Journal of Econometrics 190 (2016) 280–288

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Endogeneity in stochastic frontier models

Christine Amsler [a], Artem Prokhorov [b], Peter Schmidt [a,*]

[a] Michigan State University, USA
[b] University of Sydney, Australia

ARTICLE INFO

Article history:
Available online 23 June 2015

JEL classification:
C10
C26
C36

Keywords:
Endogeneity
Stochastic frontier
Efficiency measurement

ABSTRACT

Stochastic frontier models are typically estimated by maximum likelihood (MLE) or corrected ordinary least squares. The consistency of either estimator depends on exogeneity of the explanatory variables (inputs, in the production frontier setting). We will investigate the case that one or more of the inputs is endogenous, in the simultaneous equation sense of endogeneity. That is, we worry that there is correlation between the inputs and statistical noise or inefficiency.

In a standard regression setting, simultaneity is handled by a number of procedures that are numerically or asymptotically equivalent. These include 2SLS; using the residual from the reduced form equations for the endogenous variables as a control function; and MLE of the system that contains the equation of interest plus the unrestricted reduced form equations for the endogenous variables (LIML). We will consider modifications of these standard procedures for the stochastic frontier setting.

The paper is mostly a survey and combination of existing results from the stochastic frontier literature and the classic simultaneous equations literature, but it also contains some new results.

1. Introduction

In this paper we consider the stochastic frontier (SF) model

$$y_i = \alpha + x_i'\beta + v_i - u_i, \quad i = 1, \ldots, n, \tag{1}$$

where y_i is log output, x_i is a vector of inputs or functions of inputs, v_i is random noise distributed as $N(0, \sigma_v^2)$, and $u_i \geq 0$ represents technical inefficiency. Here i indexes firms and n is the number of firms. We are interested in the case that some of the x's may be endogenous, in the sense that they are correlated with v or u or both. This can occur when there is feedback from either statistical noise or inefficiency to the choice of inputs, or when the inputs influence the level of inefficiency as well as the frontier. Endogeneity needs to be dealt with because the usual procedures for estimating SF models depend on the assumption that the inputs are exogenous.

In a standard regression setting, simultaneity is handled by a number of procedures that are numerically or asymptotically equivalent. These include instrumental variables (2SLS); using the residual from the reduced form equations for the endogenous variables as a control function; and MLE of the system that contains the equation of interest plus the reduced form equations for the endogenous variables (LIML). We will consider modifications of these standard procedures for the SF setting. An important issue is that procedures that are numerically or asymptotically equivalent in the usual linear regression model may not be equivalent for the SF model. Another important issue is that it is definitely not appropriate to insert "fitted values" for the endogenous variables and then proceed with standard SF procedures such as the usual SF MLE.

Modification of the first three of these procedures to the SF model is straightforward. However, appropriate modification of LIML is not straightforward, because it is not clear how best to model the joint distribution of the composed error in the SF model and the error in the reduced form equations for the endogenous inputs. This is a potentially important issue because correlation between the reduced form errors and either noise or inefficiency can be helpful in the decomposition of the composed error into its noise and inefficiency components.

This paper is mostly a survey and combination of existing results from the SF literature and the classic simultaneous equations literature, but it also contains some new results. The material in this paper may be assumed to be part of the existing literature unless it is specifically claimed to be new. The plan of the paper is as follows. In Section 2 we give a brief review of estimation of stochastic frontier models, and in Section 3 we give a brief review of 2SLS

* Correspondence to: Department of Economics, Michigan State University, East Lansing, MI 48824, USA. Tel.: +1 517 355 8381.
E-mail address: schmidtp@msu.edu (P. Schmidt).

http://dx.doi.org/10.1016/j.jeconom.2015.06.013
0304-4076/

and LIML in the usual linear simultaneous equations model. In Section 4 we consider stochastic frontier models with endogeneity, and we discuss how the simple 2SLS and LIML estimators can be modified for use in the stochastic frontier model. We also discuss some issues that are relevant in the case of a translog model (or other nonlinear models). In Section 5 we give an empirical example. Finally, Section 6 gives our concluding remarks.

2. A brief review of estimation in SF models

This section will give a very brief review of the estimation of SF models under exogeneity. This is all standard material but it allows us to define some necessary notation and to summarize the relevant results for readers who are not knowledgeable about SF models.

The most common way to estimate the SF model is by MLE. Following standard terminology, we define $\varepsilon_i = v_i - u_i = y_i - \alpha - x_i'\beta$, which is the *composed error*. We will make the standard assumptions (Aigner et al., 1977) that we have random sampling (and therefore independence) over i, that x_i, v_i and u_i are mutually independent, that $v_i \sim N(0, \sigma_v^2)$, and that $u_i \sim N^+(0, \sigma_u^2)$. (That is, u_i has the so-called *half normal* distribution.) The implied density of ε_i is

$$f_\varepsilon(\varepsilon_i) = \int_0^\infty f_v(\varepsilon_i + u) f_u(u)\, du = \frac{2}{\sigma}\varphi\left(\frac{\varepsilon_i}{\sigma}\right)\Phi\left(-\frac{\lambda\varepsilon_i}{\sigma}\right), \qquad (2)$$

where: $\sigma^2 = \sigma_u^2 + \sigma_v^2$; $\lambda = \sigma_u/\sigma_v$; φ is the standard normal density function; and Φ is the standard normal cdf. We can then form the likelihood function: $\ln L = \sum_i \ln f_\varepsilon(y_i - \alpha - x_i'\beta)$.

The MLE's of the parameters of the model are obtained by maximizing the likelihood function with respect to the parameters $\alpha, \beta, \lambda, \sigma^2$ (or, equivalently, $\alpha, \beta, \sigma_u^2, \sigma_v^2$).

An alternative to MLE is corrected ordinary least squares (COLS), which was defined in Aigner et al. (1977) and Olson et al. (1980). We can make the same assumptions as above, or the slightly weaker assumptions that, conditional on x_i, the first three moments of v_i are the moments of $N(0, \sigma_v^2)$, the first three moments of u_i are the moments of $N^+(0, \sigma_u^2)$, and v_i and u_i are independent. Define $\mu = E(u) = \sqrt{\frac{2}{\pi}}\sigma_u$. Let $\hat\alpha$ and $\hat\beta$ be the OLS estimates when y is regressed on x. These are consistent estimators of $(\alpha - \mu)$ and β, respectively. Now define the OLS residuals $e_i = y_i - \hat\alpha - x_i'\hat\beta$. The second and third sample moments of the residuals are $\hat\sigma_\varepsilon^2 = \frac{1}{n}\sum_i e_i^2$ and $\hat\mu_3' = \frac{1}{n}\sum_i e_i^3$. These are consistent estimators of $\sigma_\varepsilon^2 = \sigma_v^2 + \frac{\pi-2}{\pi}\sigma_u^2$ and $\mu_3' = E[\varepsilon - E(\varepsilon)]^3 = \frac{\pi-4}{\pi}\sqrt{\frac{2}{\pi}}\sigma_u^3$. Solving for σ_u^2 and σ_v^2, in terms of sample quantities we have

$$\hat\sigma_u^2 = \left(\frac{\pi}{\pi-4}\sqrt{\frac{\pi}{2}}\hat\mu_3'\right)^{2/3}, \qquad \hat\sigma_v^2 = \hat\sigma_\varepsilon^2 - \frac{\pi-2}{\pi}\hat\sigma_u^2. \qquad (3)$$

