Estimation of Input Distance Functions: A System Approach

Tsionas, Efthymios and Kumbhakar, Subal C. and Malikov, Emir

Athens University of Economics and Business, State University of New York at Binghamton, St. Lawrence University

2015

Estimation of Input Distance Functions: A System Approach*

Efthymios G. Tsionas[1] Subal C. Kumbhakar[2] Emir Malikov[3]

[1]Department of Economics, Athens University of Economics and Business, Athens, Greece
[2]Department of Economics, State University of New York at Binghamton, Binghamton, NY
[3]Department of Economics, St. Lawrence University, Canton, NY

First Draft: September 11, 2012
This Draft: February 17, 2015

Abstract

This article offers a methodology to address the endogeneity of inputs in the input distance function (IDF) formulation of the production processes. We propose to tackle endogenous input ratios appearing in the normalized IDF by considering a flexible (simultaneous) system of the IDF and the first-order conditions from the firm's cost minimization problem. Our model can accommodate both technical and (input) allocative inefficiencies amongst firms. We also present the algorithm for quantifying the cost of allocative inefficiency. We showcase our cost-system-based model by applying it to study the production of Norwegian dairy farms during the 1991–2008 period. Among other things, we find both an economically and statistically significant improvement in the levels of technical efficiency among dairy farms associated with the 1997 quota scheme change, which a more conventional single-equation stochastic frontier model appears to be unable to detect.

Keywords: Cost Minimization, Dairy Production, Dairy Quota, Endogeneity, Input Distance Function, Stochastic Frontier

JEL Classification: C33, D24, Q12

*We would like to thank the editor, Brian Roe, and two anonymous referees for many insightful comments and suggestions that helped improve this article. Any remaining errors are our own. *Email:* tsionas@aueb.gr (Tsionas), kkar@binghamton.edu (Kumbhakar), emalikov@stlawu.edu (Malikov).

Introduction

The identification and estimation[1] of production functions using data on inputs and outputs is among the oldest empirical problems in economics dating back at least as early as the 19th century (Chambers, 1997). While "production function" formulation appears to be a natural framework of modeling single-output production processes, distance functions (Shephard, 1953, 1970) are more popular among researchers in the instances of multiple-output multiple-input production. The primary benefit of modeling the production process via distance functions is that it frees researchers from having to make a judgment call of determining which output (input) is to be treated as a left-hand-side endogenous variable in the production function (input requirement function) since the results are *not* invariant to these choices. However, like the production functions (Marschak and Andrews, 1944), distance functions suffer from an inherent endogeneity problem if inputs and/or outputs are endogenous to the firm's decision-making.[2]

When estimating the input distance function (IDF) researchers explicitly or implicitly assume that outputs are exogenous to the firm's input allocation decisions. Such an assumption is usually justified by either the premise of cost minimizing behavior by firms (e.g., Kumbhakar et al., 2008) or the structure of the industry (e.g., Das and Kumbhakar, 2012). However, the endogeneity[3] of the input ratios, which enter the IDF under the linear homogeneity normalization, is often assumed away with little or no justification (e.g., Lambert and Wilson, 2003; Karagiannis et al., 2004; Atsbeha et al., 2012). Since the assumption of exogenous outputs appears to be somewhat more reasonable[4] than that of exogenous inputs required by the output distance function (ODF) [an alternative to the IDF], in this article we focus on the issues related to the estimation of the IDF only.

In this article, we propose addressing the endogeneity of inputs in the IDF formulation of the production process from the perspective of the economic theory. More specifically, we suggest invoking the assumption of the firm's cost minimizing behavior not only to justify the treatment of outputs as exogenous but to also tackle the endogeneity of the input ratios in the IDF. We do so by augmenting the IDF of a flexible translog form with the set of independent first-order conditions (FOCs) from the cost minimization problem, which we then estimate as a system of simultaneous equations via full-information maximum likelihood (FIML). Our identification strategy relies on competitively determined input prices as a source of exogenous variation. The model we develop also accommodates both technical and (input) allocative inefficiencies amongst firms.

We show that in the presence of allocative disturbances, the input ratios are functions of not only input prices and outputs but also allocative inefficiencies. Consequently, if allocative inefficiencies are correlated with technical inefficiency and/or a random productivity shock in the IDF, the standard single-equation methods applied to the IDF are likely to produce inconsistent estimates of both the production technology and stochastic inefficiency. Even if the input ratios are uncorrelated with the composite error term in the IDF, there are advantages to using a system approach. First, since additional equations (the FOCs) do not contain any extra parameters, the system-based

[1] Throughout this article, we focus on the estimation of *stochastic* production processes only. Our discussion of endogeneity and methods of tackling it does not apply to the fitting of *deterministic* production and/or distance functions by the definition of the latter.

[2] Färe and Primont (1995) deal with theoretical underpinnings of the multi-output multi-input distance function models. However, since they provide little coverage of the estimation of distance functions, the endogeneity issue never arises in their discussion.

[3] That is, the correlation with the error term as a result of the simultaneity of the input allocation.

[4] The cost minimization premise is in line with the economic theory. Furthermore, it may be relatively easier to justify exogeneity of outputs in heavily regulated industries, such as airlines or electric power distribution, since outputs are clearly not firms' choice variables in these instances. Admittedly, the endogeneity in outputs may still continue to persist even in the case of such regulated industries should there be an omitted variable that correlates with output quantities included in the regression equation. We thank an anonymous referee for pointing this out.

estimator is more efficient. Second, technological metrics obtained via the cost system approach are likely to be more meaningful because the economic behavior is embedded into the system through the FOCs.

The endogeneity in the stochastic IDF has also been considered in a somewhat different framework by Atkinson and Primont (2002) and Atkinson et al. (2003). The two papers propose addressing the endogeneity problem using Generalized Method of Moments (GMM). In the case of a single-equation IDF, Atkinson et al. (2003) suggest using input (and possibly output) prices to instrument for inputs (and outputs). Similar to our article, Atkinson and Primont (2002) estimate a cost system of the IDF where, consistent with the cost minimization assumption, input prices and output quantities are instead used as instruments. In practice, the GMM approach is applicable only if technical inefficiency is to be treated as a fixed-effect type inefficiency. That is, one needs to assume technical inefficiency to be either time-invariant or parameterized as a firm-specific function of time in the spirit of Cornwell et al. (1990). In this article, we explore an alternative approach to modeling technical inefficiency, where the latter is treated as a time-varying random effect in the spirit of a stochastic frontier analysis. Hence, unlike Atkinson and Primont's (2002) GMM formulation, we estimate the system of the IDF and cost-minimizing FOCs via FIML.

We note that the estimation of the IDF via a FIML system approach has also been discussed in several related papers. For instance, Karagiannis et al. (2006) consider a system of the IDF along with share equations where allocative inefficiency is defined in terms of shadow input prices. Other studies whose primary interest lies in quantifying allocative errors (as opposed to resolving the endogeneity issue) via a system approach using share equations have been found to still suffer from the endogeneity problem (see Coelli et al., 2008, for an excellent review). Our article differs from the earlier work in at least two respects. As opposed to considering share equations, our primal system contains the FOCs (in explicit form) where we use observed input prices.[5] Further, we model allocative inefficiency in the spirit of Schmidt and Lovell (1978), i.e., in terms of input quantities as opposed to input prices.[6]

Our contribution to the literature is threefold. First, we offer a flexible system approach to a consistent estimation of the IDF based on the (behavioral) cost minimization assumption consistent with the economic theory, while also allowing for a random-effect type technical inefficiency. The model we develop also accommodates allocative inefficiency. Second, we present the methodology for quantifying the *cost* of allocative inefficiency obtained from the primal IDF system. Lastly, we showcase our model by applying it to study the production technology of Norwegian dairy farms during the 1991–2008 period.

Production Technology under Cost Minimization

Consider a production process in which K inputs $\mathbf{X} \in \mathbb{R}_+^K$ are transformed into M outputs $\mathbf{Y} \in \mathbb{R}_+^M$ via the production technology described by

$$Af(\theta\mathbf{X}, \lambda\mathbf{Y}) = 1 \ , \tag{1}$$

where $f(\cdot)$ is the transformation function, and A captures observed and unobserved factors which affect the transformation function neutrally. We specify the production technology in terms of the transformation function $f(\cdot)$ because the latter is more general than the production/distance/input requirement function (e.g., see Caves et al., 1981). Here, scalars $\theta \leq 1$ and $\lambda \geq 1$ capture input

[5]The unavailability of input prices is the primary reason why Karagiannis et al. (2006) and others consider a system of *share* equations, which are *not* based on the cost minimizing FOCs unlike our approach.

[6]This also differentiates our cost system from that considered by Atkinson and Primont (2002).

and output technical inefficiency, respectively. For instance, if $\theta = 0.9$, inputs are said to be 90% efficient, i.e., the use of each input could be reduced by 10% without reducing outputs if inefficiency is eliminated. Similarly, if λ is 1.1, each output could be increased by 10% without increasing the use of any input if inefficiency is eliminated.

