ECONOMIC DISCUSSION PAPERS

Efficiency Series Paper 06/2004

Endogeneity Problems in the Estimation of Multi-Output Technologies

David Roibás y Carlos Arias

Departamento de Economía

Universidad de Oviedo

Available online at: www.uniovi.es/economia/edp.htm

UNIVERSIDAD DE OVIEDO

DEPARTAMENTO DE ECONOMÍA

PERMANENT SEMINAR ON EFFICIENCY AND PRODUCTIVITY

ENDOGENEITY PROBLEMS IN THE ESTIMATION OF MULTI-OUTPUT TECHNOLOGIES

David Roibas♣ and Carlos Arias♠

Efficiency Series Paper 06/2004

Abstract: Empirical representations of multi-output technology use either quantities or ratios of outputs as explanatory variables. These specifications raise questions about the existence of endogeneity problems in the estimation of multi-output technology. We contribute to the analysis of this issue by distinguishing between two sources of endogeneity: endogeneity of the production plan and unobservability of the production plan (error in variables). Using Monte Carlo simulation, we explore the consequences of endogeneity for two empirical representations of multi-output technologies: the distance function and the transformation function. Particularly, we analyze the role played by the degree of correlation among output disturbances.

Key words: multi-output production process, output oriented distance functions, transformation function, output endogeneity, instrumental variables, Monte Carlo.

♣ Departamento de Economía, Universidad de Oviedo, Spain.

♠ Departamento de Economía, Universidad de León, Spain.

Corresponding author: David Roibás. E-mail: droibas@uniovi.es.

1. Introduction

This paper explores the role of endogeneity in the estimation of multi-output technologies. Some examples of multi-output technologies are agricultural and livestock products in farming or passengers and freight in transportation. Additionally, sometimes environmental damage is modeled as the result of a multi-output production process where pollutants are considered one output of a multi-output technology (Färe *et al*. 1996).

Empirical analysts have dealt frequently with multi-output technologies by aggregating multiple outputs into a single output. An alternative is the use of a multi-output cost function. Coelli and Perelman (2001) discuss the problems of both approaches. In the former case, the aggregation of multiple outputs into a single output can cause well known problems. In the latter case, the dual approach requires the availability of input prices plus the assumption of cost minimization.

The distance function is a primal representation of a multi-ouput technology that avoids both problems. Additionally, it naturally integrates a radial measure of technical efficiency involving all outputs in the production process. In fact, the distance function provides a convenient representation of radial efficiency measured as the inverse of distance between the observations and the frontier of the technology (Kumbhakar and Lovell, 2000). The homogeneity of degree one in output of the output oriented distance function[1] allows the factoring out of an output for **any** functional form. This property seems to be highly valued in empirical work because a dependent variable (the output which has been factored out) can be written as a function of ratios of outputs and inputs, which allows to estimate distance functions using standard econometric methods (see, for example: Lovell *et al.*, 1994; Coelli and Perelman, 2000).

Two paths have been followed in previous research to deal with the potential endogeneity of output ratios. For example, Sickles *et al.* (1996), Cuesta and Orea (1998), Rodriguez Alvarez (2000) and Atkinson *et al.* (2003) use instrumental variables to correct the endogeneity problem mentioned above. A different approach is followed by Coelli and Perelman (2001) which argue that the endogeneity problem does not

[1] We will restrict our analysis to the output oriented distance function.

appear under reasonable behavioral assumptions. Particularly, the ratios of outputs are exogenous variables if the random disturbances affecting production processes change all outputs in the same proportion. However, this seems an odd requirement in many production processes. For example, in marine fisheries, for given inputs, weather, tides and streams affect not only in different proportions but in different directions catches of different species. Therefore, if random disturbances make outputs change in different proportions, it is necessary to deal with a ratio of endogenous variables. Consistent estimation of the parameters of the technology requires finding instruments for ratios of endogenous variables which can be a difficult task.

The transformation function is a feasible alternative to estimate multi-output technologies (Mundlak, 1963; Diewert, 1973: Hall, 1973). Under some regularity conditions, the transformation function might be specified with one output on the left hand side of the equation and other outputs and inputs on the right hand side. This specification can be estimates by least squares or maximum likelihood (Weningen and Strand, 2003). In this case, only instruments for output quantities are required. The transformation function, however, is not used in empirical work measuring efficiency because the error term in the equation is only related to the output selected as a dependent variable, which in turn, implies that efficiency measures are only related to this output (Orea *et al.* 2003). To solve this limitation we propose a specification of a transformation function that includes an efficiency index related to all outputs. This specification of the model can be estimated with panel data using a fixed effect estimator in a way like the one used in the works of Atkinson and Cornwell (1994) or Orea *et al*. (2004).