This presumes that $\hat\mu_3' < 0$. (It is the case that $\mu_3' < 0$, but because of estimation error it is possible that $\hat\mu_3' > 0$.) If $\hat\mu_3' > 0$, the so-called *wrong skew problem*, we set $\hat\sigma_u^2 = 0$ (Waldman, 1982). We can now correct the intercept: $\tilde\alpha = \hat\alpha + \sqrt{\frac{2}{\pi}}\hat\sigma_u$. Then the COLS estimates are $\tilde\alpha, \hat\beta, \hat\sigma_u^2, \hat\sigma_v^2$.

There is no real case for preferring the COLS estimate to the MLE in the current setting, but as we will see it is easy to generalize to models with endogeneity.

Once the parameters have been estimated, the ultimate aim is to estimate (or, more properly, *predict*) the values of the inefficiency terms u_i. Under the assumptions that were made in the discussion of MLE above, Jondrow et al. (1982) showed that the distribution of u_i conditional on ε_i is $N^+(a_i, \sigma_*^2)$ where $a_i = -\varepsilon_i\sigma_u^2/\sigma^2$ and $\sigma_*^2 = \sigma_u^2\sigma_v^2/\sigma^2$. Then the prediction of u_i is the mean of this distribution:

$$\hat u_i = E(u_i|\varepsilon_i) = \sigma_*\left[\frac{\varphi(b_i)}{1-\Phi(b_i)} - b_i\right] \quad \text{where } b_i = \varepsilon_i\lambda/\sigma. \qquad (4)$$

To implement this formula, it must be evaluated at the estimated parameters ($\hat\alpha, \hat\beta, \hat\sigma_u^2, \hat\sigma_v^2$ and the implied values of $\hat\lambda$ and $\hat\sigma^2$) and at $\hat\varepsilon_i = y_i - \hat\alpha - x_i'\hat\beta$. (Here, with a slight abuse of notation, $\hat\alpha$, $\hat\beta$, etc. can be either the MLE or the COLS estimates.)

3. A brief review of 2SLS and LIML

This section will give a very brief review of the estimation of linear models (not SF models) when some variables may be endogenous. This is all standard material but the discussion allows us to define some necessary notation and to summarize the relevant results that will be generalized to the stochastic frontier model.

The model of interest is

$$y_i = x_i'\beta + v_i = x_{1i}'\beta_1 + x_{2i}'\beta_2 + v_i, \quad i = 1, \ldots, n. \qquad (5)$$

Here x_{1i} is *exogenous*, meaning $E(v_i|x_{1i}) = 0$ (loosely, x_{1i} is not correlated with v_i) and x_{2i} is *endogenous*, meaning $E(v_i|x_{2i}) \neq 0$ (loosely, x_{2i} is correlated with v_i). There are k_1 variables in x_{1i} and k_2 variables in x_{2i}. The intercept is part of x_{1i}. In matrix terms we write the model as $y = X\beta + v = X_1\beta_1 + X_2\beta_2 + v$ where y is $n \times 1$, X_1 is $n \times k_1$, etc.

We assume there are some *instruments* $z_i = \begin{bmatrix} x_{1i} \\ w_i \end{bmatrix}$ with w_i of dimension $k_w \geq k_2$, so there are at least as many instruments as x's. We say that the model is *exactly identified* when $k_w = k_2$ and that it is *overidentified* when $k_w > k_2$. The instruments are exogenous, in the sense that $E(v_i|z_i) = 0$. We can think in terms of a reduced form for the endogenous variables, which we write in matrix terms as

$$X_2 = Z\Pi + \eta \qquad (6)$$

where $Z = (X_1, W)$ and where η_i is uncorrelated with z_i. Then endogeneity of X_2 corresponds to $\mathrm{cov}(\eta v) \neq 0$.

The problem that endogeneity causes (*simultaneous equations bias*) is that ordinary least squares is inconsistent. This occurs because $E(v|X_2) \neq 0$, and therefore $E(y|X_1, X_2) \neq X_1\beta_1 + X_2\beta_2$, so the regression model is not valid.

We now discuss standard methods to obtain consistent estimates in the presence of endogeneity. These are the methods that we will later generalize to the SF model.

3.1. Two stage least squares (2SLS)

Let $\hat\Pi = (Z'Z)^{-1}Z'X_2$ be the ordinary least squares estimate of the reduced form (6), and let $\hat X_2 = Z\hat\Pi$ and $\hat\eta = X_2 - \hat X_2$ be the corresponding fitted values and residuals, respectively. Also define $\hat X = (X_1, \hat X_2)$. Then the 2SLS (or *instrumental variables*, IV) estimator of β in (5) is

$$\hat\beta = (\hat X'\hat X)^{-1}\hat X'y = (\hat X'X)^{-1}\hat X'y. \qquad (7)$$

This estimator is consistent if the instruments are exogenous (as defined above) and if there are enough relevant instruments (the model is identified).

An alternative approach that is equivalent to 2SLS in the linear model uses a so-called *control function*. In principle, we could control for the effect of η on v by including η in the regression (7). This would be a regression of the form $y = X\beta + \eta\xi + error$, where $\xi = \Sigma_{\eta\eta}^{-1}\Sigma_{\eta v}$ and $error = v - \eta\xi$. Least squares applied

to this equation yields a consistent estimator of β because *error* is uncorrelated with X and with η. Now, we do not observe η, but we do observe the reduced form residuals $\hat{\eta}$. We can enter them as a control function in the regression:

$$y = X\beta + \hat{\eta}\xi + error \tag{8}$$

and the resulting least squares estimator of β is consistent. For obvious reasons, this is also sometimes called the *residual inclusion* method (e.g. Terza et al., 2008).

For the linear model we are now considering, this estimate of β is the same as 2SLS. However, as is well known, this estimate is not the same as 2SLS in nonlinear models like the translog model. We will discuss this point later.

Another feature of the control function regression (8) is that it can be used to test the null hypothesis that the X variables are exogenous. We simply use a standard t or F test to test the significance of the control function variables $\hat{\eta}$. That is, the hypothesis that X is exogenous corresponds to the hypothesis that $\xi = 0$ in Eq. (8). This test is valid asymptotically.