Since in general θ and λ are not jointly identified, researchers consider either of the three special cases (depending on the choice of normalization): (i) input-oriented technical inefficiency $\theta < 1$ in the case of $\lambda = 1$; (ii) output-oriented technical inefficiency $\lambda > 1$ in the case of $\theta = 1$; and (iii) hyperbolic technical inefficiency corresponding to the case of $\lambda\theta = 1$ which seeks a simultaneous expansion in outputs and a reduction in inputs at the same rate. All three are related to one another (e.g., see Färe et al., 2002; Kumbhakar, 2013). In this article, we focus on the first case, i.e., identifying and consistently estimating the input-oriented technical inefficiency.

Specifically, defining $\widehat{\mathbf{Y}} \equiv \lambda \mathbf{Y}$ and $\widehat{\mathbf{X}} \equiv \theta \mathbf{X}$, we can rewrite the transformation function (1) in the logarithmic form, i.e.,

$$\ln A + \ln f(\widehat{\mathbf{Y}}, \widehat{\mathbf{X}}) = 0 \ . \tag{2}$$

We assume that $\ln f(\cdot)$ takes the translog functional form, i.e.,

$$\begin{aligned} \ln f(\widehat{\mathbf{Y}}, \widehat{\mathbf{X}}) = & \sum_m \alpha_m \ln \widehat{Y}_m + \frac{1}{2} \sum_m \sum_n \alpha_{mn} \ln \widehat{Y}_m \ln \widehat{Y}_n + \sum_k \beta_k \ln \widehat{X}_k \ + \\ & \frac{1}{2} \sum_k \sum_l \beta_{kl} \ln \widehat{X}_k \ln \widehat{X}_l + \sum_m \sum_k \gamma_{mk} \ln \widehat{Y}_m \ln \widehat{X}_k \ , \end{aligned} \tag{3}$$

which satisfies the following symmetry restrictions: $\beta_{kl} = \beta_{lk}$ and $\alpha_{mn} = \alpha_{nm}$.

Note that we cannot identify, and hence estimate, all parameters in (3) (even after imposing the symmetry restrictions). Specifically, there are $(M + K + 2)$ unidentified parameters. We therefore need to impose identifying restrictions on the transformation function (3). We first rewrite (3) as

$$\begin{aligned} \ln f(\widehat{\mathbf{Y}}, \widehat{\mathbf{X}}) = & \sum_m \alpha_m \ln \widehat{Y}_m + \frac{1}{2} \sum_m \sum_n \alpha_{mn} \ln \widehat{Y}_m \ln \widehat{Y}_n + \sum_{k=2} \beta_k \ln(X_k/X_1) \ + \\ & \frac{1}{2} \sum_{k=2} \sum_{l=2} \beta_{kl} \ln(X_k/X_1) \ln(X_l/X_1) + \sum_m \sum_{k=2} \gamma_{mk} \ln \widehat{Y}_m \ln(X_k/X_1) \ + \\ & \left[\sum_k \beta_k\right] \ln \widehat{X}_1 + \sum_k \left[\sum_l \beta_{kl}\right] \ln \widehat{X}_k \ln \widehat{X}_1 + \sum_m \left[\sum_k \gamma_{mk}\right] \ln \widehat{Y}_m \ln \widehat{X}_1 \ . \end{aligned} \tag{4}$$

Imposing the following $(M + K + 2)$ normalizations on (4):

$$\begin{aligned} & \sum_k \beta_k = 1; \qquad && \sum_l \beta_{kl} = 0 \ \forall \ k \\ & \sum_k \gamma_{mk} = 0 \ \forall \ m; \qquad && \lambda = 1 \end{aligned} \tag{5}$$

yields an identified IDF representation of the production technology (2), i.e.,

$$\begin{aligned} -\ln X_1 = & \alpha_0 + \sum_{k=2} \beta_k \ln \widetilde{X}_k + \frac{1}{2} \sum_{k=2} \sum_{l=2} \beta_{kl} \ln \widetilde{X}_k \ln \widetilde{X}_l + \sum_m \alpha_m \ln Y_m \\ & \frac{1}{2} \sum_m \sum_n \alpha_{mn} \ln Y_m \ln Y_n + \sum_m \sum_{k=2} \gamma_{mk} \ln Y_m \ln \widetilde{X}_k - u + v_1 \ , \end{aligned} \tag{6}$$

where $\widetilde{X}_k \equiv X_k/X_1 \;\forall\; k = 2,\ldots,K$, $\ln A = \alpha_0 + v_1$, and $u \equiv -\ln\theta \geq 0$ measures technical *in*efficiency. Note that the normalizations in (5) except for $\lambda = 1$ are most commonly referred to as the linear homogeneity (in inputs) property of IDF. The remaining restriction $\lambda = 1$ provides an *input*-oriented interpretation of the inefficiency term u.

Cost Minimization with No Allocative Inefficiency

Under the assumption of cost-minimizing behavior, according to which outputs and competitively determined input prices are exogenous to the firm's input allocation, the firm's objective is defined as

$$\min_{\mathbf{X}} \quad \mathbf{W}'\mathbf{X} : \; Af(\widehat{\mathbf{X}}, \widehat{\mathbf{Y}}) = 1 \;, \tag{7}$$

where $\mathbf{W} \in \mathbb{R}_+^K$ is a vector of input prices. The above objective function yields $(K-1)$ independent first-order conditions (FOCs):

$$\frac{W_k X_k}{W_1 X_1} = \frac{\partial \ln f\left(\widehat{\mathbf{Y}}, \widehat{\mathbf{X}}\right)/\partial \ln X_k}{\partial \ln f\left(\widehat{\mathbf{Y}}, \widehat{\mathbf{X}}\right)/\partial \ln X_1} \quad \forall \quad k = 2,\ldots,K \;. \tag{8}$$

Define, for convenience, $\widetilde{X}_k \equiv X_k/X_1$. Given the IDF representation of the production technology in (6), the FOCs (8) take the following form:

$$\frac{W_k}{W_1}\widetilde{X}_k = \frac{\beta_k + \sum_{l=2}\beta_{kl}\ln\widetilde{X}_l + \sum_m \gamma_{mk}\ln Y_m}{1 - \sum_{k=2}\beta_k - \frac{1}{2}\sum_{k=2}\sum_{l=2}\beta_{kl}\left(\ln\widetilde{X}_k + \ln\widetilde{X}_l\right) - \sum_{k=2}\sum_m \gamma_{mk}\ln Y_m} \quad \forall \quad k = 2,\ldots,K \;, \tag{9}$$

which can be rewritten in a more compact form using the identifying restrictions in (5), i.e.,

$$\frac{W_k}{W_1}\widetilde{X}_k = \frac{\beta_k + \sum_{l=2}\beta_{kl}\ln\widetilde{X}_l + \sum_m \gamma_{mk}\ln Y_m}{\beta_1 + \sum_{l=2}\beta_{1l}\ln\widetilde{X}_l + \sum_m \gamma_{m1}\ln Y_m} \quad \forall \quad k = 2,\ldots,K \;. \tag{10}$$

We note that equation (10) by no means implies that parameters β_1, β_{1l} and γ_{1m} are identified and can be estimated. Rather, they can be recovered using the (deterministic) identifying restrictions in (5).

The system of the above FOCs along with the translog IDF in (6) can be used to solve for the input demand functions conditional on outputs. These solutions will be functions of input price ratios and outputs only. Since both $\mathbf{W}$ and $\mathbf{Y}$ are assumed to be exogenous in a cost minimizing framework, the input ratios $\widetilde{X}_k \;\forall\; k = 2,\ldots,K$ are not affected by either the inefficiency u or stochastic productivity shock v_1 (which is absorbed in parameter A). Hence, one can argue in support of treating the regressors in IDF (6) as exogenous (in the absence of allocative disturbances) thereby justifying the use of standard stochastic frontier methods in order to estimate production technology and inefficiency via a single-equation IDF.[7]

Cost Minimization with Allocative Inefficiency

The cost minimization framework can be relaxed to allow for the presence of allocative inefficiencies, i.e., optimization errors. Following Schmidt and Lovell (1978), we augment the FOCs in (10) with

[7]Coelli (2000) reaches the same conclusion.

allocative disturbances, i.e.,

$$\frac{W_k}{W_1}\widetilde{X}_k \times \exp(v_k) = \frac{\beta_k + \sum_{l=2}\beta_{kl}\ln\widetilde{X}_l + \sum_m \gamma_{mk}\ln Y_m}{\beta_1 + \sum_{l=2}\beta_{1l}\ln\widetilde{X}_l + \sum_m \gamma_{m1}\ln Y_m} \quad \forall \quad k = 2, \ldots, K \ , \tag{11}$$

where $v_k \in \mathbb{R} \ \forall \ k = 2, \ldots, K$ represents the allocative inefficiency for the input pair (X_k, X_1).

Unlike in the case of full allocative efficiency (as considered in the previous subsection), the solution to the system of the FOCs in (11) along with the IDF can be shown to yield conditional input demand functions that are functions of not only input price ratios and outputs but also allocative inefficiencies. Consequently, if the allocative disturbances $v_k \ \forall \ k = 2, \ldots, K$ are correlated with technical inefficiency u and/or the productivity shock v_1 in the IDF, the standard stochastic frontier methods applied to a single-equation IDF (6) would produce inconsistent estimates of both the technology and inefficiency.