There are two contributions in the present paper. First, by carefully distinguishing between planned and observed output we clarify the problems of endogeneity in the estimation of multioutput technologies. Precisely, we are able to characterize the unlikely conditions under which distance functions can be estimated by plain least squares. In any other circumstances, instrumental variables are required for consistent estimation. Second, we develop a transformation function which incorporates a global efficiency term and we use Monte Carlo simulation in order to compare the relative performance of both primal approaches to the technology: transformation and distance functions. Particularly, we analyze the role played by different degrees of correlation among output disturbances. We find that under different degrees of correlation between outputs the techniques used to correct endogeneity problems are more effective in the transformation function than in the distance function, so the transformation function dominates the distance function on empirical performance.

The structure of the paper is the following. In Section 2, we discuss the estimation of multi-output technologies using, alternatively, the distance and transformation function. In Section 3, we describe a Monte-Carlo experiment that shows the empirical performance of the two alternative representations of technology. In Section 4, we present the results of the Monte-Carlo experiment. Finally, some concluding comments are included in Section 5.

2. Estimation of multi-output technologies

In this section, we analyze the estimation of multi-output technologies using, alternatively, the distance and the transformation function. We limit our analysis to the one-input two-output case since it is the simplest one that contains the empirical difficulties mentioned in the previous section.

In the empirical estimation of single output production processes inputs are the basic explanatory variables. Very often, endogeneity is written-off by arguing that observed inputs were chosen ex–ante and are not affected by random noise (Zellner *et al.*, 1966). In multi-output production processes things are a bit more complicated. Now, some of the outputs (or ratios of outputs) are explanatory variables and it is harder to argue that observed outputs are chosen ex–ante.

In this section, we explicitly distinguish between planned production (unobserved) and actual production (observed). This distinction proves to be essential to analyze the problems that arise in the estimation of multi-output technologies. It is worth noting that planned production, and input endowments, could be chosen ex-ante. Therefore, it is reasonable to assume that they are independent of ex-post random shocks. On the other hand, observed outputs are possibly affected by contemporaneous random shocks. We claim that inclusion of the planned (unobserved) output as explanatory variable in the empirical representation of multi-output technology can avoid biases due to endogeneity. However, since planned output has to be approximated by observed output we will likely be facing a problem of error in variables akin to endogeneity. We will develop these ideas carefully in the next two subsections.

2.1. The distance function

Shephard (1953) proposes the distance function as a primal representation of multi-output technologies. The output oriented distance function can be written as[2]:

$$D\left(x, y_1^*, y_2^*\right) = \min\left\{\theta \;\middle|\; \left(\frac{y_1^*}{\theta}, \frac{y_2^*}{\theta}\right) \text{can be produced with } x\right\} \tag{1}$$

where *x* denotes quantity of input, y_1^* and y_2^* are the planned quantities of two outputs that can be produced with input *x* and θ ($0 \le \theta \le 1$) is an output oriented index of technical efficiency.

The distance function is homogeneous of degree one in outputs (Coelli *et al.*,1998). Therefore, dividing both outputs by y_1^* we have that:

$$\frac{\theta}{y_1^*} = D\left(x, 1, \frac{y_2^*}{y_1^*}\right) \Rightarrow \ln y_1^* = -\ln D\left(x, 1, \frac{y_2^*}{y_1^*}\right) + \ln\theta \tag{2}$$

The primal representation of a multi-output technology in (2) has two convenient properties. First, the linearity on the logarithm of technical efficiency greatly simplifies the estimation. In fact, this is the standard specification in the literature on output-oriented technical efficiency. Second, one of the outputs can be factored out easily which facilitates the use of standard econometric techniques (e.g. least squares or maximum likelihood).

In the estimation of distance functions planned output is usually substituted by observed output in equation (2), and a random term v_1 is added.

$$\ln y_1 = -\ln D\left(x, 1, \frac{y_2}{y_1}\right) + \ln\theta + v_1 \tag{3}$$

---

[2] Alternatively, the input oriented distance function can be written as: $D\left(x, y_1, y_2\right) = \min\left\{\theta \mid \left(y_1, y_2\right) \text{ can be produced with } \theta x\right\}$.