3.2. Limited information maximum likelihood (LIML)

Consider the system of equations made up of the equation of interest plus the reduced form equations for the endogenous regressors. Define $\psi_i \equiv \begin{bmatrix} v_i \\ \eta_i \end{bmatrix}$, the vector of errors from this system. Suppose that we assume that, conditional on the instruments z_i, $\psi_i \sim N(0, \Omega)$ where $\Omega = \begin{bmatrix} \sigma_v^2 & \Sigma_{v\eta} \\ \Sigma_{\eta v} & \Sigma_{\eta\eta} \end{bmatrix}$. (This implies that ψ_i is independent of z_i.) Then we can form a likelihood based on the multivariate normal density for ψ_i. Maximizing this likelihood with respect to the parameters $(\beta, \Pi, \sigma_v^2, \Sigma_{v\eta}, \Sigma_{\eta\eta})$ yields the limited information maximum likelihood estimator of β.

There are several other ways to derive the LIML estimator. See, for example, Schmidt (1976), pp. 169–195. The one just given is easiest to generalize to the SF model.

The LIML estimator is numerically equivalent to the 2SLS estimator if Eq. (5) is exactly identified. In the overidentified case, it is different. In either case it is asymptotically equivalent to (has the same asymptotic distribution as) the 2SLS estimator. Importantly, its consistency and asymptotic distribution do not depend on the correctness of the normality assumption for ψ_i or on the correct specification of the reduced form (6). The LIML estimator is consistent and has the same asymptotic distribution as 2SLS under the same assumptions as were needed for 2SLS, namely, $E(v_i|z_i) = 0$, plus the identification condition and some basic regularity conditions on the distribution of v.

3.3. Other sources of identification

In the discussion above, we had exogenous regressors x_{1i}, endogenous regressors x_{2i} and some "outside instruments" w_i. Our identifying assumptions are that x_{1i} and w_i are uncorrelated with the error v_i and that there are enough instruments (the dimension of w_i is at least as big as the dimension of x_{2i}). However, under additional assumptions we can create more instruments (moment conditions), either to create identification when we do not have enough instruments, or to improve efficiency of estimation even in cases where we do already have enough instruments for identification. This will generally be possible because if the expectation of v_i conditional on x_{1i} (or w_i) equals zero, then any function of x_{1i} (or w_i) will be uncorrelated with v_i. Alternatively, additional moment conditions may be based on assumptions about higher moments of the data or the form of the error distribution.

An early example is Lewbel (1997). He has (in our notation) an exogenous regressor x_{1i} and an endogenous regressor x_{2i}, and the endogenous regressor is endogenous because it is a mismeasured version of a second regressor x_{2i}^* that would be exogenous but is unobservable. He then creates instruments using assumptions on the higher moments of the data. For example, one of his instruments is $(x_{1i}-\bar{x}_1)(x_{2i}-\bar{x}_2)$, and this is a legitimate instrument if $E(v_i|x_{1i}, x_{2i}^*) = 0$ and if x_{1i} is uncorrelated with the *square* of the measurement error.

Lewbel (2012) follows somewhat similar lines. In this paper he has a system of two equations with exogenous variables x and errors $\varepsilon = (\varepsilon_1, \varepsilon_2)'$. Lewbel creates additional moment conditions by restricting the correlations of $\varepsilon\varepsilon'$ with x. These moment conditions are useful so long as $E(\varepsilon\varepsilon'|x)$ is not constant.

Hansen et al. (2010) attempt to increase the efficiency of the 2SLS estimator by using moment conditions that arise from assuming a flexible parametric form for the error density. This would lead to a likelihood that could be maximized to obtain a consistent estimator, if the regressors were exogenous. That is, if f is the assumed density, the first order condition for its maximization with respect to β is $\sum_i x_i'\rho\left(y_i - x_i'\beta\right) = 0$ where $\rho(\varepsilon) = \partial \ln f(\varepsilon)/\partial\varepsilon$, and it reflects the moment conditions that $E\left[x_i'\rho(\varepsilon_i)\right] = 0$. This moment condition does not hold if x_i is endogenous. Therefore we replace the endogenous regressors x_i with the exogenous instruments z_i, assumed to be independent of ε_i, to obtain the moment conditions $Ez_i'\rho\left(y_i - x_i'\beta\right) = 0$. Estimation of β can be based on these moment conditions alone, or these moment conditions could be used together with the 2SLS moment conditions $Ez_i'\left(y_i - x_i'\beta\right) = 0$. If we use them as extra moment conditions, these extra moment conditions will generally improve efficiency, except when $\rho(\varepsilon)$ is proportional to ε, as it is in the normal case. This is relevant to the SF case because in the SF model we do have a specific non-normal error distribution.

Finally, there are papers that claim to establish identification and consistent estimation without any exogeneity assumptions (a truly instrument-free approach). An example is Park and Gupta (2012). They assume that the marginal distribution of ε_i is normal, and they use nonparametric density estimation to identify the marginal distribution of x_i. Then they assume a Gaussian copula to create a joint distribution for ε_i and x_i that has the appropriate marginal distributions. This joint distribution implies a likelihood that can be maximized, presumably to obtain a consistent estimator of β. However, no proof of consistency is given. This approach is suspect because it misses the point that the problem with endogeneity is fundamentally an identification problem. The issue is not that we cannot construct a likelihood, but rather that different parameter values give the same (maximal) value of the likelihood.

4. SF models with endogeneity

We now wish to embed the endogeneity problem into the SF model. So we consider the SF model

$$y_i = \alpha + x_i'\beta + \varepsilon_i = \alpha + x_{1i}'\beta_1 + x_{2i}'\beta_2 + \varepsilon_i, \tag{9}$$

where, as in Section 2, $\varepsilon_i = v_i - u_i = y_i - \alpha - x_i'\beta$. In matrix terms,

$$y = 1_n\alpha + X_1\beta_1 + X_2\beta_2 + \varepsilon. \tag{10}$$

where 1_n is an $n \times 1$ vector of ones (representing the constant in the regression, which we now distinguish from the rest of X_1). We assume, as we did in Section 2, that we have random sampling (and therefore independence) over i, that v_i and u_i are mutually independent, that $v_i \sim N(0, \sigma_v^2)$, and that $u_i \sim N^+(0, \sigma_u^2)$. However, we now distinguish the exogenous X_1 from the endogenous X_2. We assume a set of instruments $Z = (1_n, X_1, W)$, and we assume that the identification condition holds, which requires $k_w \geq k_2$, as in Section 3. We assume that u_i and v_i are independent of the instruments z_i (but not of the endogenous

variables x_{2i}), though this assumption can be weakened for some of the estimators we are about to consider.

Some of the estimators that we will consider require specification of the reduced form (6). For the moment we put no restrictions on the relationship between the reduced form error η_i and either v_i or u_i.

For the SF model, neither the MLE nor COLS will be consistent when we have endogeneity. We now proceed to consider modifications of these procedures that are consistent.