Even if the input ratios $\widetilde{X}_k \ \forall \ k = 2, \ldots, K$ are uncorrelated with both error terms in the IDF, thereby implying that allocative inefficiencies are uncorrelated with both technical inefficiency and the productivity shock (stochastic noise in the IDF), there are advantages to using a system approach. First, since additional equations (the FOCs) do not contain any extra parameters, the system-based parameter estimates are likely to be more precise. Second, technological metrics obtained based on the estimated IDF are likely to be more meaningful because the economic behavior is embedded into the system through the FOCs. That is, if one believes in the economic behavior of the producers, the FOCs ought to be used in the estimation.

However, firms may pursue economic objectives that differ from the cost minimization. One such example is revenue maximization, under which outputs are endogenous to the firms' decision-making whereas inputs are exogenous. Under this scenario, for a system approach to remain valid, one solely needs to replace inputs with outputs and vice versa. That is, the IDF needs to be replaced with an equivalent ODF based representation of the production process, and the FOCs are to be derived for outputs from the revenue maximization problem. Mechanically, the latter implies switching X and Y in our discussion above.

If both inputs and outputs are endogenous to the firm's decisions, then the endogeneity problem remains no matter whether one estimates a single-equation IDF or the cost system we propose in this article. Given that the joint endogeneity of inputs and outputs corresponds to profit-maximizing behavior by firms, a plausible identification strategy would be to augment the IDF with the FOCs obtained from the profit maximization problem. We do not consider such an instance here, leaving the latter for a different paper.

Econometric Model

We consider the econometric model consisting of the translog IDF (6) and the FOCs with allocative inefficiency in (11). Assuming the existence of panel data and denoting the log values of variables by their lower-case versions (except for the time trend), e.g., $\widetilde{x}_k \equiv \ln\widetilde{X}_k$, the model (system) takes the following form:

$$\begin{aligned}
-x_{1,it} = {} & \alpha_0 + \sum_{k=2}\beta_k\widetilde{x}_{k,it} + \frac{1}{2}\sum_{k=2}\sum_{l=2}\beta_{kl}\widetilde{x}_{k,it}\widetilde{x}_{l,it} + \sum_m \alpha_m y_{m,it} + {} \\
& \frac{1}{2}\sum_m\sum_n \alpha_{mn} y_{m,it} y_{n,it} + \sum_m\sum_{k=2}\gamma_{mk} y_{m,it}\widetilde{x}_{k,it} + {} \\
& \delta_t t + \frac{1}{2}\delta_{tt}t^2 + \sum_{k=2}\varphi_k\widetilde{x}_{k,it}t + \sum_m \psi_m y_{m,it}t - u_{it} + v_{1,it}
\end{aligned} \tag{12a}$$

$$-\widetilde{x}_{k,it} + \ln\left(\beta_k + \sum_{l=2}\beta_{kl}\widetilde{x}_{l,it} + \sum_m \gamma_{mk}y_{m,it} + \varphi_k t\right) - \ln\left(\beta_1 + \sum_{l=2}\beta_{1l}\widetilde{x}_{l,it} + \sum_m \gamma_{m1}y_{m,it} + \varphi_1 t\right)$$
$$= w_{k,it} + v_{k,it} \quad \forall \quad k = 2,\ldots,K \ , \tag{12b}$$

where, for convenience, we have denoted $w_{k,it} = \ln(W_{k,it}/W_{1,it})$; variable t is the time trend to allow for technical change which captures a temporal shift of the production frontier; and subscripts $i = 1,\ldots,N$ and $t = 1,\ldots,T$ index firms and time periods, respectively.

Under the assumptions of competitive factor markets and the cost-minimizing behavior by firms, which are said to minimize cost subject to *given* outputs and input prices, the system in (12) is exactly identified. Specifically, the number of endogenous variables — in our case, an input x_1 and $(K-1)$ input ratios $\widetilde{x}_k \ \forall\ k = 2,\ldots,K$ — equals the number of independent simultaneous equations, i.e., $(K-1)$ FOCs and the IDF itself. Hence, the order condition is satisfied. In addition to exogenous variation generation by K competitively determined input prices $W_k \ \forall\ k = 1,\ldots,K$, the distributional assumption about a K-variate vector of stochastic disturbances $\mathbf{v}_{it}$ provides an additional source of identification. Our identification strategy can also be explained in a more conventional instrumental variable (IV) framework. In particular, K endogenous inputs $\mathbf{X}$ are being instrumented for by K exogenous input prices $\mathbf{W}$, whereas exogenous outputs $\mathbf{Y}$ and a time trend t instrument for themselves.

Estimation

We estimate the cost system in (12) using the FIML method. To implement it, we assume

$$\mathbf{v}_{it} = [v_{1,it}, v_{2,it},\ldots,v_{K,it}]' \equiv \left[v_{1,it}, \mathbf{v}_{it}^{*\prime}\right]' \sim \text{i.i.d. } \mathbb{N}(0,\boldsymbol{\Sigma}) \ . \tag{13}$$

The first issue we need to deal with during the estimation is the non-unit Jacobian of the transformation when $\mathbf{x}_{it} = [x_{1,it}, \widetilde{x}_{2,it},\ldots,\widetilde{x}_{K,it}]' \equiv [x_{1,it}, \widetilde{\mathbf{x}}_{it}{}']'$ are the endogenous variables. Then the Jacobian of the transformation from $\mathbf{v}_{it}$ to $\mathbf{x}_{it}$ is equal to the Jacobian from $\mathbf{v}_{it}^*$ to $\widetilde{\mathbf{x}}_{it}$ since $\frac{\partial v_{1,it}}{\partial x_{1,it}} = -1$, and $\frac{\partial v_{k,it}}{\partial x_{1,it}} = 0 \ \forall\ k = 2,\ldots,K$. To express the Jacobian in a more compact form, we define

$$\left[\beta_k + \sum_{l=2}\beta_{kl}\widetilde{x}_{l,it} + \sum_m \gamma_{mk}y_{m,it} + \varphi_k t\right]^{-1} \equiv A_{k,it},\ k = 2,\ldots,K; \qquad \mathbf{a}_{it}^* \equiv [A_{2,it},\ldots,A_{K,it}]'$$
$$\mathbf{B}^* \equiv [\beta_{kl};\ k,l = 2,\ldots,K]; \qquad \mathbf{b} \equiv [\beta_{1k};\ k = 2,\ldots,K]$$
$$\mathbf{Q}_{it} \equiv \mathbf{a}_{it}^* \otimes \boldsymbol{\iota}_{K-1}'; \qquad \mathbf{R} \equiv \boldsymbol{\iota}_{K-1} \otimes \mathbf{b}' \ ,$$

where $\boldsymbol{\iota}_{K-1}'$ is the row vector of ones with $(K-1)$ elements.
Using the above notation we can write the Jacobian as

$$J_{it}(\boldsymbol{\theta}, \mathbf{x}_{it}, \mathbf{y}_{it}) = \|\mathbf{B}^*\mathbf{Q}_{it} - A_{1,it}\mathbf{R} - \mathbf{I}_{K-1}\| \ , \tag{14}$$

where $\mathbf{I}_{K-1}$ is the identity matrix of a $(K-1)$ dimension. While the Jacobian is easy to compute in this form, it is still a function of all parameters as well as a function of the data.

The density function of endogenous variables $\mathbf{x}_{it}$ conditional, among other things, on technical inefficiency u_{it} can be written as

$$f(\mathbf{x}_{it}|\boldsymbol{\theta}, \boldsymbol{\Sigma}, u_{it}, \boldsymbol{\Xi}) = (2\pi)^{-K/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\left\{-\frac{1}{2}\boldsymbol{\xi}_{it}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\xi}_{it}\right\} \times J_{it}(\boldsymbol{\theta}, \mathbf{x}_{it}, \mathbf{y}_{it}) \ , \tag{15}$$

where $\boldsymbol{\xi}_{it} \equiv \mathbf{v}_{it} - \begin{bmatrix} u_{it} \\ \mathbf{0}_{K-1} \end{bmatrix}$, $\boldsymbol{\Xi}$ denotes the data (including input prices), and $\boldsymbol{\theta}$ collectively denotes the parameters α_m, α_{mn}, β_k, β_{kl}, γ_{km} and φ_k $\forall\ k, l = 2, \ldots, K;\ m = 1, \ldots, M$.
Next, we write the system in (12) in the following general notation:

$$\mathbf{F}_{it}(\boldsymbol{\theta}) = \mathbf{v}_{it} - u_{it}\boldsymbol{j}_K \equiv \boldsymbol{\xi}_{it}\ , \tag{16}$$

where $\boldsymbol{j}_K = [1, 0, \ldots, 0]$ is a $(K \times 1)$ vector, and the elements of $\mathbf{F}_{it} = [F_{1,it}, F_{2,it}, \ldots, F_{K,it}]'$ are given by

$$\begin{aligned} F_{1,it} \equiv\ & -x_{1,it} - \alpha_0 - \sum_{k=2} \beta_k \widetilde{x}_{k,it} - \frac{1}{2}\sum_{k=2}\sum_{l=2} \beta_{kl}\widetilde{x}_{k,it}\widetilde{x}_{l,it} - \sum_m \alpha_m y_{m,it} - \\ & \frac{1}{2}\sum_m\sum_n \alpha_{mn} y_{m,it} y_{n,it} - \sum_m\sum_{k=2} \gamma_{mk} y_{m,it}\widetilde{x}_{k,it} - \\ & \delta_t t - \frac{1}{2}\delta_{tt}t^2 - \sum_{k=2}\varphi_k \widetilde{x}_{k,it} t - \sum_m \psi_m y_{m,it} t \end{aligned} \tag{17a}$$

$$\begin{aligned} F_{k,it} \equiv\ & -\widetilde{x}_{k,it} + \ln\left(\beta_k + \sum_{l=2}\beta_{kl}\widetilde{x}_{l,it} + \sum_m \gamma_{mk} y_{m,it} + \varphi_k t\right) - \\ & \ln\left(\beta_1 + \sum_{l=2}\beta_{1l}\widetilde{x}_{l,it} + \sum_m \gamma_{m1} y_{m,it} + \varphi_1 t\right) - w_{k,it} \quad \forall \quad k = 2, \ldots, K\ . \end{aligned} \tag{17b}$$