This approach raises difficult issues about the correlation between the ratio of observed outputs and the random disturbance v_1. Some authors have pointed out to the endogeneity of the output ratio and propose the use of instrumental variables (Sickles *et al.*, 1996; Cuesta and Orea, 1998; Rodriguez, 2000; and Atkinson *et al.*, 2003). Other authors claim that under reasonable behavioral assumptions the ratio of outputs is an exogenous variable (Coelli and Perelman, 2001).

We claim that a careful analysis of the substitution of observed by planned output in equation (2) can shed some light into the nature of the problem and the solution. In an empirical setting, it is reasonable to assume that outputs are affected by variables not under the control of the producer. We deal with this issue by distinguishing between observed production (y_1, y_2) and planned production (y_1^*, y_2^*) and assuming the differences between observed and planned production can be modeled as a multiplicative random disturbance. In this case, the relation between the planned and observed output can be written as:

$$\begin{aligned} y_1 &= y_1^* e^{v_1} \\ y_2 &= y_2^* e^{v_2} \end{aligned} \tag{4}$$

where the random disturbances v_1 and v_2 are assumed to be distributed as a multivariate random variable with the following characteristics:

$$\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \sim N(0, \Omega) \quad \Omega = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix} \tag{5}$$

Substituting (4) in (2) we have that:

$$\ln y_1 = -\ln D\left(x, 1, \frac{y_2 e^{-v_2}}{y_1 e^{-v_1}} \right) + \ln\theta + v_1 \tag{6}$$

In general, the random disturbances inside the distance function can not be factored-out. Therefore, a careful substitution of planned by observed output does not leads to the empirical specification in expression (3).

As an alternative, we analyze the case in which the substitution of planned by observed output is done only for the factored-out output:

$$\ln y_1 = -\ln D\left(x, 1, \frac{y_2^*}{y_1^*}\right) + \ln\theta + v_1 \tag{7}$$

The main feature of expression (7) is the use of the ratio of planned outputs as explanatory variable versus the ratio of observed outputs in expression (3). This difference has important empirical consequences. For once, it is not difficult to argue that planned output is independent of the random disturbances. In fact, that is the case in models that assume that firms maximize an expected objective function (e.g. expected profit) instead of the actual objective function. In this case, planned output is not stochastic and there is not an endogeneity problem in expression (7) (Marshak and Andrews, 1944; Hoch, 1958; Zellner *et al.*, 1966; Coelli, 2000). However, since the ratio of planned outputs is unobserved it is necessary to face the difficult task of getting a good proxy.

A natural candidate is the ratio of observed outputs. From expression (4), we have that:

$$\frac{y_2}{y_1} = \frac{y_2^* e^{v_2}}{y_1^* e^{v_1}} = \frac{y_2^*}{y_1^*} e^{v_2 - v_1} \tag{8}$$

Therefore, the ratio of observed outputs is not correlated with the error term in equation (7) when the covariance between $(v_2 - v_1)$ and (v_1) is null. This condition implies:

$$C(v_2 - v_1, v_1) = C(v_2, v_1) - V(v_1) = 0 \Rightarrow \frac{C(v_2, v_1)}{V(v_1)} = 1 \tag{9}$$

In turn, expression (9) implies that:

$$v_2 = \alpha + v_1 \text{ where } \alpha \in \Re \tag{10}$$

Finally, the requirement of zero expectation for the random disturbances of observed output (v_1 and v_2) implies that $\alpha = 0$.

Therefore, the ratio of observed outputs is a good measure of the ratio of planned outputs only if the random disturbances that affect each output are equal (alternatively, if random shocks affect all outputs in the same proportion). Otherwise, we need to rely on a procedure akin to instrumental variables. We need to find a set of random variables (z) uncorrelated with the random disturbances v_1 and v_2. Taking natural logarithms and expectations with respect to z in expression (8) we have that:

$$\mathrm{E}\left[\ln\left(\frac{y_1}{y_2}\right) \mid z\right] = \ln\left(\frac{y_1^*}{y_2^*}\right) + \mathrm{E}\left[v_1 - v_2 \mid z\right] = \ln\left(\frac{y_1^*}{y_2^*}\right) \tag{11}$$

The log of the ratio of planned output equals the expected value of observed output with respect to the variables z. The best predictor of this expectation is the fitted value of least squares of $\ln\left(\frac{y_1}{y_2}\right)$ on z. Therefore, the fitted value of the ratio of actual outputs is used instead of the ratio of planned outputs in the estimation of equation (7) (Greene, 1993; chapters 9 and 20).

In summary, our setting for the estimation of a distance function splits the endogeneity problem in two parts. First, we consider the endogeneity of the planned output. Secondly, we recognize that planned output is an unobserved variable that leads, in many instances, to a problem of error in variables that requires an econometric approach similar to instrumental variables.