4.1. Corrected 2SLS (C2SLS)

This is a straightforward generalization of COLS, which perhaps surprisingly does not appear to have been discussed in the literature. For COLS, we estimated the model by OLS. Then we estimated σ_u^2 and σ_v^2 from the second and third moments of the residuals, and finally we corrected the intercept using the estimate of σ_u. For C2SLS, the first step is to estimate the model (9) by 2SLS, using the instruments Z. Let the 2SLS estimates be $\hat{\alpha}$, $\hat{\beta}_1$ and $\hat{\beta}_2$. These are consistent estimators of $\alpha - \mu$, β_1 and β_2, where as before $\mu = E(u)$. The second step is to construct the 2SLS residuals e_i, calculate the second and third moments of the residuals, and then calculate the implied estimates $\hat{\sigma}_u^2$ and $\hat{\sigma}_v^2$ as in Eq. (3). An important detail is to use the right residuals. The correct residuals to use are

$$e_i = y_i - \hat{\alpha} - x_1'\hat{\beta}_1 - x_2'\hat{\beta}_2. \tag{11}$$

Note that this is *not* the same as $y_i - \hat{\alpha} - x_1'\hat{\beta}_1 - \hat{x}_2'\hat{\beta}_2$, which would be incorrect. Finally, the third step is to correct the intercept: $\tilde{\alpha} = \hat{\alpha} + \sqrt{\frac{2}{\pi}}\hat{\sigma}_u$

A somewhat similar procedure was considered by Guan et al. (2009). They estimated a stochastic frontier model using a generalized method of moments approach that is a generalization of 2SLS. Also their application involved panel data and therefore some details are considerably different than here.

4.2. The approach of Hansen, McDonald and Newey

An alternative to C2SLS is the approach of Hansen et al. (2010). Like C2SLS, it depends only on the correct specification of the SF model, but it makes use of the distributional assumptions for v_i and u_i in a way that is more analogous to the MLE estimator that would be used in the case of exogeneity.

Unlike Hansen, McDonald and Newey, in the SF context we have a "true" likelihood, not just an approximate one. The first order conditions for the maximization of the likelihood under exogeneity are given by Aigner et al. (1977, p. 27). The (population) moment conditions that are the analogue to these first order conditions are:

$$E\left[-1 + \frac{1}{\sigma^2}(y_i - x_i'\beta)^2\right] = 0 \tag{12a}$$

$$E\left[\left(y_i - x_i'\beta\right)\frac{\varphi_i}{1-\Phi_i}\right] = 0 \tag{12b}$$

$$E\left[\frac{1}{\sigma}x_i\left(y_i - x_i'\beta\right) + \lambda x_i\frac{\varphi_i}{1-\Phi_i}\right] = 0 \tag{12c}$$

where φ_i and Φ_i are the standard normal density and cdf, evaluated at $\frac{\lambda}{\sigma}(y_i - x_i'\beta)$. Under exogeneity, the MLE is the generalized method of moments (GMM) estimator based on these moment conditions.

The moment conditions (12a) and (12b) depend on the assumptions about the errors, not about the relationship of the errors to the x_i, and thus they hold under exogeneity or endogeneity of the x_i. The moment condition (12c) would not hold under endogeneity. However, given the instruments z_i, we can replace it with the moment condition

$$E\left[\frac{1}{\sigma}z_i\left(y_i - x_i'\beta\right) + \lambda z_i\frac{\varphi_i}{1-\Phi_i}\right] = 0, \tag{12d}$$

which corresponds to $Ez_i'\rho(y_i - x_i'\beta) = 0$ in the notation of Section 3.3. An important point is that the validity of this moment condition requires independence of z_i and ε_i, not just $E(\varepsilon_i|z_i) = 0$. If we are willing to make this independence assumption, we can estimate the parameters of the model consistently by GMM based on the moment conditions (12a), (12b) and (12d), assuming that enough of the instruments are relevant, exactly as was required for C2SLS.

A difference between this procedure and the procedure of Hansen, McDonald and Newey is that, because we are assuming a correctly specified parametric likelihood, there is no need for a separate estimation step to estimate the nuisance parameters in the distribution of the errors. GMM simply estimates these along with β.

4.3. LIML: v is correlated with η, but u is independent of v and η

We now consider MLE of the system consisting of the SF model (9) and the reduced form equation (6). Endogeneity occurs when the reduced form error η_i is correlated with either v_i or u_i or both. In this subsection we will consider the analytically tractable case that u_i is independent of $\psi_i \equiv \begin{bmatrix} v_i \\ \eta_i \end{bmatrix}$, so that endogeneity is due to correlation between v_i and η_i.

More explicitly, we assume that, conditional on z_i, $\psi_i \sim N(0, \Omega)$ where $\Omega = \begin{bmatrix} \sigma_v^2 & \Sigma_{v\eta} \\ \Sigma_{\eta v} & \Sigma_{\eta\eta} \end{bmatrix}$; $u_i \sim N^+\left(0, \sigma_u^2\right)$, and u_i and ψ_i are independent. This model has been considered by Kutlu (2010), Karakaplan and Kutlu (2013), Tran and Tsionas (2013) and Tsionas et al. (2013). The derivation of the likelihood by Kutlu (2010), Karakaplan and Kutlu (2013) and Tran and Tsionas (2013) was based on the Cholesky decomposition of the variance matrix. In Appendix A, we derive the likelihood using a conditioning argument that is novel and perhaps more intuitive. This corresponds to the factorization of the density of the endogenous variables conditional on the instruments: $f(y, X_2|Z) = f(y|X_2, Z) \cdot f(X_2|Z)$. Whichever derivation is used, the likelihood is:

$$\ln L = \ln L_1 + \ln L_2 \tag{13a}$$

where

$$\ln L_1 = -\frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_i(y_i - \alpha - x_i'\beta - \mu_{ci})^2 + \sum_i \ln\left[\Phi\left(\frac{-\lambda(y_i - \alpha - x_i'\beta - \mu_{ci})}{\sigma}\right)\right] \tag{13b}$$

and

$$\ln L_2 = -\frac{n}{2}\ln\left|\Sigma_{\eta\eta}\right| - \frac{1}{2}\sum_i\left(x_{2i}' - z_i'\Pi\right)\Sigma_{\eta\eta}^{-1}\left(x_{2i} - \Pi' z_i\right), \tag{13c}$$

where Φ is the standard normal cdf and:

$$\mu_{ci} = \Sigma_{v\eta}\Sigma_{\eta\eta}^{-1}(x_{2i} - \Pi' z_i), \qquad \sigma^2 = \sigma_u^2 + \sigma_c^2,$$
$$\lambda = \frac{\sigma_u}{\sigma_c}, \qquad \sigma_c^2 = \sigma_v^2 - \Sigma_{v\eta}\Sigma_{\eta\eta}^{-1}\Sigma_{\eta v}. \tag{13d}$$