Formal integration with respect to u_{it} in (15) produces a density that will be unconditional of technical inefficiency. For the sake of exposition, suppose $u_{it} \sim$ i.i.d. $\mathbb{N}_+\left(0, \sigma_u^2\right)$ independently of the data and the elements of $\mathbf{v}_{it}$.[8] Then, it is straightforward to show that

$$\begin{aligned} f\left(\mathbf{x}_{it}|\boldsymbol{\theta}, \boldsymbol{\Sigma}, \boldsymbol{\Xi}\right) = 2(2\pi)^{-K/2}\left|\boldsymbol{\Sigma}\right|^{-1/2}\left(\sigma_u\sigma_*\right)^{-1}\Phi\left(\widehat{u}_{it}/\sigma_*\right) \times \\ \exp\left\{-\frac{1}{2}\left(\mathbf{F}_{it}'\boldsymbol{\Sigma}^{-1}\mathbf{F}_{it} - \sigma_*^2\left(\mathbf{F}_{it}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\iota}_K\right)^2\right)\right\} \times J_{it}\left(\boldsymbol{\theta}, \mathbf{x}_{it}, \mathbf{y}_{it}\right), \end{aligned} \tag{18}$$

where $\sigma_*^2 \equiv \frac{\sigma_u^2}{1+\sigma_u^2\boldsymbol{\iota}_K'\boldsymbol{\Sigma}^{-1}\boldsymbol{\iota}_K}$, $\widehat{u}_{it} \equiv -\sigma_*^2\mathbf{F}_{it}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\iota}_K$, and $\Phi(\cdot)$ is the standard normal distribution function. Note that $\mathbf{F}_{it}'\boldsymbol{\Sigma}^{-1}\mathbf{F}_{it}$ is a familiar "sum of squares" from a standard treatment of the nonlinear simultaneous equations model or nonlinear seemingly unrelated regressions.

The density in (18) is used to obtain the likelihood function which is maximized with respect to $(\boldsymbol{\theta}, \boldsymbol{\Sigma})$. In (18), the term $\mathbf{F}_{it}'\boldsymbol{\Sigma}^{-1}\mathbf{F}_{it} - \sigma_*^2\left(\mathbf{F}_{it}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\iota}_K\right)^2 = \mathbf{F}_{it}'\boldsymbol{\Sigma}^{-1}\mathbf{M}\mathbf{F}_{it}$, where $\mathbf{M} \equiv \mathbf{I}_K - \sigma_*^2\mathbf{J}_K\boldsymbol{\Sigma}^{-1}$ and $\mathbf{J}_K$ is a square matrix of dimension K, all elements of which are equal to unity. Although formal concentration with respect to $\boldsymbol{\Sigma}^{-1}$ is not feasible, we can use the Cholesky decomposition $\boldsymbol{\Sigma}^{-1} = \mathbf{C}'\mathbf{C}$ and treat the non-zero elements of $\mathbf{C}$ below the main diagonal as unrestricted parameters.

Further, we can show that $u_{it}|\ \boldsymbol{\theta}, \boldsymbol{\Sigma}, \boldsymbol{\Xi} \sim \mathbb{N}_+\left(\widehat{u}_{it}, \sigma_*^2\right)$ for which we extend the Jondrow et al. (1982) formula to obtain observation-specific estimates of technical inefficiency $\mathbb{E}\left[u_{it}|\boldsymbol{\theta}, \boldsymbol{\Sigma}, \boldsymbol{\Xi}\right]$ which we can compute using properties of the truncated normal distribution.

In our discussion of the system approach to the (consistent) estimation of the IDF, we model technical inefficiency u_{it} as a random-effect type inefficiency in the spirit of the stochastic frontier analysis (Kumbhakar and Lovell, 2000). However, given the IV interpretation of our identification strategy outlined in the previous subsection, one may wonder if system (12) can be alternatively

[8]In our application, we consider a more general case of $u_{it} \sim$ i.i.d. $\mathbb{N}_+\left(\mu_{u,it}, \sigma_u^2\right)$ allowing the mean of inefficiency to vary with the dairy quota regime in Norway. For more details, see application.

estimated via GMM as done in Atkinson and Primont (2002). The latter is meaningful only if u_{it} is to be treated as a fixed-effect type technical inefficiency. That is, u_{it} is to be assumed to be either time-invariant ("the" fixed effect) or parameterized as a firm-specific function of time in the spirit of Cornwell et al. (1990). For instance, Atkinson and Primont (2002) pursue the latter route. The random-effect treatment of (one-sided) technical inefficiency would however require a two-stage GMM procedure, in which u_{it} is to be recovered via maximum likelihood (ML) under the distributional assumption about stochastic disturbances. The two stages can surely be combined into a one-step multiple-equation GMM framework by augmenting the IDF by the log-likelihood score functions from the second stage. Regardless whether a one– or two-step GMM procedure is employed, it would still require a distributional assumption about u_{it} and the random productivity shocks. Given the asymptotic equivalence of ML and GMM under this scenario, employing ML has the benefit of superior efficiency. Hence, in the stochastic frontier framework (as in this article), the use of FIML which we opt for seems to be optimal.

Computing Cost of Allocative Inefficiency

Given system (12) where we allow for allocative disturbances, we can also compute the *cost* of allocative inefficiency (CAI) which we define as the predicted difference between actual and frontier cost, computed as a fraction of the predicted frontier cost. Both predicted actual and frontier costs are "optimal" in the sense of being associated with the firm's optimal input allocation. Predicted frontier cost is the cost of (predicted) optimal input quantities in the *absence* of allocative inefficiencies, while predicted actual cost is the cost of optimal input quantities in the *presence* of allocative inefficiencies. In this section, we show how to compute these optimal values of input quantities with and without allocative inefficiency which we then use to evaluate CAI.

Suppressing the (i,t) subscripts for notational convenience, we rewrite the set of the FOCs in (12b) as

$$\ln\left(c_k + \sum_{l=2} \beta_{kl}\widetilde{x}_l\right) - \ln\left(c_1 + \sum_{l=2} \beta_{1l}\widetilde{x}_l\right) = \widetilde{x}_k + w_k + v_k \quad \forall \quad k = 2, \ldots, K \ , \tag{19}$$

where $c_\kappa \equiv \beta_\kappa + \sum_m \gamma_{m\kappa} y_m + \varphi_\kappa t$ for $\kappa = 1, ..., K$. The (sub)system in (19) may be written as

$$G_k\left(\widetilde{\mathbf{x}}\right) = \widetilde{x}_k + w_k + v_k \quad \forall \quad k = 2, \ldots, K \tag{20}$$

or

$$\mathbf{G}\left(\widetilde{\mathbf{x}}\right) = \widetilde{\mathbf{x}} + \mathbf{w}^* + \mathbf{v}^*, \ \mathbf{G} : \mathbb{R}^{K-1} \to \mathbb{R}^{K-1} \ , \tag{21}$$

where $\mathbf{G}(\cdot) = [G_2(\cdot), \ldots, G_K(\cdot)]'$ the elements of which are the left-hand sides of the FOCs in (19), and $\mathbf{w}^* = [w_2, \ldots, w_K]'$.

We next show the existence and uniqueness of an optimal input-allocation solution. We first derive the Jacobian matrix for the system in (21): $\mathcal{J} = [\partial G_k\left(\widetilde{\mathbf{x}}\right)/\partial \widetilde{x}_l;\ k, l = 2, \ldots, K]$. From (19), it follows that $\partial G_k\left(\widetilde{\mathbf{x}}\right)/\partial \widetilde{x}_l = [\widehat{s}_1\widehat{s}_k]^{-1}\left(\beta_{kl}\widehat{s}_1 - \beta_{1l}\widehat{s}_k - \delta_{kl}\widehat{s}_1\widehat{s}_k\right)$, where $\widehat{s}_\kappa = c_\kappa + \sum_{l=2}\beta_{\kappa l}\widetilde{x}_l$ for $\kappa = 1, \ldots, K$, and δ_{kl} is the Kronecker *delta*. Further, using the normalizing restrictions in (5) we can show that $\sum_{k=2}\exp\left\{G_k\left(\widetilde{\mathbf{x}}\right)\right\} = 1$.[9] Hence, we can sum all FOCs in (19) in exponential form to obtain

$$\frac{1 - \left(c_1 + \sum_{l=2}\beta_{1l}\widetilde{x}_l\right)}{c_1 + \sum_{l=2}\beta_{1l}\widetilde{x}_l} = \sum_{l=2}\exp\left\{\widetilde{x}_l + w_l + v_l\right\} \ . \tag{22}$$

[9] The equality is evident from equation (9).