2.2. The transformation function

The multi-output technology can be represented by the following transformation function:

$$F\left(x, y_1^*, y_2^*\right) = 0 \tag{12}$$

where x denotes quantity of input, y_1^* and y_2^* are the maximum quantities of two outputs that can be produce with input x and F is a transformation function with the usual properties (Chambers, 1988). If the function in (12) is continuously differentiable and has nonzero first derivatives with respect to y_1 the transformation function may be specified as follows using the implicit function theorem:

$$y_1^* = f\left(x, y_2^*\right) \tag{13}$$

Felthoven and Paul (2004) argue that, in empirical applications, this specification avoids endogeneity problems: “The outputs in this function are specified in levels, whereas distance functions estimated by stochastic production frontier methods are typically specified in terms of output ratios (in order to impose homogeneity). Estimating the transformation function thus avoids possible econometric endogeneity problems associated with having the dependent variable (the numeraire output) also appear in arguments of the function (the denominators of the output ratios), which may be correlated with the error term and violate standard independence assumptions."

Using expression (4) and taking natural logs in (13), the transformation function to be estimated can be written as:

$$\ln y_1 = \ln f\left(x,\ y_2^*\right) + v_1 \tag{14}$$

The firm production plan $\left(x, y_1^*, y_2^*\right)$ is uncorrelated with the disturbance v_1 if the firm maximizes expected profit. Therefore, strictly speaking there is no problem of endogeneity in (14). However, the planned output y_2^* is unobservable and using observed output $\left(y_2 = y_2^* e^{v_2}\right)$ as a proxy to estimate (14) involves a problem of error in variables[3]. Again, it is necessary to use instrumental variables for consistent estimation.

The transformation function is not used frequently in empirical analysis when technical efficiency is involved. The reason being that including a simple additive efficiency term in equation (14) implies that the efficiency measure refers only to the output set as dependent variable. However, a multi-output output oriented index of technical efficiency can be included in the specification of a transformation function treating efficiency as a fixed effect (Atkinson and Cornwell, 1994; Orea *et al.*, 2004):

$$\ln\left(\frac{y_1}{\theta}\right) = \ln f\left(x,\ \frac{y_2^*}{\theta}\right) + v_1 \tag{15}$$

where θ is an output oriented index of technical efficiency. Moving the logarithm of *TE* to the right hand side yields:

[3] See Greene (1993), chapter 9.

$$\ln y_1 = \ln f\left(x,\ \frac{y_2^*}{\theta}\right) + \ln\theta + v_1 \tag{16}$$

The model in (16) can be estimated with panel data assuming that efficiency is a time invariant individual effect.

Our rationale to propose the estimation of the transformation function in (16) is that in many empirical applications output prices, which are correlated with output levels, are frequently available and they are natural instrumental variables. However, finding a set of variables correlated to output ratios is more difficult. At best, it can be done through output price ratios, but correlation between output ratios and output prices can be weaker than correlation between output levels and prices so output prices seem to be a better instrumental variable for output levels than price ratios for output ratios. Therefore, it is natural to expect that an empirical representation of the technology where the variables are the output levels (transformation function) produces estimates closer to the true technology and efficiency levels than those representations of the technology where the variables are the output ratios.

Additionally, in applications with at least one undesirable output (Färe et al., 1996), it can be difficult to find a instrumental variables correlated with the undesirable output. However, using a transformation function allows to use the undesirable output as dependent variable, eliminating the need of an instrumental variable for it.

In summary, in this section we have discussed the properties of two alternative feasible empirical representations of multi-output technologies. We have found that the estimation of the transformation function requires of instrumental variables for outputs. On the other hand, unless the random disturbances are equal for all outputs, the estimation of the distance function requires instrumental variables for ratios of outputs. In the following two sections, we explore the estimation of both the distance and the transformation function.

3. Monte Carlo experiment

The starting point of data generation is the following distance function for planned output $\left(y_1^*, y_2^*\right)$:

$$
\begin{aligned}
-\ln y_{1it}^* &= c_1 + c_2 \ln x + c_3 \ln\left(\frac{y_{2it}^*}{y_{1it}^*}\right) + c_4 \ln x \ln\left(\frac{y_{2it}^*}{y_{1it}^*}\right) \\
&+ c_5 \ln\left(\frac{y_{3it}^*}{y_{1it}^*}\right) + c_6 \ln x \ln\left(\frac{y_{3it}^*}{y_{1it}^*}\right) + c_7 \left(\ln x\right)^2 - \ln D_i
\end{aligned}
\tag{17}
$$

