{nft}$. Farm-specific shocks ($\varepsilon_{nft}$) are relatively small, so as to mimic the small cross-sectional variability observed in netput prices.

The calibration of farm-specific shocks ($\varepsilon_{nft}$) implies firm-specific prices with a coefficient of variation of 0.08, which doubles that of firm prices in USDA-ARMS dataset and, if anything, favors parameter identification. Also to favor identification of parameters in estimation, netput prices are assumed to be continuous$^{21}$ and independent from firm size.

A.1. Exogenous “National” Prices ($p_{US,t}^{**}$)

Exogenous “national” prices for the $n$th netput ($p_{n,US,t}^{**}$) are generated by assuming that they follow the AR(1) process

$$\ln(p_{n,US,t}^{**}) = \theta_{0n} + \theta_{1n} \ln(p_{n,US,t-1}^{**}) + \zeta_{n,t},$$

where $\theta_{0n}$ and $\theta_{1n}$ are parameters, and $\zeta_{n,t} \sim \text{Normal}(0, \sigma_{\zeta}^2)$ is an error term. Table A.1 reports the parameter estimates obtained by fitting regression (A.1) to the observed time series of futures crop prices from the CME, and input prices from Eldon Ball’s (USDA-ERS) dataset. The

$^{20}$ Analogous procedures are used to compute $p_{ft}^{**}$.

$^{21}$ The observed firm-level prices for a netput in the USDA-ARMS dataset are concentrated in about four to five values in each period, which contrasts with the continuum of firm-level values computed in the simulation.
parameters in Table 1 are used to simulate prices matching the mean, standard deviation, and serial autocorrelation of the observed series, by (a) setting the value for the first iteration equal to the unconditional mean \( \ln( P_{n,US,t=0}^{**} ) = E[ \ln( P_{n,US,t}^{**} ) ] = \theta_0/(1 - \theta_1) \), (b) taking 10,000 random draws from a Normal(0, \( \sigma^2_{\epsilon_t} \)) distribution, (c) plugging them into expression (A.1) to obtain a log-price series by iteration, and (d) keeping the last 50 values as the desired exogenous “national” netput prices \( P_{n,US,t}^{**} \).

A.2. Endogenous “National” Prices (\( P_{US,t}^* \))

Simulated endogenous “national” prices (\( P_{US,t}^* \)) are obtained by assuming that aggregate changes in netput quantities across individual firms lead to changes in market prices. This endogeneity induces correlation between prices and the error term in system (4.2), which violates the orthogonality assumption required by Ordinarily Least Squares.

Endogeneity is achieved by postulating the system of isoelastic netput demands and supplies

(A.2) \( Q_t = \Phi \ p_{US,t}^\eta \).

The \( N \)-vector \( Q_t \) consists of the aggregate output demands and input supplies faced by firms at time \( t \), \( p_{US,t}^\eta \equiv [ p_{1,US,t}^\eta, \ldots, p_{N,US,t}^\eta ]^T \) denotes an \( N \)-vector of “national” netput prices, each raised to the power of the calibrated netput-specific own-price elasticity of demand or supply (\( \eta \)), and \( \Phi \) is an \((N \times N)\) diagonal matrix of supply and demand netput-specific time-varying positive scalars \( \phi_{ht} \). Based on the FAPRI Elasticities Database and other sources, own-price elasticities are set equal to \([ \eta_1, \ldots, \eta_N ] = [-0.25, -0.21, -0.75, 0.90, 0.87, 0.85, 0.83, 0.80] \).

The objective is to find farm-specific netput prices \( p_{f,t}^* \) such that the optimal netput quantities aggregated across firms (\( y_t^* = \Sigma_f y_{f,t}^* \)) equal the aggregate output demands and input supplies.

---

22 Note that individual firms are price takers, and therefore consider price as exogenous when making decisions.
faced by firms ($Q_t$). To this end, consider the solution to the firm-specific optimization problem (3.1), which yields the firms’ optimal aggregate output supplies and input demands:

(A.3) \[ y_t^* \equiv \Sigma_f [(A_{11f})^{-1} (p_{ft} - A_{1f} + A_{12f} K_f^*) + \psi_{ft}], \]

where $p_{ft}$ is a vector comprising farm-specific prices $p_{nft} \equiv p_{n,US,t} d_f \varepsilon_{nrt} \varepsilon_{f,n,t}$ (explained above), and $\psi_{ft}$ is the vector of heteroskedastic production residuals (2.5) (e.g., optimization mistakes, weather shocks, deviation of prices from expected values, etc.) independent from the price shocks.