Dividing (19), after exponentiating, by the above expression yields

$$g_k(\widetilde{\mathbf{x}}) \equiv \frac{c_k + \sum_{l=2} \beta_{kl}\widetilde{x}_l}{1 - \left(c_1 + \sum_{l=2} \beta_{1l}\widetilde{x}_l\right)} = \frac{\exp\{\widetilde{x}_k + w_k + v_k\}}{\sum_{l=2} \exp\{\widetilde{x}_l + w_l + v_l\}} \quad \forall \quad k = 2, \ldots, K \ . \tag{23}$$

We use (23) to show the existence and uniqueness of a solution to the cost minimization system. Specifically, since $g_k(\widetilde{\mathbf{x}}) = \frac{c_k + \sum_{l=2} \beta_{kl}\widetilde{x}_l}{\sum_{k=2}\left(c_k + \sum_{l=2} \beta_{kl}\widetilde{x}_l\right)}$, we have that $g_k : \mathbb{R}^{K-1} \to (0,1)$. Showing that $g_k(\widetilde{\mathbf{x}})$ attains any value in $(0,1)$ would enable us to claim that there exists an optimal input-allocation solution.

Suppose $\boldsymbol{\tau} = [\tau_2, \ldots, \tau_K]' \in (0,1)^{K-1}$. Then, consider a system $g_k(\widetilde{\mathbf{x}}) = \tau_k$ for $k = 2, \ldots, K$, which can be written in the form of $\mathbf{D}\widetilde{\mathbf{x}} = \mathbf{p}$, where $\mathbf{D} = [a_{kl}]$ and $\mathbf{p} = [b_k]$ with $a_{kl} = \beta_{kl} + \tau_k\beta_{1l}$, $b_k = \tau_k(1 - c_1) - c_k$ for $k, l = 2, \ldots, K$. Equations $\sum_{l=2} a_{kl}\widetilde{x}_l = b_k$ are equivalent to $\sum_{l=2} \beta_{kl}\widetilde{x}_l = \frac{\tau_k(1-c_1)-c_k}{1+\tau_k}$ for $k = 2, \ldots, K$ after substituting the restrictions. This system has a *unique* solution provided the $(K-1)\times(K-1)$ matrix $\mathbf{B}^* = [\beta_{kl}, \ k, l = 2, \ldots, K]$ is invertible. In fact, this is so by the translog regularity conditions. This concludes a proof of the existence of a unique input-allocation solution. We next proceed to the discussion of how to compute it.

Let $\widetilde{\mathbf{x}}(\mathbf{v}^*)$ denote the solution to (19). We can approximate $\widetilde{\mathbf{x}}(\mathbf{v}^*)$ using a Taylor expansion as $\widetilde{\mathbf{x}}(\mathbf{v}^*) \simeq \widetilde{\mathbf{x}}(\mathbf{0}_{K-1}) + \nabla\widetilde{\mathbf{x}}(\mathbf{0}_{K-1})\,\mathbf{v}^*$, where $\mathbf{0}_{K-1}$ denotes the zero vector of dimension $(K-1)$, and $\widetilde{\mathbf{x}}(\mathbf{0}_{K-1})$ is a solution to the system (19) at $\mathbf{v}^* = \mathbf{0}_{K-1}$ (i.e., in the absence of allocative inefficiencies) which exists and is unique as we have shown above.

Denote the system in (19) as $\mathbf{H}(\widetilde{\mathbf{x}}) = \mathbf{v}^*$ so that the solution satisfies the identity: $\mathbf{H}(\widetilde{\mathbf{x}}(\mathbf{v}^*)) = \mathbf{v}^*$. Differentiating the latter identity with respect to $\mathbf{v}^*$ we obtain: $\nabla\mathbf{H}(\widetilde{\mathbf{x}}(\mathbf{v}^*)) \times \nabla\widetilde{\mathbf{x}}(\mathbf{v}^*) = \mathbf{I}_{K-1}$, from which it follows that $\nabla\widetilde{\mathbf{x}}(\mathbf{v}^*) = \nabla\mathbf{H}(\widetilde{\mathbf{x}}(\mathbf{v}^*))^{-1}$. It further follows that $\nabla\widetilde{\mathbf{x}}(\mathbf{0}_{K-1}) = \nabla\mathbf{H}(\widetilde{\mathbf{x}}(\mathbf{0}_{K-1}))^{-1}$ assuming that the Jacobian of the system $\mathcal{J}_0 = \nabla\mathbf{H}(\widetilde{\mathbf{x}}(\mathbf{0}_{K-1}))$ is non-singular. Therefore, we have

$$\widetilde{\mathbf{x}}(\mathbf{v}^*) \simeq \widetilde{\mathbf{x}}(\mathbf{0}_{K-1}) + \mathcal{J}_0^{-1}\mathbf{v}^* \ . \tag{24}$$

From (24), it follows that the input distortions are

$$\boldsymbol{\zeta}^* \triangleq \widetilde{\mathbf{x}}(\mathbf{v}^*) - \widetilde{\mathbf{x}}(\mathbf{0}_{K-1}) \simeq \mathcal{J}_0^{-1}\mathbf{v}^* \tag{25}$$

to the first order of approximation around the expected value of $\mathbf{v}^*$, where $\boldsymbol{\zeta}^* = [\zeta_2, \ldots, \zeta_K]'$. Since the Jacobian is already available from the FIML estimation, the additional computation involved in (25) is minimal. Also note that one does not need to obtain $\widetilde{\mathbf{x}}(\mathbf{v}^*)$ separately for $\mathbf{v}^* \neq \mathbf{0}_{K-1}$ and $\mathbf{v}^* = \mathbf{0}_{K-1}$, since we can obtain their difference $\boldsymbol{\zeta}^*$ directly from (25).

We also obtain $\mathbf{v}^* = \mathcal{J}_0\boldsymbol{\zeta}^*$ from (25), whereas from (24) we can obtain $\widetilde{\mathbf{x}}(\mathbf{v}^*)$ provided we have $\widetilde{\mathbf{x}}(\mathbf{0}_{K-1})$. We can get the latter by solving (21) for $\widetilde{\mathbf{x}}$ at $\mathbf{v}^* = \mathbf{0}_{K-1}$, i.e.,

$$\mathbf{G}\big(\widetilde{\mathbf{x}}(\mathbf{0}_{K-1})\big) - \mathbf{w}^* = \widetilde{\mathbf{x}}(\mathbf{0}_{K-1}) \ . \tag{26}$$

We solve system (26) using a fixed-point iteration starting from the observed value of $\widetilde{\mathbf{x}}$. We obtain $x_1(\mathbf{v}^*)$ and $x_1(\mathbf{0}_{K-1})$ from (24) using the IDF in (12a) by a simple substitution assuming $u = 0$. This provides us with $\zeta_1 = x_1(\mathbf{v}^*) - x_1(\mathbf{0}_{K-1})$, i.e., the allocative distortion of the *numeraire* input. We then compute the (observation-specific) cost of allocative inefficiency as

$$\text{CAI} = \sum_k w_k\zeta_k \ . \tag{27}$$

Table 1: Summary statistics for Norwegian dairy production, 1991–2008

Variable	Units	Mean	Median	SD	2.5%	97.5%
Outputs						
Milk	*Liters*	97,414.12	88,216.51	49,248.58	33,751.70	227,419.40
Other Outputs	*Real EUR*	43,917.51	39,552.45	22,182.90	17,026.88	99,786.35
Inputs						
Land	*Hectares (ha)*	21.75	19.40	11.32	7.50	49.90
Feed	*Real EUR*	15,819.04	13,877.61	9,312.38	4,787.15	37,570.52
Labor	*Hours (hr)*	3,661.18	3,515.00	1,127.00	1,922.70	6,327.27
Materials	*Real EUR*	10,665.31	9,572.84	5,655.30	3,492.88	24,187.41
Capital	*Real EUR*	26,607.75	24,021.42	13,905.56	9,250.16	61,869.77
Input Prices						
Price of Land	*Real EUR/ha*	128.601	117.322	71.876	36.777	334.330
Price of Feed	*Real EUR*	1.063	1.016	0.095	0.990	1.335
Price of Labor	*Real EUR/hr*	14.151	13.749	2.538	10.615	18.561
Price of Materials	*Real EUR*	1.036	1.0001	0.089	0.944	1.270
Price of Capital	*Real EUR*	0.984	0.972	0.099	0.830	1.164

Data

We focus on the analysis of dairy farms in Norway. The data are a large unbalanced panel with 7,866 observations on 1,104 farms observed during the 1991–2008 period. Our farm-level data come from the Norwegian Farm Accountancy Survey administered by the Norwegian Agricultural Economics Research Institute. The survey includes the information on farm production and economic data collected annually from about 1,000 farms from different regions, farm size classes and types of farms. Participation in the survey is voluntary. There is no limit on the number of years a farm may be included in the survey. Approximately 10% of the farms surveyed are replaced every year. The farms are classified according to their main category of farming, defined in terms of the standard gross margins of the farm. Hence, the dairy farms in our sample are the farms, the largest share of the total standard gross margin of which is attributed to dairy production.