We can obtain the MLE by maximizing the likelihood with respect to the parameters $(\alpha, \beta, \sigma_v^2, \sigma_u^2, \Sigma_{v\eta}, \Sigma_{\eta\eta}, \Pi)$. Tran and Tsionas (2013) consider a GMM procedure based on the score of

the likelihood, which is very similar to the MLE. Alternatively, as suggested by Kutlu (2010), we can use a two-step procedure. Let $\theta_1 \equiv (\alpha, \beta, \sigma_v^2, \sigma_u^2, \Sigma_{v\eta})$ be the "first set" of parameters and $\theta_2 \equiv (\Sigma_{\eta\eta}\Pi)$ be the "second set". Then Step 1 is to estimate the second set of parameters from the reduced form equations, that is by maximizing $\ln L_2$. This yields $\hat{\Pi} =$ OLS of X_2 on Z and $\hat{\Sigma}_{\eta\eta} = \frac{1}{n}\sum_i (x_{2i} - \hat{\Pi}' z_i)(x_{2i} - \hat{\Pi}' z_i)'$. Step 2 is to estimate the first set of parameters by maximizing $\ln L_1$ taking the estimates of the second set of parameters as given. This is essentially a control function approach where the control function in the SF model equation is $\Sigma_{\eta\eta}^{-1}(x_{2i} - \Pi' z_i)$ and the coefficients are $\Sigma_{v\eta}$ (Or the control function is $(x_{2i} - \Pi' z_i)$ and the coefficients are $(\Sigma_{v\eta}\Sigma_{\eta\eta}^{-1})$.)

The two-step procedure is generally different from the MLE because it ignores the information about $\Sigma_{\eta\eta}$ and Π contained in $\ln L_1$. That is, $\Sigma_{\eta\eta}$ and Π appear implicitly in $\ln L_1$ through μ_{ci} and σ_c^2. A practical implication is that the conventionally-calculated standard errors from the Step 2 estimation are not correct. They need to be adjusted to reflect the fact that $\Sigma_{\eta\eta}$ and Π have been estimated. Kutlu (2010) suggests constructing standard errors using bootstrapping, which is a valid suggestion. Alternatively, an analytical approach is possible. See Murphy and Topel (2002) or Wooldridge (2010), Section 12.4.2, equation (12.41).

An exception to the conclusion of the previous paragraph is the case that $E\nabla_{\theta_1\theta_2} \ln L_1 = 0$. This will hold when $\Sigma_{v\eta} = 0$. In this case the estimation error in $\hat{\theta}_2$ does not affect the asymptotic distribution of $\hat{\theta}_1$, the two-step estimator is asymptotically as efficient as the MLE, and the conventionally-calculated standard errors for the second step estimation of θ_1 are valid. A practical implication is that we can test the null hypothesis of exogeneity of x_{2i}, which corresponds to $\Sigma_{v\eta} = 0$, using a standard F-test of the hypothesis that the coefficients of the control function $(x_{2i} - \Pi' z_i)$ equal zero.

We will call the MLE just discussed the "LIML" estimator because it is logically similar to the LIML estimator discussed in Section 3.3. However, there is an important difference, because in the discussion of Section 3.3 the consistency of the LIML estimator did not require the correctness of the linear reduced form specification or normality of its error, whereas here it does require both of these things. In fact, we now have a completely specified likelihood-based model for which we are calculating the MLE. The only sense in which we are operating with limited information is that we have not specified a structural (economic based) specification for the distribution of the endogenous variables. We will return to this point later.

4.4. Prediction of u_i

The usual predictor of u_i is $\hat{u}_i = E(u_i|\varepsilon_i)$, as suggested by Jondrow et al. (1982). An original and important observation is that given the model of Section 4.3 we can define a better predictor of u_i, namely $\tilde{u}_i = E(u_i|\varepsilon_i, \eta_i)$. Even though u_i is independent of η_i, η_i is correlated with, and therefore informative about, v_i. Therefore, conditional on $\varepsilon_i = v_i - u_i$, η_i is informative about u_i.

Suppose that we transform $(u_i, \varepsilon_i, \eta_i)$ into $(u_i, \tilde{\varepsilon}_i, \eta_i)$ where

$$\tilde{\varepsilon}_i = \varepsilon_i - \mu_{ci} \quad \text{where } \mu_{ci} = \Sigma_{v\eta}\Sigma_{\eta\eta}^{-1}\eta_i. \tag{14}$$

Then η_i is independent of $(u_i, \tilde{\varepsilon}_i)$, so that $E(u_i|\varepsilon_i, \eta_i) = E(u_i|\tilde{\varepsilon}_i, \eta_i) = E(u_i|\tilde{\varepsilon}_i)$. We show in Appendix B that the distribution of u_i conditional on $\tilde{\varepsilon}_i$ (or conditional on ε_i and η_i) is $N^+(\mu_{*i}, \sigma_*^2)$ where $\mu_{*i} = -\frac{\sigma_u^2}{\sigma^2}\tilde{\varepsilon}_i$ and $\sigma_*^2 = \frac{\sigma_u^2\sigma_c^2}{\sigma^2}$. This is essentially the same as Theorem 1 of Jondrow et al. (1982), with σ_c^2 replacing their σ_v^2 in these definitions and in the definitions of λ and σ^2, and with $\tilde{\varepsilon}_i$ replacing their ε_i. This leads to the explicit expression:

$$\tilde{u}_i = E(u_i|\varepsilon_i, \eta_i) = E(u_i|\tilde{\varepsilon}_i) = \sigma_* [\Lambda(h_i) - h_i] \tag{15}$$

where $h_i = \frac{\lambda}{\sigma}\tilde{\varepsilon}_i$ and where $\Lambda(h) = \varphi(h)/[1 - \Phi(h)]$ is the standard normal hazard function. This is similar to equation (3) of Jondrow et al. (1982).

This is a better predictor than the former predictor, $\hat{u}_i = E(u_i|\varepsilon_i)$, because $\sigma_c^2 < \sigma_v^2$. More explicitly, consider the following well-known identity, which holds for any random variable u and for any set of random variables m that are used to predict u:

$$\text{var}(u) = \text{var}_m [E(u|m)] + E_m [\text{var}(u|m)]. \tag{16}$$

(Here E_m and var_m are respectively the mean and the variance over the distribution of m.) This is the usual decomposition of the variance of u into its explained (by m) and unexplained parts. When we enlarge the set m the explained variance increases and the unexplained variance decreases. In our specific case, $\tilde{u} = E(u|\varepsilon, \eta)$ and $\hat{u} = E(u|\varepsilon)$, so on average $\text{var}(\tilde{u}) < \text{var}(\hat{u})$.

The obvious disadvantage of the new estimator of u_i is that now the reduced form must in fact be a correctly specified model with normal errors. Unlike in the standard linear model this is now a substantive assumption.

Either $\tilde{u}_i$ or $\hat{u}_i$ has to be evaluated at the estimated parameters. Any consistent estimator can be used (e.g. C2SLS or LIML). Given that the reduced form model must be correctly specified for $\tilde{u}_i$ to make sense, it is natural to think of using the LIML estimates, but this is not required. All that is required is that the model that would be assumed for LIML be correctly specified.