Since the firm-specific time-invariant matrices of production coefficients $A_{1f}, A_{12f},$ and $A_{11f}$ are generated by introducing deviations from the respective starting production matrices $A_1, A_{12},$ and $A_{11},$ and such deviations are independent from the price shocks ($\varepsilon_{nrt}$ and $\varepsilon_{f,n,t}$) and the production shocks ($\psi_{ft}$), for a sufficiently large number of farms ($F$) and by the law of large numbers, $y_t^*$ converges in distribution to a Normal random variable whose mean is

(A.4) \[ \bar{y}_t = F (A_{11})^{-1} p_{US,t} - F (A_{11})^{-1} A_1 + F (A_{11})^{-1} A_{12} K_t^* + F \bar{\psi}_t. \]

This expression depends only on the known “average” production parameters and “national” time-$t$ prices $p_{US,t},$ which are the same as those in the isoelastic demand or supply function (A.2) faced by firms. Thus, the time-$t$ endogenous “national” netput prices that clear the markets ($Q_t = y_t^*$) is obtained by numerically solving for $p_{US,t}^*$ the system

(A.5) \[ \Phi^0 p_{US,t}^* = F (A_{11})^{-1} p_{US,t}^* - F (A_{11})^{-1} A_1 + F (A_{11})^{-1} A_{12} K_t^* + F \bar{\psi}_t, \]

which is derived by equating the left-hand-sides of expressions (A.2) and (A.4).
According to equation (A.5), the endogenous “national” prices are determined by the time-varying scalars $\phi_{nt}$ comprised in matrix $\Phi$. These scalars, whose variability represents systemic shocks, are modeled as auto-correlated and log-normally distributed:

\[
\ln(\phi_{nt}) = \rho_{0n} + \rho_{1n} \ln(\phi_{nt-1}) + \xi_{nt},
\]

where $\xi_{nt} \sim \text{Normal}(0, \sigma_{\xi_{nt}}^2)$. Parameters $\rho_{0n}$, $\rho_{1n}$, and $\sigma_{\xi_{nt}}^2$ are calibrated so that they yield “national” prices $p_{US,t}^*$ with descriptive statistics comparable to the observed output and input prices (see Table A.2 for the specific parameter values). To generate the shock series, we set $\ln(\phi_{n,t=0}) = \rho_{0n}(1 - \rho_{1n})$ (i.e., the unconditional mean of $\ln(\phi_{n,t=0})$), take 10,000 draws from a $\text{Normal}(0, \sigma_{\xi_{nt}}^2)$ distribution, plug them into expression (A.6) to generate iteratively the systematic shocks ($\phi_{nt}$), and keep the last 50 of them as the series of shocks used to solve for the endogenous “national” netput prices $p_{US,t}^*$ from equality (A.5).23

It is worth noting that price variability is critical to recover production parameters, because it contributes to the identification of a bigger portion of the production function. The simulated systematic shocks are independent from each other because random draws from $\text{Normal}(0, \sigma_{\xi_{nt}}^2)$ are independent; however, when plugged into system (A.6) correlation between national prices is induced through matrix $A_{11}$. This DGP ultimately generates national netput prices that exhibit higher variance and lower cross-correlations than the CME future crop prices and Eldon Ball’s input prices. These two features favor identification in estimation when prices are explanatory variables, as it is our case.

---

23 To obtain the calibrated parameters in Table A.2, we rely upon equation (A.5) because the $\phi_{nt}$ shocks are not directly observed, but time series of netput prices are. So we first plug the 10,000 exogenous “national” prices $p_{US,t}^{**}$ (described in section A.1) into system (A.5), and solve for starting values of $\Phi$. In this manner we compute time series of $\phi_{nt}$ that allow us to “learn” about their unconditional means and variances, given prices, “average” production parameters, and elasticities. Since the unconditional mean and variance of $\ln(\phi_{nt})$ are given by $E[\ln(\phi_{nt})] = \rho_{0n}(1 - \rho_{1n})$ and $\sigma_{\phi_{nt}}^2 = \rho_{1n}(1 - \rho_{1n})$, respectively, and there are three unknown parameters ($\rho_{0n}$, $\rho_{1n}$, $\sigma_{\phi_{nt}}^2$), we arbitrarily fix $\rho_{0n} = 0.5$ to guarantee a stationary AR(1) process, and compute $\rho_{1n}$ and $\sigma_{\phi_{nt}}^2$ from the mean and variance for each netput $n$ (as reported in Table A.2).
Table A.1. Estimation results of the OLS regression model used to generate random exogenous “national” prices from equation (A.1).