Dairy farms are often involved not only in the production of milk but also in other farm production activities such as the production of various types of meat, crop, etc. Following Sipiläinen et al. (2014), we consider two outputs: Y_1 – milk, measured in liters sold, and Y_2 – a single measure of all other outputs, which includes cattle and crop products. Since Y_2 includes several outputs we measure it in monetary value terms, i.e., revenue from all these outputs. To covert the nominal value into the real terms, we first deflate Y_2 to real 2000 Norwegian Kroner (NOK) using a weighted price index for cattle and crops, which we then convert from NOK to Euros (EUR) using the average exchange rate. Further, we specify the following five inputs: X_1 – land, measured in hectares, X_2 – own and hired farm labor, measured in hours, X_3 – purchased feed, X_4 – materials, which include the cost of fertilizer, pesticides, preservatives, cost related to animal husbandry, etc., and X_5 – physical capital, which includes farm machinery. Similar to Y_2, purchased feed, materials and capital are measured in real 2000 EUR. All are deflated using a respective price index. The information on input prices $W_k \ \forall \ k = 1, \ldots, 5$ is taken from either the farm survey when available (i.e., the price of land follows the regional average rents, the price of labor is obtained from the value of paid labor) or the agricultural sector of the national accounts. Table 1 reports summary statistics for the variables used in our analysis.

Dairy farming in Norway is highly regulated (Jervell and Borgen, 2000). Throughout the past decades, various regulatory schemes have been set up to align aggregate milk production to domestic demand. A quota-based regulatory scheme was set up in 1983. From 1991, quotas of dairy farmers exiting the industry were used to reduce national milk supply, and there was no redistribution of quotas. The individual quotas have been reduced on several occasions in order to adjust total supply to domestic demand. The exit rate from the dairy sector was however slow for many years. Partly as a reaction to this outcome, a limited quota-trading scheme with quota prices defined by the government as well as the administrative reallocation of quotas was introduced in 1997. The objective of the change was to introduce greater flexibility to the quota system and to encourage structural changes such as higher efficiency and productivity. To maintain the regional distribution of production, the country was divided into milk-trade regions, and quota transfers were restricted to a given region only. However, for many years, only a small fraction of the milk quota was reallocated due to the reduction in the total quota. Since 2002, a farmer interested in selling (a portion of) quota had to sell a portion to the government to be reallocated.

Empirical Results

We start by examining whether endogeneity is indeed a problem in our application. To do so, we consider the IDF in (12a) alone [without the FOCs in (12b)]. For the time being, we also abstract away from the technical inefficiency term in the equation, provided the latter is distributed independently from the data (right-hand-side covariates). A common method to test for endogeneity of the covariates in the IDF is to employ a Wu-Hausman test performed on the difference between the OLS and two-stage least squares (2SLS) [or IV] estimates of the IDF parameters. To obtain the 2SLS estimates of the IDF, one would need to first obtain the fitted values of $\widetilde{x}_{k,it}\ \forall\ k = 2, \ldots, K$ from the respective reduced form equations (in terms of exogenous $y_{m,it}\ \forall\ m = 1, \ldots, M$, $w_{k,it}\ \forall\ k = 2, \ldots, K$ and t), which would then be used to replace $\widetilde{x}_{k,it}$ in (12a). However, since the IDF and the reduced-form equations are nonlinear it is unclear whether this is the best approach. In fact, Terza et al. (2008) advocate for a more robust alternative: a two-stage residual inclusion (2SRI) estimation of Hausman (1978).[10]

In accordance with 2SRI, the first-stage residuals from the reduced-form equations for each of the potentially endogenous covariates are to be included as additional regressors while keeping all original covariates intact (i.e., no replacement with first-stage fitted values). Intuitively, the first-stage residuals are the estimates of all the other factors that affect the endogenous variables in addition to the exogenous instruments that are explicitly accounted for. Those factors are also likely to appear in the random error of the main equation (here, the IDF). Hence, by including these residuals in the IDF, we are able to control for the correlation between endogenous regressors and the random error thereby, essentially, avoiding the omitted variable problem. The test for endogeneity then boils down to a joint F-test on the first-stage residuals. In the case of linear models, 2SRI and 2SLS are equivelent; in *nonlinear* models (like ours) 2SRI is generically consistent whereas 2SLS is not.

The four included residuals from the reduced-form equations for $\widetilde{x}_k\ \forall\ k = 2, \ldots, 5$ are highly significant both individually and jointly. The joint exclusion F-statistic is 48.69. Since the test is not exact, we use pair panel bootstrap to obtain critical values. Specifically, we use 10,000 random draws (with replacement) of cross-sections from our data to reestimate both stages and recompute the F-statistic for the joint exclusion of the first-stage residuals from the IDF. The 99% bootstrap critical value, which we compute as the 99th percentile of bootstrap distribution of the F-statistic,

[10] The 2SRI procedure is similar in its nature to Petrin and Train's (2010) control function approach to handling endogeneity in consumer choice models.

Table 2: Estimates of selected technological metrics

Metric	Mean	Median	SD	2.5%	97.5%
Single-Equation IDF					
SE	0.5801	0.5754	0.2839	0.0324	1.1340
TC	0.0125	0.0127	0.0199	−0.0268	0.0518
Tech. Ineff.	0.0238	0.0213	0.0166	0.0009	0.0604
Cost System of IDF					
SE	0.8486	0.8480	0.0740	0.7058	0.9932
TC	0.0132	0.0131	0.0199	−0.0258	0.0530
Tech. Ineff.	0.0523	0.0429	0.0402	0.0019	0.1451
CAI	0.0953	0.0690	0.0801	0.0040	0.3019

is 3.21. As a robustness check, we repeat the test while also controlling for fixed effects in the IDF. The corresponding F-statistic is even larger and equals 97.10. Thus, the data lend strong support in favor of endogeneity present in the IDF.[11]

We next proceed to the discussion of the main results. Instead of looking at the parameter estimates, we rather focus on more meaningful technological metrics such as scale economies (SE), technical change (TC), technical inefficiency and cost of allocative inefficiency (CAI). Observation-specific estimates of technical inefficiency are obtained from the estimated conditional mean of u in the spirit of Jondrow et al. (1982), and the estimates of CAI are obtained as described earlier. Measures of scale economies and technical change are computed based on the estimated IDF, the available data and the duality results (Färe and Primont, 1995; Karagiannis et al., 2004) as follows:

$$\mathrm{SE} \equiv \sum_m \frac{\partial \ln C}{\partial \ln Y_m} = -\sum_m \frac{\partial \ln f(\widehat{\mathbf{Y}}, \widehat{\mathbf{X}}, t)}{\partial \ln Y_m} = -\sum_m \left[\alpha_m + \sum_n \alpha_{mn} y_n + \sum_{k=2} \gamma_{mk} \widetilde{x}_k + \psi_m t \right]$$

$$\mathrm{TC} \equiv -\frac{\partial \ln C}{\partial t} = \frac{\partial \ln f(\widehat{\mathbf{Y}}, \widehat{\mathbf{X}}, t)}{\partial t} = \delta_t + \delta_{tt} t + \sum_{k=2} \varphi_k \widetilde{x}_k + \sum_m \psi_m y_m \ .$$

Scale economies are defined as the sum of output elasticities of cost thus capturing an increase in the cost (in %) resulting from a proportional one-percent increase in all outputs. The instance of SE $>$ 1 corresponds to decreasing returns to scale, while SE $<$ 1 indicates increasing returns to scale. Lastly, technical change is defined as a secular decline in cost over time.

We estimate two models: *(i)* the cost system (12) via FIML and *(ii)* a single-equation IDF (12a) via ML. We estimate a single-equation model, which suffers from simultaneity, in order to examine the degree to which results get distorted should the endogeneity of the input ratios remain unaddressed. We however acknowledge that this comparison is meaningful only if our identification strategy is valid.

As discussed earlier, dairy farming in Norway is regulated via a quota-based scheme, which underwent a (policy) regime change in the late 1990s. Specifically, a quota-trading system was introduced in 1997 with the objective of encouraging structural changes that would lead to higher efficiency and productivity in this agricultural sector. To investigate whether the farm-level technical inefficiency has responded to this regime change,[12] when estimating the system in (12) or a single-equation IDF in (12a) we model inefficiency u_{it} in a way that permits its mean to vary with the

[11] We note that we are unable to test for exogeneity of the input price ratios and/or outputs (our instruments) since our model is exactly identified, and the test for over-identifying restrictions would require additional instruments.

[12] We thank the editor and one of the referees for this insightful suggestion.

quota regime. Formally, we assume that $u_{it} \sim \mathbb{N}_+ \left(\mu_{u,it}, \sigma_u^2\right)$ with the mean $\mu_{u,it}$ parameterized as $\mu_{u,it} = \eta_0 + \eta_1 H_{it}$, where H_{it} is an indicator which equals one for years from 1997 onward and zero otherwise.[13]

Table 2 reports summary statistics of the technological metrics computed from both the cost system of the IDF and a single-equation IDF. In order to not confine our analysis to mean estimates only (which are not that informative), we also provide the distributions for each of the metrics in figures 1 and 2. The kernel densities are constructed using the second-order Epanechnikov kernel with the optimal bandwidth selected via the data-driven least-squares cross-validation (Silverman, 1986).