4.5. LIML: η is correlated with both v and u

We now wish to allow η to be correlated with both v and u. This case has not been considered in the literature. We will consider the slightly more general case in which η, v and u are all potentially correlated. The issue we face is to find a joint distribution such that (i) the marginal distribution of u is half-normal; (ii) the marginal distribution of $\psi = \binom{v}{\eta}$ is multivariate normal; and (iii) u and ψ are correlated in a sensible way.

4.5.1. Two things that do not work

There are (at least) two intuitive approaches to this problem that do not work.

The first approach that does not work is to specify a multivariate normal distribution with zero mean for η, v and u, and then truncate this distribution by requiring $u \geqslant 0$. This does generate a half-normal marginal for u. However, the marginal for ψ is no longer normal. See, e.g., Horrace (2005, p. 214). In fact, the marginal distribution of ψ is skew-normal (Azzalini, 2005, p. 163).

A second approach that does not work is to specify a multivariate normal distribution with zero mean for η, v and u^*, and then define $u = |u^*|$. This generates the correct marginals, but now u is not correlated with ψ. See Schmidt and Lovell (1980), who used this construction in a context where they wanted u and ψ to be dependent but not correlated.

4.5.2. Specification of a joint distribution using a copula

A copula is a joint distribution whose marginal distributions are uniforms. The copula captures the dependence in a joint distribution. Suppose that we have a joint distribution of two random variables (x_1, x_2) with density $h(x_1, x_2)$, and suppose that the marginal density and cdf of x_j are $f_j(x_j)$ and $F_j(x_j)$ respectively, $j = 1, 2$. Then a well-known result, Sklar's Theorem (e.g. Nelsen (2006, p. 15)) says that

$$h(x_1, x_2) = c(F_1(x_1), F_2(x_2)) \cdot f_1(x_1) \cdot f_2(x_2). \tag{17}$$

Here "c" is the density of the copula distribution, and its arguments are the uniform random variables $z_1 = F_1(x_1)$ and $z_2 = F_2(x_2)$. More explicitly, Sklar's Theorem says that the joint distribution

h determines the marginal distributions and also the copula c (and c is determined uniquely if h is a continuous distribution). Conversely, if we specify the marginals and we specify a copula c, this determines the joint distribution h, and h does have the marginals with which we started.

The result in (17) extends in the obvious way to the case of more than two random variables. In our case there are $k_2 + 2$ random variables: v, u and η.

In our context, we begin with a half-normal marginal for u and normal marginals for the elements of $\psi = (v, \eta')'$. Then we need to pick a copula, with the obvious candidate being the multivariate normal (or "Gaussian") copula. This is partly a matter of convenience, but also it implies that the distribution of ψ is multivariate normal, which is what we want to assume. The assumed marginals and copula imply a joint density for u and ψ; that is, for u, v and η. By a simple substitution this gives us the joint density for u, $\varepsilon + u$ and η. However, to form the likelihood we need the joint density of ε and η; that is, we need to integrate out u from the joint density for u, $\varepsilon + u$ and η. Perhaps for some clever choice of copula it would be possible to do this analytically, but we have not found such a choice.

An alternative is to evaluate the joint density by numerical methods. We want to calculate

$$\begin{aligned} f_{\varepsilon,\eta}(\varepsilon, \eta) &= \int_0^\infty f_{u,v,\eta}(u, \varepsilon + u, \eta)\, du \\ &= \int_0^\infty \left[\frac{f_{u,v,\eta}(u, \varepsilon + u, \eta)}{f_u(u)}\right] f_u(u) du \\ &= E_u\left[\frac{f_{u,v,\eta}(u, \varepsilon + u, \eta)}{f_u(u)}\right]. \end{aligned} \tag{18}$$

Here E_u denotes the expectation over the distribution of u. We can evaluate (approximate) this by the averaging over draws from the distribution of u. That is, if we draw u_s, $s = 1, \ldots, S$, from the distribution of u, our simulated value of $f_{\varepsilon,\eta}(\varepsilon, \eta)$ is

$$\hat{f}_{\varepsilon,\eta}(\varepsilon, \eta) = \frac{1}{S}\sum_{s=1}^{S}\left[\frac{f_{u,v,\eta}(u_s, \varepsilon + u_s, \eta)}{f_u(u_s)}\right]. \tag{19}$$

This leads us to the simulated log likelihood

$$\ln \hat{L} = \sum_{i=1}^{n} \ln \hat{f}_{\varepsilon,\eta}(\varepsilon_i, \eta_i), \tag{20}$$

where, as before, $\varepsilon_i = y_i - \alpha - x_i'\beta$ and $\eta_i = x_{2i} - \Pi' z_i$. Very similar applications of simulation to evaluate a likelihood can be found in Greene (2005, p. 24), and Amsler et al. (2014).

The maximum simulated likelihood estimator (MSLE) is obtained by maximizing the simulated likelihood with respect to the parameters. The parameters are α, β, Π, σ_u^2, σ_v^2, $\sigma_{\eta 1}^2, \ldots, \sigma_{\eta k_2}^2$ (where $\sigma_{\eta j}^2$ is the variance of η_j) plus whatever parameters appear in the copula. If we use the Gaussian copula for u, v and η, and if we let R denote the correlation matrix in this copula, the parameters in the copula that we need to estimate are the $\frac{1}{2}(k_2 + 1)(k_2 + 2)$ distinct off-diagonal elements of R.

Standard results (e.g. Gourieroux and Monfort, 1993) indicate that the MSLE is consistent and has the same asymptotic distribution as the MLE based on $f_{\varepsilon,\eta}(\varepsilon, \eta)$, provided that $n \to \infty$, $S \to \infty$ and $S/\sqrt{n} \to \infty$, and suitable regularity conditions hold. That is, the randomness due to the simulation of the likelihood can be ignored if we use a large enough number of random draws in the simulation. Unfortunately the Gourieroux and Monfort result does not give us practical guidance as to how large S should be for a given sample size n.

A remaining issue is to calculate $\tilde{u}_i = E(u_i|\varepsilon_i, \eta_i)$. Following Amsler et al. (2014), we can do this based on draws from the joint distribution of v, u, η, by kernel methods. More explicitly, since we have specified the marginal distributions and a copula, we have defined the joint distribution of v, u, η. If the copula is simple enough that we can draw simulated observations from it (and this would be true of the Gaussian copula), we can then draw simulated observations from the joint distribution of v, u, η. Call these v_s, u_s, η_s, $s = 1, \ldots, S$. We can then calculate $\varepsilon_s = v_s - u_s$ so that we have a sample on ε_s, u_s, η_s. From this sample we can estimate $E(u_i|\varepsilon_i, \eta_i)$ nonparametrically, by nearest neighbors or kernel methods. This estimator is consistent as $S \to \infty$.

A potential advantage of this approach is that we should be able to obtain more precise predictions of u_i when it is correlated with ψ_i than when it is not. In the LIML model of Section 4.3, u_i was not correlated with η_i, and the only reason that η_i was useful in predicting u_i is that u_i is correlated with η_i *conditional on* ε_i. In the present model, however, u_i and η_i are correlated "directly" (not just through ε_i) and, intuitively speaking, the case for the usefulness of η_i in predicting u_i is stronger.