where *i* denotes firms (*i = 1, ... ,5*), *t* denotes time period (*t = 1, ... ,100*). The total number of observations is 500. The alternative representation for the distance function in (17) is the following transformation function:

$$
\begin{aligned}
\ln y_{1it}^* &= -\frac{c_1 + c_2 \ln x_{it} + c_3 \ln y_{2it}^* + c_4 \ln x_{it} \ln y_{2it}^*}{1 - c_3 - c_5 - \left(c_4 + c_6\right) \ln x} \\
&- \frac{c_5 \ln y_{3it}^* + c_6 \ln x_{it} \ln y_{3it}^* + c_7 \left(\ln x\right)^2 + \ln \theta}{1 - c_3 - c_5 - \left(c_4 + c_6\right) \ln x}
\end{aligned}
\tag{18}
$$

This function is obtained by factoring out $\ln y_{1it}^*$ in expression (17).

The distributions of the independent variables are:

$$
\ln x \sim U\left[0,2\right] \;\; \ln y_{2it}^* \sim N\left(0,1\right) \;\; \ln y_{3it}^* \sim N\left(0,1\right) \tag{19}
$$

The value of the parameters are shown in Table 1.

Table 1. Parameter values for the simulation

Parameter Name	Parameter Value	Parameter Name	Parameter value
$\ln D_1$	-0.4	c_1	-2
$\ln D_2$	-0.3	c_2	-1
$\ln D_3$	-0.2	c_3	0.35
$\ln D_4$	-0.1	c_4	-0.1
$\ln D_5$	0	c_5	0.15
		c_6	-0.15
		c_7	-0.5

Finally, values of $\ln y_{1it}^{*}$ are generated using expression (18).

The values of observed output (y_1, y_2) (2500 repetitions) are generated using expression (4) and the following random disturbances for observed output:

1. Observed output disturbances are perfectly correlated standard normal random variables

$$\begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \sim N(0,\Omega) \quad \Omega = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \tag{20}$$

2. Observed output disturbances are independently distributed standard normal random variables

$$\begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \sim N(0,\Omega) \quad \Omega = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{21}$$

3. Observed output disturbances are correlated standard normal variables

$$\begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \sim N(0,\Omega) \quad \Omega = \begin{bmatrix} 1 & -1 & -1 \\ -1 & 1 & 1 \\ -1 & 1 & 1 \end{bmatrix} \tag{22}$$

Finally, we generate a set of instrumental variables for observed output defined as:

$$\begin{aligned} \ln Z_1 &= \ln y_{1it}^{*} + u_1 \quad u_1 \sim N(0,1) \\ \ln Z_2 &= \ln y_{2it}^{*} + u_2 \quad u_2 \sim N(0,1) \\ \ln Z_3 &= \ln y_{3it}^{*} + u_3 \quad u_3 \sim N(0,1) \end{aligned} \tag{23}$$

Since the random variables (u_1, u_2, u_3) in (23) are generated independently from the random disturbances of the distance function (v_1, v_2, v_3), then each instrumental variable is correlated with the corresponding planned output but not with the random disturbances of the transformation function in (18).

4. Results

We estimate the distance function in (17) and the transformation function in (18) alternatively by:

a. OLS
b. Using the set of Instrumental Variables *1*(IV_1): lnZ_2 and lnZ_3.
c. Using the set of Instrumental Variables *2* (IV_2): the logarithms of the ratio of outputs (lnZ_2- lnZ_1) and (lnZ_3- lnZ_1).

The parameters of the distance and transformation functions are estimated for the 2500 samples of observed output. The means and standard deviations (with respect to the true value) of the estimated parameters are shown in Tables 2, 3 and 4. Table 2 shows the results of the three different estimation approaches when random output shocks are perfectly correlated. The estimation of the transformation function by instrumental variables gives good results. The case of identical output random shocks is well suited for the distance function because, in this case, the ratio of outputs are exogenous variables. As a result, the estimation of the distance function by ordinary least squares provides consistent estimates of the parameters. Finally, the estimation of the distance function with instrumental variables for the output and for the ratio of outputs gives poorer results than plain least squares.

The transformation function produces nine out of eleven parameters estimates closer to the true parameter value than the distance function with the set of instrumental variables IV_1. Additionally the transformation function produces ten out of eleven parameters estimates closer to the true parameter value than the distance function with the set of instrumental variables IV_2. Comparing the estimates of the distance function with both sets of instrumental variables we find that the set of instrumental variables IV_1 produces six out of eleven parameters estimates closer to the true value than the instrumental variable set IV_2.