Comparing the scale economies estimates from the system and single-equation IDF model, we find that a single-equation approach commonly employed in the literature (e.g., Lambert and Wilson, 2003; Karagiannis et al., 2004; Atsbeha et al., 2012), which takes endogeneity of the input ratios for granted, produces unreasonable estimates of SE. In particular, the mean single-equation-based estimate of SE is 0.58 with the corresponding 2.5%–97.5% interval ranging from 0.03 to 1.13. That is, dairy farms are predicted, on average, to enjoy increasing returns to scale of an astounding magnitude 1.72(=1/0.58). Furthermore, about 2% of the single-equation-based estimates of SE are negative (see figure 1) as a result of the violated IDF regularity conditions. In contrast, the system-based estimates of SE are uniformly positive and suggest the presence of increasing returns to scale (i.e., mostly SE < 1) of reasonable magnitudes. For instance, the median system-based estimate of SE is 0.85 with the standard deviation of 0.07, which is significantly larger than more dispersedly distributed estimates from the single-equation model. To sum, empirical evidence suggests that increasing the scale of farms may considerably improve farms' performance.

In stark contrast to the results on scale economies, the estimates of TC are virtually indistinguishable across the two models. Both the system and single-equation models suggest that the Norwegian dairy sector has experienced technical change (progress) at an average rate of 1.3% per annum over the course of the 1991–2008 sample period. The Li (1996) test confirms the equality of the two empirical distributions of the TC estimates.

We next consider empirical results on the cost of allocative inefficiency. Figure 2 plots the distribution of the CAI estimates obtained from the cost system (12) estimated via FIML. The distribution is highly skewed to the right. The CAI estimates average at 9.5% and range from 0.4% to 30.2% (see table 2). These estimates indicate economically significant misallocation of inputs which, if eliminated, may potentially reduce farms' costs by the median of 6.9%. Also note that no CAI estimates are obtained from the single-equation IDF model since the latter does not allow the identification of allocative disturbances due to the omission of the cost minimizing FOCs.

We conclude our analysis by examining the results on technical inefficiency among farms and studying the potential effect (if any) of the 1997 quota regime change on the former. When pooling the technical inefficiency estimates over the entire sample period, we find that, on average, the system FIML approach produces estimates of technical inefficiency of values twice larger than those from a standard single-equation ML approach: 5.2% vs. 2.4% with the corresponding standard deviations of 4.3% and 2.1% (see table 2). The system-based estimates are considerably more dispersed and skewed to the right, as can also be seen in figure 1.

However, when differentiating between the technical inefficiency estimates corresponding to the pre-1997 period (prior to regime change) from those corresponding to the period from 1997 onward (post regime change), we see a substantially different picture. Figure 3 presents box-plots of the distributions of the technical inefficiency estimates for the two time periods across the two models we estimate. Based on our preferred cost system FIML model, we document both an economically

[13]The above specification of $\mu_{u,it}$ essentially models a "structural break".

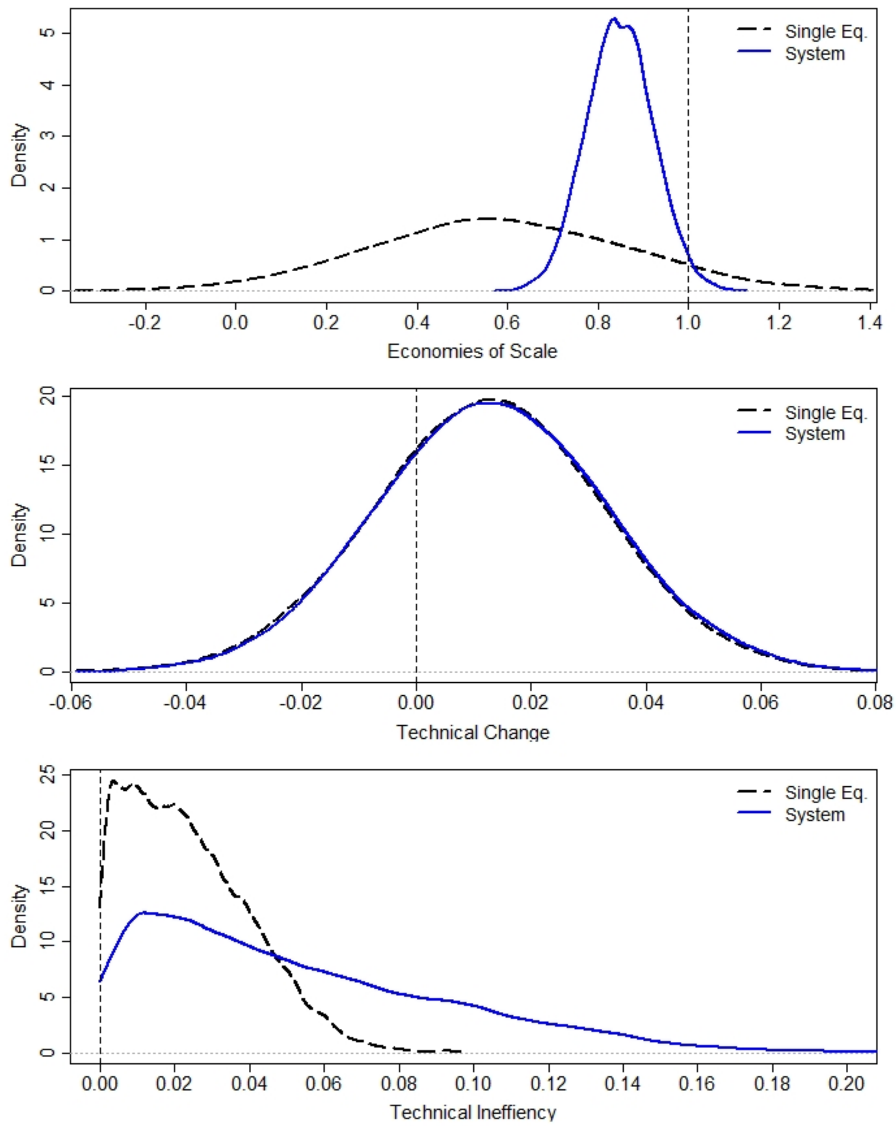

Figure 1: Scale economies, technical change and technical inefficiency

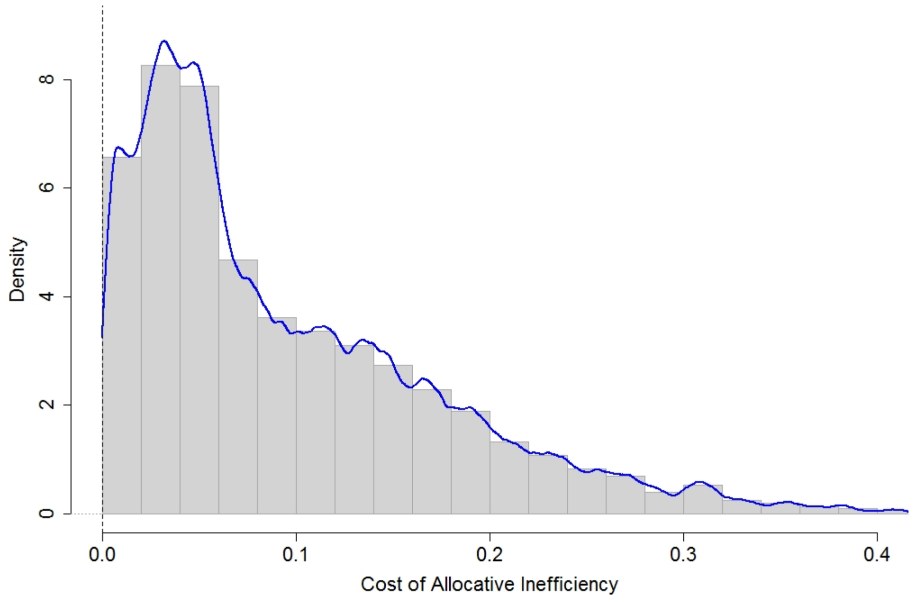

Figure 2: Cost of allocative inefficiency

and statistically significant improvement in the levels of technical efficiency among dairy farms associated with the 1997 quota scheme change. The mean estimate of technical inefficiency has declined from 8.5% in the pre-1997 period to 3.5% in the post-1997 period. This decline is non-negligible as also confirmed by the statistical significance of the coefficient η_1 on H_{it} in the equation for the mean of u_{it}. Table 3 reports the estimates of the mean function parameters with their corresponding standard errors. From table 3 and figure 3, it is evident that the single-equation ML model however fails to detect this structural improvement in the farms' efficiency levels. Admittedly, these differences in the results may not necessarily be due to the failure of a single-equation model to account for the endogeneity of inputs. One may alternatively argue that the post-1997 period is characterized by greater freedom for producers to choose milk outputs, which can potentially lead to the endogeneity in outputs thus compromising the consistency of our system-based estimator. It is possible that a single-equation method, which fails to account for endogeneity of either inputs or outputs and thus is always inconsistent, is more insulated to such changes than our preferred system approach, which becomes inconsistent only during the post-1997 period. The detected improvements in technical efficiency may therefore be argued to be spurious.[14] However, for many years after the regulatory regime change only a small fraction of the milk quota was reallocated due to the reduction in the total quota. Furthermore, since 2002, a farmer interested in selling (a portion of) quota had to sell it to the government to be reallocated. In the light of the above, we believe it is somewhat unlikely that the structural improvement in the farms' efficiency levels that our preferred system-based model estimates is spurious.