One obvious difficulty with the approach of this section is the need to specify a copula. If u and ψ have a continuous joint distribution, then there is a true copula, but we do not know what it is. The Gaussian copula is often used for reasons of convenience, and in our case it does imply the desired multivariate normal distribution for ψ, but there is no good reason to think that it properly captures the dependence between u and ψ. It is well known that the Gaussian copula is "comprehensive" in the sense that it can accommodate any range of correlation between the uniform random variables $F_u(u)$ and $F_{\psi_i}(\psi_i)$. Also, because the implied distribution of ψ is multivariate normal, it can accommodate any range of correlation in ψ. However, it cannot necessarily accommodate any range of correlation between u and ψ. Still, assuming a Gaussian copula is at least more general than assuming that u and ψ are independent, as we did in Section 4.3, because the Gaussian copula contains the independence copula as a special case (correlations equal to zero).

Another difficulty is that, given the current state of computing technology, the application of copulas is effectively limited to the case of a small number of variables. For example, in our empirical application of Section 5, there were no serious computational problems, but we had only four endogenous inputs.

4.6. The translog case

All of the discussion so far has been for the linear case (e.g. Cobb–Douglas production function). Now we will discuss the translog case, though the same issues that arise here would arise in other nonlinear specifications.

For expositional simplicity we will restrict our attention to the case of two inputs, both of which are endogenous. The translog model is

$$\begin{aligned} \ln y_i = {} & \alpha + \beta_1 \ln x_{1i} + \beta_2 \ln x_{2i} + \gamma_1 (\ln x_{1i})^2 \\ & + \gamma_2 (\ln x_{2i})^2 + \delta (\ln x_{1i})(\ln x_{2i}) + \varepsilon_i. \end{aligned} \tag{21}$$

Note that although we have only two endogenous inputs, we have five endogenous variables, because functions of endogenous variables are in general endogenous. Also note that to make the correspondence to the SF model we need to interpret ε_i as being demeaned, that is, as equal to $v_i - u_i^*$ where $u_i^* = u_i - E(u_i)$, and correspondingly α is equal to the original intercept minus $E(u_i)$, as in the discussion of COLS in Section 2.

In this case we can use 2SLS but it is cumbersome. We need to have five valid instruments, not just two. Valid here means that $E(\varepsilon_i|z_i) = 0$. Of course, we can create five instruments out of two because, if $E(\varepsilon_i|z_i) = 0$, then ε_i is uncorrelated with any function of z_i. For example, if input prices are exogenous, then we

could use their logarithms, the squares of their logarithms, and the cross-product of their logarithms, thus essentially copying the translog functional form in the reduced form equations. One might reasonably worry about the marginal strength of some of these instruments.

A well-known but important result is that, for this nonlinear model, a control function regression is not the same as 2SLS. More precisely, a control function regression in which we used five control functions, namely the residuals from the five reduced form equations, would be the same as 2SLS. However, under some additional assumptions, we can obtain consistent estimators using only two control functions, not five. This point has been made by a number of people, including Blundell and Powell (2004), Terza et al. (2008) and Wooldridge (2010). Our discussion will follow Wooldridge. Consider the system consisting of the translog function (21) plus two reduced form equations:

$$q_i = \Pi z_i + \eta_i \tag{22}$$

where z_i is the vector of instruments and where $q_i = \begin{bmatrix} \ln x_{1i} \\ \ln x_{2i} \end{bmatrix}$. We need to assume that ε_i and η_i are independent of z_i, which is a stronger assumption than the usual exogeneity condition that $E(\varepsilon_i|z_i) = 0$ and the usual linear projection condition that z_i and η_i are uncorrelated. We also need to assume that $E(\varepsilon_i|\eta_i)$ is linear in η_i; that is, $E(\varepsilon_i|\eta_i) = \eta_i'\rho$. This last assumption would follow if v_i and η_i are jointly normal and independent of u_i, but it is technically slightly weaker since joint normality is sufficient but not necessary for linearity of the conditional expectation. Under these assumption, the argument in Wooldridge (2010, pp. 128–129) establishes that

$$E(\ln y_i|\ln x_{1i}, \ln x_{2i}) = \alpha + \beta_1 \ln x_{1i} + \beta_2 \ln x_{2i} + \gamma_1 (\ln x_{1i})^2 + \gamma_2 (\ln x_{2i})^2 + \delta (\ln x_{1i})(\ln x_{2i}) + \eta_i'\rho. \tag{23}$$

The implication is that $\hat{\eta}_i$, the vector of reduced form residuals from least squares estimation of (22), can be used as a control function in an operational version of (23) to obtain consistent estimators of the translog parameters. Note that the dimension of $\hat{\eta}_i$ is two; we have two control functions, not five.

Under the assumptions made above, the control function estimates are more efficient than the 2SLS estimates. But they do rely on more assumptions and they will not be consistent if those assumptions are violated.

The LIML estimation of the system consisting of (22) and (23) has not been considered in the literature, but similar results would apply. In particular we have only two reduced form equations, not five. Now we make the same assumptions as were made in Section 4.3, so that v_i and η_i are jointly normal, and u_i is half-normal and independent of v_i and η_i. As noted above, these assumptions are slightly stronger than are needed for the control function approach. Then the joint density of ε_i and η_i is as given in Eq. (A.4) of Appendix A, and we obtain the likelihood by substituting $\eta_i = q_i - \Pi z_i$ and $\varepsilon_i = y_i - \alpha - \beta_1 \ln x_{1i} - \beta_2 \ln x_{2i} - \gamma_1 (\ln x_{1i})^2 - \gamma_2 (\ln x_{2i})^2 - \delta (\ln x_{1i})(\ln x_{2i})$. (The Jacobian of this transformation equals one despite the nonlinearity in variables.) This LIML estimator should be more efficient than the control function estimator, because it takes into account the composed error in the production function, but we will not attempt to prove that in this paper.

4.7. Structural approaches

In the discussion above, the source of the instruments Z and the justification for the linearity of the reduced form equations for X_2 are vague. An alternative to this kind of unrestricted approach is to make behavioral assumptions that indicate how the endogenous inputs are generated. (This might reasonably be called a "full-information" or a "structural" approach.) In the frontiers literature, an early example of this approach is Schmidt and Lovell (1979), who assumed a Cobb–Douglas production function in which the output level and input prices were exogenous, the input levels were endogenous, and the firm minimized cost. They estimated the system consisting of the production function and the first-order conditions for cost minimization. The first order conditions for cost minimization are linear in output and input prices, and more fundamentally there are no new parameters in these equations; the parameters in them are functions of the parameters in the production function. A somewhat similar early paper is Greene (1980), who considered a translog cost function and the associated cost-minimizing share equations. There have been many papers since then along these same lines, but with different functional forms for the equation of interest and different assumptions about the behavior of the firm. Examples include Kumbhakar (1987), Atkinson and Cornwell (1994), Kumbhakar (1997), Kumbhakar and Tsionas (2005, 2009), to name just a few.