An important result is that the standard deviations of estimated parameters are always smaller in the estimation of the transformation function than in the estimation of distance function using both sets of instrumental variables.

Table 2. Random shocks are perfectly correlated

Parameter	True	Transformation Function (IV)	Distance Function (OLS)	Distance Function (IV_1)	Distance Function (IV_2)
lnD_1	-0.4	-0,3883	-0,4012	-0,4047	-0,4222
		(0,108)	(0,144)	(0,148)	(0,147)
lnD_2	-0.3	-0,3182	-0,3029	-0,3212	-0,3422
		(0,109)	(0,146)	(0,149)	(0,148)
lnD_3	-0.2	-0,2277	-0,2009	-0,2059	-0,2161
		(0,108)	(0,141)	(0,144)	(0,143)
lnD_4	-0.1	-0,0993	-0,1009	-0,1044	-0,1074
		(0,107)	(0,145)	(0,144)	(0,144)
c_1	-2	-1,9624	-2,0046	-1,4360	-1,8016
		(0,281)	(0,314)	(0,575)	(0,485)
c_2	-1	-1,0478	-0,9907	-1,1022	-1,1102
		(0,321)	(0,371)	(0,441)	(0,428)
c_3	0.35	0,3559	0,3498	0,4623	0,4204
		(0,071)	(0,070)	(0,101)	(0,111)
c_4	-0.1	-0,1026	-0,0989	-0,1936	-0,1384
		(0,070)	(0,062)	(0,087)	(0,096)
c_5	0.15	0,1607	0,1505	0,2450	0,1665
		(0,080)	(0,082)	(0,134)	(0,154)
c_6	-0.15	-0,1809	-0,1510	-0,2371	-0,2227
		(0,085)	(0,074)	(0,112)	(0,131)
c_7	-0.5	-0,5652	-0,5038	-0,8176	-0,6950
		(0,166)	(0,175)	(0,222)	(0,202)

Table 3 shows the results of the three different estimation techniques when random output shocks are independent. The estimation of the transformation function by instrumental variables gives good results. However, in this case, ordinary least squares is not a reasonable procedure to estimate the distance function since the ratios of outputs used as explanatory variables are not exogenous. Finally, instrumental variables for output applied to the distance function provide reasonable results in this case.

The transformation function produces six out of the eleven parameters closer to the true value than in the distance function with IV_1. The transformation function produces eight out of the eleven parameters closer to the true value than in the distance function with IV_2. The distance function with IV_2 gives seven parameters closer to the true value than distance function with IV_1. Again, the transformation function with instrumental variables produces smaller standard errors of the parameter estimates than the distance function with instruments.

Table 3: Random shocks are independent

Parameter	**True**	**Transformation Function (IV)**	**Distance Function (OLS)**	**Distance Function (IV_1)**	**Distance Function (IV_2)**
lnD_1	-0.4	-0,4321	-0,2465	-0,4141	-0,4347
		(0,109)	(0,088)	(0,123)	(0,129)
lnD_2	-0.3	-0,3238	-0,1403	-0,2894	-0,3102
		(0,107)	(0,087)	(0,119)	(0,125)
lnD_3	-0.2	-0,2681	-0,0677	-0,2207	-0,2492
		(0,108)	(0,086)	(0,124)	(0,130)
lnD_4	-0.1	-0,1356	0,0041	-0,1070	-0,1269
		(0,104)	(0,086)	(0,114)	(0,122)
c_1	-2	-2,0184	-1,6597	-1,5296	-1,9025
		(0,264)	(0,151)	(0,524)	(0,442)
c_2	-1	-1,0057	-0,2819	-1,0014	-1,0429
		(0,313)	(0,206)	(0,431)	(0,416)
c_3	0.35	0,3493	0,3197	0,4482	0,4121
		(0,061)	(0,030)	(0,084)	(0,096)
c_4	-0.1	-0,1004	-0,0224	-0,1861	-0,1401
		(0,067)	(0,029)	(0,080)	(0,092)
c_5	0.15	0,1516	0,2487	0,2243	0,1399
		(0,078)	(0,033)	(0,127)	(0,150)
c_6	-0.15	-0,1586	-0,0223	-0,2053	-0,1873
		(0,089)	(0,031)	(0,114)	(0,134)
c_7	-0.5	-0,5329	-0,2167	-0,7935	-0,6676
		(0,164)	(0,101)	(0,231)	(0,219)

Table 4 shows the result of the different estimation approaches when there is a substantial correlation between the random output shocks. In this case, both the output used as explanatory variable in the transformation function and the ratio of outputs used in the distance function are endogenous variables. As expected, ordinary least squares does not provide reasonable estimates in the transformation function or the distance function. In turn, the transformation function with instrumental variables provides consistent estimates of the parameters. In the case of the distance function the use of instrumental variables allows consistency in the estimation.