<table>
<thead>
<tr>
<th></th>
<th>$n = 1$</th>
<th>$n = 2$</th>
<th>$n = 3$</th>
<th>$n = 4$</th>
<th>$n = 5$</th>
<th>$n = 6$</th>
<th>$n = 7$</th>
<th>$n = 8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_{n0}$</td>
<td>-0.031</td>
<td>-0.065</td>
<td>-0.012</td>
<td>-0.031</td>
<td>-0.001</td>
<td>-0.057</td>
<td>0.041</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>(0.038)</td>
<td>(0.032)</td>
<td>(0.030)</td>
<td>(0.064)</td>
<td>(0.035)</td>
<td>(0.035)</td>
<td>(0.024)</td>
<td>(0.037)</td>
</tr>
<tr>
<td>$\theta_{n1}$</td>
<td>0.680</td>
<td>0.34</td>
<td>0.67</td>
<td>0.902</td>
<td>0.861</td>
<td>0.60</td>
<td>0.843</td>
<td>0.923</td>
</tr>
<tr>
<td></td>
<td>(0.094)</td>
<td>(0.14)</td>
<td>(0.11)</td>
<td>(0.079)</td>
<td>(0.080)</td>
<td>(0.12)</td>
<td>(0.080)</td>
<td>(0.054)</td>
</tr>
<tr>
<td>$\sigma_{\xi}^2$</td>
<td>0.0680</td>
<td>0.0342</td>
<td>0.0372</td>
<td>0.0340</td>
<td>0.0439</td>
<td>0.0392</td>
<td>0.0207</td>
<td>0.0237</td>
</tr>
</tbody>
</table>

Note: Standard errors in parenthesis. Number of observations in each regression: 44.

Table A.2. Calibrated parameter values for market shocks ($\phi_n$) used in equation (A.7)

<table>
<thead>
<tr>
<th></th>
<th>$n = 1$</th>
<th>$n = 2$</th>
<th>$n = 3$</th>
<th>$n = 4$</th>
<th>$n = 5$</th>
<th>$n = 6$</th>
<th>$n = 7$</th>
<th>$n = 8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho_{on}$</td>
<td>5.1096</td>
<td>5.2822</td>
<td>5.1794</td>
<td>4.764</td>
<td>4.2696</td>
<td>4.4937</td>
<td>4.4506</td>
<td>4.6259</td>
</tr>
<tr>
<td>$\rho_{on}$</td>
<td>0.5000</td>
<td>0.5000</td>
<td>0.5000</td>
<td>0.5000</td>
<td>0.5000</td>
<td>0.5000</td>
<td>0.5000</td>
<td>0.5000</td>
</tr>
<tr>
<td>$\sigma_{\xi_n}^2$</td>
<td>0.1779</td>
<td>0.078</td>
<td>0.1406</td>
<td>0.3053</td>
<td>0.3605</td>
<td>0.4664</td>
<td>0.7051</td>
<td>0.269</td>
</tr>
</tbody>
</table>
Appendix B. Simulation of Initial Wealth ($W_{ft,0}$) and Coefficient of Risk Aversion ($\lambda_{ft}$)

A firm’s initial wealth is postulated to be a function of its value of production, because initial wealth measured as total net assets ($TNA_{ft}$) is strongly associated to the value of production ($VP_{ft}$) in the USDA-ARMS database. More specifically, panel A of Table B.1 shows that the coefficient estimates of the following regression fitted with such data are highly significant\(^{24}\)

(B.1) \[ TNA_{ft} = \gamma_0 + \gamma_1 VP_{ft} + \gamma_2 VP_{ft}^2 + \tau_{ft}. \]

Following Wooldridge (2003), heteroskedasticity in the residual term $\tau_{ft}$ is modeled as $ln(\sigma_{\tau_{ft}}^2) = \delta_0 + \delta_1 VP_{ft} + \delta_2 VP_{ft}^2 + \epsilon_{ft}$; panel B of Table B.1 reports the corresponding estimates.