Conclusion

In this article, we offer a methodology to address the endogeneity of inputs in the IDF formulation of the multi-output multi-input production process from the perspective of economic theory. We

[14]We would like to thank the editor for bringing this alternative explanation to our attention.

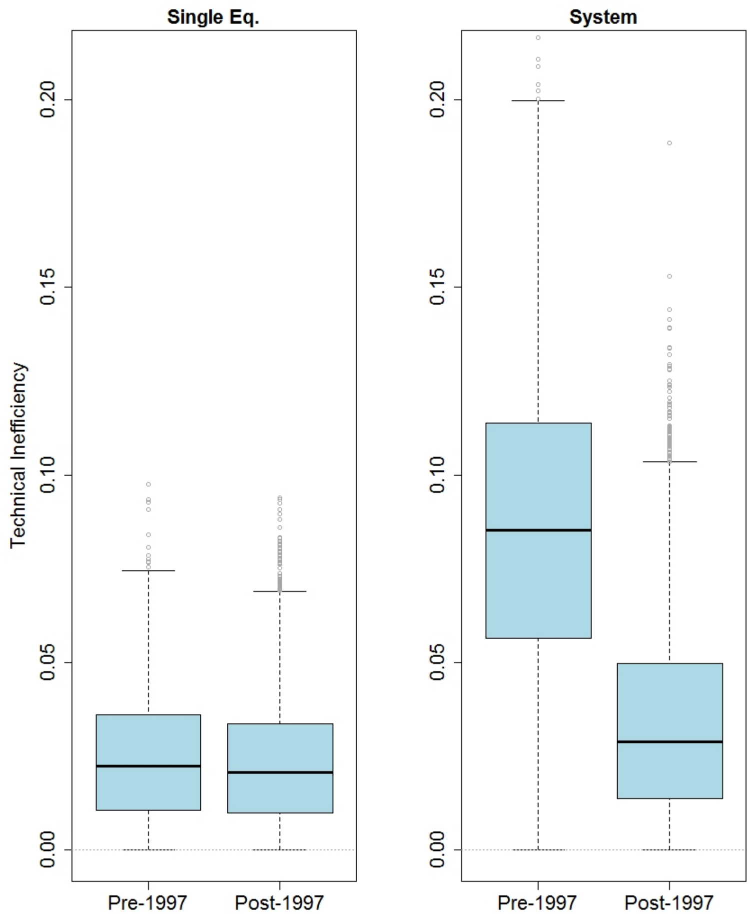

Figure 3: Technical inefficiency pre- and post-regime change

Table 3: Mean of technical inefficiency (μ_u)

Coeff.	Estimate	SE
Single-Equation IDF		
η_0	–0.02103	0.01032
η_1	0.00172	0.02732
Cost System of IDF		
η_0	0.08453	0.00442
η_1	–0.07345	0.01723

suggest invoking the assumption of the firm's cost minimizing behavior not only to justify the treatment of outputs as exogenous but to also tackle endogeneity of the input ratios in the IDF. We do so by considering a flexible (simultaneous) system of the translog IDF along with the FOCs from the firm's cost minimization problem. The model we develop also accommodates both technical and (input) allocative inefficiencies amongst firms. There are advantages to using our system approach over the standard single-equation model even if the input ratios are uncorrelated with the random productivity shocks in the IDF. Since additional equations (the FOCs) do not contain any extra parameters, the system-based estimator is more efficient. Furthermore, technological metrics obtained via the cost system approach are likely to be more meaningful because the economic behavior is embedded into the system through the FOCs.

We showcase our model by applying it to study the production of dairy farms in Norway during the 1991-2008 period. Having failed to reject the presence of endogeneity in the data, we find that the single-equation IDF model produces unreasonable estimates of scale economies as well as fails to detect a change in the level of technical inefficiency associated with the quota scheme change in 1997. In contrast, the technological estimates obtained from the cost system indicate the evidence of increasing returns to scale of reasonable magnitudes (the average of 1.18). Using the system approach, we are also able to document both an economically and statistically significant improvement in the levels of technical efficiency among dairy farms associated with the 1997 quota scheme change. The mean estimate of technical inefficiency has declined from 8.5% in the pre-1997 period to 3.5% in the post-1997 period. The median cost of allocative inefficiency is estimated to be 6.9%.

References

Atkinson, S. E., Cornwell, C., and Honerkamp, O. (2003). Measuring and decomposing productivity change: Stochastic distance function estimation versus data envelopment analysis. *Journal of Business and Economics Statistics*, 21(2):284–294.

Atkinson, S. E. and Primont, D. (2002). Stochastic estimation of firm technology, inefficiency, and productivity growth using shadow cost and distance functions. *Journal of Econometrics*, 108:203–225.

Atsbeha, D. M., Kristofersson, D., and Rickertsen, K. (2012). Animal breeding and productivity growth of dairy farms. *American Journal of Agricultural Economics*, 94:996–1012.

Caves, D. W., Christensen, L. R., and Swanson, J. A. (1981). Productivity growth, scale economies, and capacity utilization in US railroads, 1955-74. *American Economic Review*, 71(5):994–1002.

Chambers, R. G. (1997). *Applied Production Analysis: A Dual Approach*. Cambridge University Press.

Coelli, T. J. (2000). On the econometric estimation of the distance function representation of a production technology. Discussion Paper 2000/42, Center for Operations Research and Econometrics, University Catholique de Louvain.

Coelli, T. J., Hajargasht, G., and Lovell, C. A. K. (2008). Econometric estimation of an input distance function in a system of equations. Working Paper, University of Queensland.

Cornwell, C., Schmidt, P., and Sickles, R. C. (1990). Production frontiers with cross-sectional and time-series variation in efficiency levels. *Journal of Econometrics*, 46(1–2):185–200.

Das, A. and Kumbhakar, S. C. (2012). Productivity and efficiency dynamics in Indian banking: An input distance approach incorporating quality of inputs and outputs. *Journal of Applied Econometrics*, 27:205–234.

Färe, R., Grosskopf, S., and Zaim, O. (2002). Hyperbolic efficiency and returns to the dollar. *European Journal of Operational Research*, 136:671–679.

Färe, R. and Primont, D. (1995). *Multi-Output Production and Duality: Theory and Applications*. Kluwer Academic Publishers, Boston, MA.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica*, 46(6):1251–1271.

Jervell, A. M. and Borgen, S. O. (2000). Distribution of dairy production rights through quotas: The Norwegian case. In Schwarzweller, H. K. and Davidson, A. P., editors, *Dairy Industry Restructuring*. Elsevier Science Inc., New York.

Jondrow, J., Lovell, C. A. K., Materov, I. S., and Schmidt, P. (1982). On the estimation of technical inefficiency in the stochastic frontier production function model. *Journal of Econometrics*, 19:233–238.

Karagiannis, G., Midmore, P., and Tzouvelekas, V. (2004). Parametric decomposition of output growth using a stochastic input distance function. *American Journal of Agricultural Economics*, 86:1044–1057.

Karagiannis, G., Tsionas, E. G., and Kumbhakar, S. C. (2006). Estimation of input distance functions in the presence of technical and allocative inefficiency: A system approach. Working Paper, Binghamton University.

Kumbhakar, S. and Lovell, C. A. K. (2000). *Stochastic Frontier Analysis*. Cambridge University Press, New York, NY.

Kumbhakar, S. C. (2013). Specification and estimation of multiple output technologies: A primal approach. *European Journal of Operational Research*, 231:465–473.

Kumbhakar, S. C., Lien, G., Flaten, O., and Tveteräs, R. (2008). Impacts of Norwegian milk quotas on output growth: A modified distance function approach. *Journal of Agricultural Economics*, 59:350–369.

Lambert, D. K. and Wilson, W. W. (2003). Valuing varieties with imperfect output quality measurement. *American Journal of Agricultural Economics*, 85:95–107.

Li, Q. (1996). Nonparametric testing of closeness between two unknown distribution functions. *Econometric Reviews*, 15:261–274.

Marschak, J. and Andrews, W. H. (1944). Random simultaneous equations and the theory of production. *Econometrica*, 12(3–4):1263–1298.

Petrin, A. and Train, K. (2010). A control function approach to endogeneity in consumer choice models. *Journal of Marketing Research*, 47(1):3–13.

Schmidt, P. and Lovell, C. A. K. (1978). Estimating technical and allocative inefficiency relative to stochastic production and cost frontiers. *Journal of Econometrics*, 9:343–366.

Shephard, R. W. (1953). *Cost and Production Functions*. Princeton University Press, Princeton, NJ.

Shephard, R. W. (1970). *Theory of Cost and Production Functions*. Princeton University Press, Princeton, NJ.

Silverman, B. W. (1986). *Density Estimation*. Chapman and Hall, London.

Sipiläinen, T., Kumbhakar, S. C., and Lien, G. (2014). Performance of dairy farms in Findland and Norway from 1991 to 2008. *European Review of Agricultural Economics*, 41:63–86.

Terza, J. V., Basu, A., and Rathouz, P. J. (2008). Two-stage residual inclusion: Addressing endogeneity in health econometric modeling. *Journal of Health Economics*, 27:531–543.