Table 4: Correlated random shocks

Parameter	**True**	**Transformation Function (IV)**	**Distance Function (OLS)**	**Distance Function (IV_1)**	**Distance Function (IV_2)**
lnD_1	-0.4	-0,4026	-0,3244	-0,4129	-0,4375
		(0,105)	(0,012)	(0,075)	(0,089)
lnD_2	-0.3	-0,2931	-0,2354	-0,3103	-0,3242
		(0,105)	(0,012)	(0,071)	(0,086)
lnD_3	-0.2	-0,2094	-0,1151	-0,2012	-0,2171
		(0,106)	(0,012)	(0,072)	(0,086)
lnD_4	-0.1	-0,0821	-0,0954	-0,1063	-0,1073
		(0,104)	(0,012)	(0,064)	(0,080)
c_1	-2	-1,9902	-1,9866	-1,6175	-1,9737
		(0,250)	(0,019)	(0,499)	(0,452)
c_2	-1	-1,0254	-0,3512	-0,9752	-1,0049
		(0,331)	(0,030)	(0,459)	(0,448)
c_3	0.35	0,3559	0,3439	0,4496	0,4173
		(0,053)	(0,004)	(0,074)	(0,085)
c_4	-0.1	-0,1090	-0,0278	-0,1866	-0,1460
		(0,067)	(0,005)	(0,080)	(0,089)
c_5	0.15	0,1415	0,1422	0,1950	0,1143
		(0,074)	(0,005)	(0,121)	(0,147)
c_6	-0.15	-0,1506	0,0013	-0,1801	-0,1625
		(0,091)	(0,007)	(0,114)	(0,133)
c_7	-0.5	-0,5308	-0,1819	-0,7524	-0,6380
		(0,167)	(0,016)	(0,229)	(0,213)

Again, we can observe that the transformation function produces parameter estimates closer to the true value than distance function with IV_1. Eight out of the eleven parameters are closer to the true value in the former case. We can also observe that nine of the parameters are closer to the true value in the transformation function estimation than in the distance function estimation with IV_2. As in the previous case, we observe that seven parameters are closer to the true value in the estimation of distance function with IV_2 than in the estimation with IV_1.

The results in Table 4 show that the standard errors of the parameters related to efficiency are smaller in the distance functions with both sets of instrumental variables than in the transformation function. However, the set of parameters common to all firms have a smaller standard error in the estimation of the transformation function.

5. Conclusions

In this paper, we have analyzed the role of endogeneity in the estimation of multi-output technologies. Our analysis starts by considering two sources of endogeneity. One of them is given by possible correlations between random shocks and the production plan of the firm. In this sense, classical results in the literature establish that if the firm maximizes expected profits the production plan is not correlated with random shocks and no endogeneity problem arises. The second source of endogeneity is related to the fact that the production plan is unobservable and, therefore, actual production, which incorporates random shocks, is used for the empirical analysis.

Our analysis shows that this second source of endogeneity affects differently the two empirical representations of a multi-output techonology considered: the distance function and the transformation function. Basically, endogeneity is not an issue in the estimation of distance functions if random disturbances affect all outputs in the same proportion. On the other hand, endogeneity is not a problem in the estimation of the transformation function if random disturbances of outputs are not correlated. However, if random disturbances of outputs have an arbitrary degree of correlation, endogeneity problems arises in both empirical models: the distance function and the transformation function.

Finally, we explore by Monte-Carlo simulation the role of correlation among output disturbances in the relative performance of the proposed empirical models. In order to do this, we develop a specification of the transformation function in which technical efficiency is related to all outputs in the production set. Instrumental variables used in simulation to correct endogeneity problems are designed to be correlated with output levels, assuming that the natural instrumental variables in empirical work, output prices, are correlated with output levels rather than with output ratios.

The main results of the Monte-Carlo simulation are the following:

a) The best performance of the distance function corresponds to the estimation that uses instrumental variables for all outputs (including the output factored out).
b) The transformation function performs better than the distance function in estimating technology and efficiency when output random disturbances are arbitrarily correlated.

These results suggest that, independently of the empirical tool used, instrumental variables are necessary when the random disturbances of outputs are arbitrarily correlated. On the other hand, the estimations obtained with the proposed transformation function are more precise than those obtained using a distance function. At least, if instrumental variables are expected to be correlated with output levels rather than with output ratios.