Based on the parameter estimates for regression (B.1), denoted by hats, 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: $VP_{ft} = y_{ft}^{***} p_{ft}^*$. Endogenous prices $p_{ft}$ are described in Appendix A, and netput quantities $y_{ft}^{***}$ are the solution to problem (2.3) under risk-neutrality. Risk neutrality is assumed at this stage, because solving the expected utility problem requires conditioning on initial wealth, which is what we are trying to calculate. Netput quantities $y_{ft}^{***}$ are only used to compute the firm’s initial wealth, and are not used anywhere else in the analysis.

Step 2: Take a draw from $Normal(0, \hat{\sigma}_e^2)$ to obtain $\epsilon_{ft}$, and use it to calculate $\hat{\sigma}_{\tau_{ft}}^2 = exp(\hat{\delta}_0 + \hat{\delta}_1 VP_{ft} + \hat{\delta}_2 VP_{ft}^2 + \epsilon_{ft})$.

Step 3: Take a draw from a $Normal(0, \hat{\sigma}_{\tau_{ft}}^2)$ for the error term $\tau_{ft}$, and plug it into the analog of (B.1) to obtain initial wealth as $W_{ft,0} = \hat{\gamma}_0 + \hat{\gamma}_1 VP_{ft} + \hat{\gamma}_2 VP_{ft}^2 + \tau_{ft}$.

---

\(^{24}\) $TNA$ is computed as “value of total farm financial assets” minus “total farm financial debt,” and $VP$ is calculated as “all crops – value of production” plus “all livestock – value of production.”
Values of the absolute risk aversion coefficient $\lambda_f$ are computed as the ratio between a relative risk aversion coefficient uniformly distributed in the interval $[2, 4]$ (Pennacchi, 2008 p. 16) and the terminal wealth $W_{f,1}$ (the initial wealth $W_{f,0}$ plus firm- and time-specific profits).
Table B.1. Parameter estimates of regression (B.1), and the form of its heteroskedasticity.

A. Dependent variable: Total Net Assets (TNA)

<table>
<thead>
<tr>
<th>Explanatory Variables:</th>
<th>Region 1</th>
<th>Region 2</th>
<th>Region 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant ($\gamma_0$)</td>
<td>0.724</td>
<td>0.843</td>
<td>0.892</td>
</tr>
<tr>
<td></td>
<td>(0.040)</td>
<td>(0.058)</td>
<td>(0.058)</td>
</tr>
<tr>
<td>$VP$ ($\gamma_1$)</td>
<td>1.279</td>
<td>1.138</td>
<td>0.574</td>
</tr>
<tr>
<td></td>
<td>(0.064)</td>
<td>(0.062)</td>
<td>(0.042)</td>
</tr>
<tr>
<td>$VP^2$ ($\gamma_2$)</td>
<td>-0.066</td>
<td>-0.019</td>
<td>-0.010</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.002)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.160</td>
<td>0.195</td>
<td>0.092</td>
</tr>
</tbody>
</table>

Note: $VP$: Value of Production. Standard errors within parentheses.

B. Dependent variable: $ln(\hat{\sigma}_\tau^2)$, where $\hat{\sigma}_\tau^2$ is the sample estimate of $\sigma^2_{\tau}$

<table>
<thead>
<tr>
<th>Explanatory Variables:</th>
<th>Region 1</th>
<th>Region 2</th>
<th>Region 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant ($\delta_0$)</td>
<td>-2.062</td>
<td>-1.925</td>
<td>-1.5069</td>
</tr>
<tr>
<td></td>
<td>(0.045)</td>
<td>(0.058)</td>
<td>(0.048)</td>
</tr>
<tr>
<td>$VP$ ($\delta_1$)</td>
<td>1.544</td>
<td>0.964</td>
<td>0.416</td>
</tr>
<tr>
<td></td>
<td>(0.071)</td>
<td>(0.063)</td>
<td>(0.035)</td>
</tr>
<tr>
<td>$VP^2$ ($\delta_2$)</td>
<td>-0.105</td>
<td>-0.026</td>
<td>-0.006</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.002)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>$\hat{\sigma}_c^2$</td>
<td>0.160</td>
<td>0.113</td>
<td>0.080</td>
</tr>
<tr>
<td>$R^2$</td>
<td>4.343</td>
<td>4.710</td>
<td>4.040</td>
</tr>
</tbody>
</table>

Note: $VP$: Value of Production. Standard errors within parentheses.
Appendix C. Simulation of Noisy Dataset

The following subsections provide details about the generation of the noisy data, including the procedures used to maximize expected utility, introduce price and production shocks, aggregate across netputs, and incorporate measurement errors.