References

Atkinson, S. E. and Cornwell, C. (1994), "Estimation of Output and Input Technical Efficiency Using a Flexible Functional Form and Panel Data", *International Economic Review* 35: 245-256.

Atkinson, S. E., Fare, R. and Primont, D. (2003) "Stochastic Measurement of Firm Inefficiency Using Distance Functions" *Southern Economic Journal* 69: 596-611.

Chambers, R.G. (1988), *Applied Production Analysis: A Dual Approach*, Cambridge University Press.

Coelli, T., Pradada-Rao, D.S. and Battese, G.E. (1988), *An Introduction to Efficiency and Productivity Analysis*, Kluwer Academic Publishers.

Coelli, T. and Perelman, S. (2000), "Technical Efficiency of European Railways: A Distance Funcion Approach", *Applied Economics*, 32, 1967-1976

Coelli, T. and Perelman, S. (2001), "Medición de la eficiencia técnica en contextos multiproducto", in *La medición de la eficiencia y la productividad*. Antonio Alvarez, editor.

Cuesta, R.A. and Orea, L. (1998), "Technical Efficiency, Distance Functions And Endogeneity: UIT Application to Spanish Savings Banks", paper presented at the 3rd Biennal Georgia Productivity Workshop, University of Georgia, October, 1998.

Diewert, W. E. (1973), "Functional Forms for Profit and Transformation Functions", *Journal of Economic Theory* 6: 285-304.

Färe, R.S., Grosskopf, S. and Tyteca, D. (1996), "An Activity Analysis Model of the Environment Performance of Firms: Application to Fossil-Fuel-Fired-Utilities", *Ecological Economics*, 18, 161-175.

Felthoven, R. G. and Paul, C. J. (2004), "Multi-Output, Nonfrontier Primal Measures of Capacity and Capacity Utilization", *American Journal of Agricultural Economics*, 86(3): 619-633.

Greene, W.H. (1993), *Econometric Analysis*. Second edition. Macmillan. New York.

Hall, R. (1973), "The specificaton of Technology with Several Kinds of Output", *Journal of Political Economy* 81: 878:92.

Hoch, I. (1958) "Simultaneous Equation Bias in the Context of the Cobb-Douglas Production Function", *Econometrica*, 26: 566-578.

Kumbharkar, S. and Lovell, C.A.K. (2000), *Stochastic Frontiers Analysis*, Cambridge University Press

Marshak, J. and Andrews, W. (1944) "Random Simultaneous Equations and the Theory of Production", *Econometrica*, 12: 143-205.

Mundlak, Y. (1963), "Specification and Estimation of Multiproduct Production Functions", *Journal of Farm Economics* 45: 433-443.

Lovell, C. A. K., Richardson, P., Travers, P. and Wood, L. L. (1994), "Resources and Funcionings: A New View of Inequality in Australia", in W. Eichhorn (ed), *Models and Measurement of Welfare and Inequality*, Berlin, Springer-Verlag.

Orea, L., Álvarez, A. and Paul, C. J. (2003), "Modelling and Measuring Production Processes for a Multi-Species Fishery: Alternative Technical Efficiency Estimates for the Northerrn Spain Hake Fishery", Manuscript.

Orea, L., Roibás, D. and Wall, A. (2004), "Choosing the Technical Efficiency Orientation to Analyze Firms' Techhnology: A Model Selection Test Approach", *Journal of Productivity Analysis* 22: 51-71.

Rodríguez Alvarez, A. (2000), "Measuring Allocative Efficency in a Bureaucracy: The Spanish Public Hospitals Sector", Unpublished PhD Thesis, Department of Economics, University of Oviedo.

Sickles, R.C., D.H. Good and L. Getachew (1996), “Specification of Distance Functions using Semi- and Non-parametric Methods with an Application to the Dynamic Performance of Eastern and Western European Air Carriers”, 2nd Biennial Georgia ProductivityWorkshop, University of Georgia.

Shephard, R. W. (1953), *Cost and Production Functions*. Princeton: PrincetonUniversity Press

Weninger, Q. and Strand, I.E. (2003), “An Empirical Analysis of Production Distortions in the Mid-Atlantic Surf Clam And Ocean Quahog Fishery”, *Applied Economics*, 35: 1191-1197.

Zellner, A., Kmenta, J. and Drèze, J. (1966) “Specification and Estimation of Cobb-Douglas Production Function Models”, *Econometrica* 34: 784-795.