C.1. Expected Utility Maximization

Uncertainty introduces noise because the duality theorem assumes a deterministic problem whose solution is generally different from the expected utility case. We solve optimization problem (2.3) for the vector of expected variable netput quantities \( y^*_\beta, K^*_\beta, \lambda_{\beta}, W_{\beta,0}; \alpha^*_f \) conditional on expected netput prices, quasi-fixed netput quantities, the levels of absolute risk aversion and initial wealth, and the starting production parameters.

Optimization problem (2.3) is solved by employing numerical methods and Gaussian quadrature. In the present application, the numerical integration must take into account that the objective function is multi-dimensional, and that the uncertainty stems from random variables that have nonstandard distributions and are correlated with each other. Given these requirements, we created a routine to calculate nodes and weights used in the objective function approximation.\(^{25}\) We then used MATLAB’s \textit{fmincon} function to optimize the approximated objective function, by passing the necessary first- and second-order conditions for optimization to the numerical routine as equality and inequality constraints, respectively. The solution is the vector of expected netput quantities for each farm and time, which we denote as \( y^*_\beta \).

C.2. Realized Price and Production Shocks

\(^{25}\) We generated four independent log-normal nodes and weights for each of the three output price random variables using the MATLAB function \textit{qwnorm} (Miranda and Fackler, 2011), that calculates standard Normal nodes and weights. Similarly, based on the function \textit{qwnbeta}, that calculates standard Beta nodes and weights, we computed four independent beta nodes and weights in the interval of interest for the three output quantity random variables. Then, using the Iman and Conover (1982) method, we imposed correlations directly to the nodes (correlation between output prices and quantities equal to -0.30, and correlation within them equal to 0.90); these transformations do not affect the weights.
Farmers solve the maximization problem given a set of output prices that reflects their expectations of harvest prices. It is commonly accepted that prediction errors make this difference relevant. Even in the presence of forward contracts, it might be the case that not all of the production is sold under this type of arrangements. 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. We model this feature by assuming that realized log-prices are equal the log-prices used for optimization plus noise,

\[(C.1) \quad \ln(p) = \ln(p^*) + \epsilon.\]

The \(N\)-vector \(\epsilon\) consists of realizations of \(\text{Normal}(0, \sigma^2)\) random variables with \(\sigma^2 = 0.2^2\) for outputs (Lence, 2009) and \(\sigma^2 = 0.1^2\) for inputs, implying smaller deviations from decision values for inputs than for outputs. Shocks \(\epsilon\) are systematic, in that they affect all firms by the same proportion at a given time.

We also let actual netput quantities \(y^*_\beta\) differ from the optimal quantities that solve problem (2.3) \(y^*_\beta\) (e.g., due to uncertain events in agricultural production, such as weather), as follows:

\[(C.2) \quad y^* = y^*_\beta + g(y^*_\beta) v^*.\]

Shock \(v^*\) is a realization of the random variable controlling production errors \(\psi^\beta\) given by expression (2.5).

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. These correlations are induced by means of the Iman and Conover (1982) method.
C.3. 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 to preserve degrees of freedom, or because they are not the objective of the study. To incorporate this feature, we compute value-weighted aggregates across netput quantities and prices:

(C.3) \[ y_{it} = \sum_{\omega \in \Omega} w_{nft}^i \times y_{nft} \, , \]

(C.4) \[ p_{it} = \sum_{\omega \in \Omega} w_{nft}^i \times p_{nft} \, , \]

where \( \Omega_i \) is the \( i \)th subset of netputs, and \( w_{nft}^i \equiv (p_{nft} \times y_{nft})/(\sum_{\omega \in \Omega} p_{nft} \times y_{nft}) \) is netput \( n \)’s share of subset \( \Omega_i \)’s value.

C.4. 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 this literature, we add noise to the series. This noise is distributed as standard Beta(2, 2), and we calibrate its interval to yield the desired standard deviations of 0.05 around the “true” value for netput prices, 0.08 for variable netput quantities, and 0.10 for quasi-fixed netputs. These standard deviations are smaller than or equal to the ones reported in the literature, especially in the case of prices.