5. Empirical example

Our empirical example is an analysis of data on dairy farms in Northern Spain. The data set is an extension of the data analyzed by Alvarez and Arias (2004), to whom we are grateful for providing the data. They analyzed data on 196 farms over six years (1993–1998) whereas our data is a balanced panel (except for a few exclusions due to missing data) on 137 farms over 12 years (1999–2010). However, we ignore the panel nature of the data in order to make the empirical analysis better reflect the techniques described earlier in the paper.

We use a Cobb–Douglas production function. (They used a translog, but we wanted in the first instance to keep things simple.) The dependent variable is (the log of) milk production in liters. There are five inputs: Labor, in man-equivalent units; Cows, the number of milking cows; Feed, number of kilograms of feedstuff fed to the cows; Land, the number of hectares of land devoted to pasture and crops; and Rough, expenses in euros incurred in producing roughage on the farm. For more detail on the data, see Alvarez and Arias (2004). We also include an intercept and 11 dummy variables for year in the production functions specification, so that we have 17 regressors.

The first two pairs of columns in Table 1 give the results for COLS and the SF MLE, both of which assume exogeneity. These two methods give rather similar results (a common empirical finding). Returns to scale are about 1.06, which is not unreasonable. The inputs with the biggest coefficients (elasticities) by far are Cows and Feed. The coefficient of Labor is small and only marginally significant (which agrees with the result of Alvarez and Arias). The coefficient of Land is negative, and small in magnitude but significantly different from zero. Zero would not be hard to believe, but negative seems unreasonable. Nevertheless, a negative coefficient for Land will occur for all of our methods of estimation, so we will just accept it.

Next we allow for possible endogeneity of the five inputs. We have 22 variables that we assume to be exogenous: PMilk, the price of milk; PFeed, the price of feed; Landown, the percentage of land owned by the farmer; seven dummy variables reflecting membership in one of eight agricultural cooperatives; plus the constant and the 11 time dummies.

We test the exogeneity of the inputs using t-tests for the significance of the control functions (reduced form residuals), as described at the end of Section 3.1. We find that Land is exogenous and the other four inputs are endogenous, an intuitively reasonable outcome. So now we have 23 exogenous variables.

The C2SLS estimates are strikingly different from COLS and SF MLE. Returns to scale are higher, about 1.18. Now the coefficient of Cows is just slightly less than one. The coefficient of Labor is higher

Table 1
Estimates of the production function parameters.

	COLS		SF MLE		C2SLS		Hansen et al.		LIML		2-Step	Copula	
	Est	Std Err	Est	Std Err	Est	Std Err	Est	Std Err	Est	Std Err	Est	Est	Std Err
Const	0.0595		0.0602	0.0127	0.0679		0.0409	0.0231	0.0399	0.0282	0.0368	0.0375	0.0390
Labor	0.0250	0.0121	0.0184	0.0118	0.1246	0.0628	0.1166	0.0513	0.1298	0.0675	0.1169	0.1168	0.0677
Cows	0.5802	0.0205	0.6354	0.0216	0.9704	0.0827	0.9666	0.0669	0.9925	0.0870	0.9685	0.9684	0.0877
Feed	0.3723	0.0122	0.3438	0.0125	0.0889	0.0472	0.1119	0.0357	0.0722	0.0582	0.1156	0.1154	0.0605
Land	−0.0234	0.0094	−0.0268	0.0089	−0.1136	0.0177	−0.1081	0.0151	−0.1208	0.0187	−0.1047	−0.1048	0.0204
Rough	0.1042	0.0064	0.0982	0.0063	0.1108	0.0260	0.0858	0.0182	0.1125	0.0286	0.0911	0.0909	0.0278
D00	0.0167	0.0161	0.0165	0.0153	0.0093	0.0202	0.0107	0.0196	0.0032	0.0223	0.0096	0.0096	0.0217
D01	0.0213	0.0158	0.0207	0.0149	0.0256	0.0221	0.0256	0.0204	0.0197	0.0247	0.0244	0.0245	0.0247
D02	0.0462	0.0159	0.0400	0.0151	0.0406	0.0244	0.0365	0.0218	0.0310	0.0274	0.0354	0.0355	0.0271
D03	0.0246	0.0166	0.0203	0.0157	0.0163	0.0250	0.0144	0.0231	0.0090	0.0278	0.0132	0.0132	0.0274
D04	0.0404	0.0158	0.0394	0.0149	0.0430	0.0256	0.0434	0.0246	0.0372	0.0274	0.0423	0.0424	0.0278
D05	0.0830	0.0168	0.0772	0.0160	0.1005	0.0371	0.0942	0.0352	0.0920	0.0405	0.0933	0.0934	0.0400
D06	0.1053	0.0168	0.1039	0.0161	0.1320	0.0401	0.1313	0.0393	0.1321	0.0425	0.1305	0.1306	0.0422
D07	0.0979	0.0169	0.1016	0.0161	0.1294	0.0412	0.1326	0.0391	0.1335	0.0437	0.1302	0.1303	0.0428
D08	0.0826	0.0169	0.0883	0.0161	0.1135	0.0403	0.1161	0.0397	0.1177	0.0428	0.1151	0.1152	0.0416
D09	0.0698	0.0169	0.0717	0.0161	0.1031	0.0419	0.1033	0.0406	0.1072	0.0456	0.1025	0.1026	0.0436
D10	0.0951	0.0168	0.0990	0.0161	0.1194	0.0405	0.1227	0.0382	0.1263	0.0450	0.1222	0.1223	0.0430
σ	0.1738		0.1761	0.0060	0.2166		0.2024	0.0254	0.1967	0.0081	0.1910	0.2713	0.0357
λ	1.7675		1.8758	0.1993	1.7735		1.8098	0.1881	1.2488	0.1537	1.2766	1.8816	0.2592
σ_u^2	0.0229		0.0241		0.0356		0.0314		0.0236		0.0226	0.0574	
σ_v^2	0.0073		0.0069		0.0113		0.0096		0.0151		0.0139	0.0162	
$E(u)$	0.1207		0.1240		0.1505		0.1413		0.1225		0.1200	0.1911	

than before, and the coefficient of Feed is smaller. These general results are true for all of the methods we consider that allow for endogeneity. So allowing for endogeneity makes a substantial difference in this application.

Unsurprisingly, the standard errors for C2SLS are substantially larger than for COLS or SF MLE. Our reduced form regressions had values of R^2 that ranged from 0.48 to 0.64, so our instruments are not really "weak", but here as elsewhere allowing for endogeneity carries a price, in terms of variability of the estimates.

We implemented the method of Hansen et al. (2010) using continuously updated GMM. The results are very similar to those for C2SLS and need not be discussed separately, other than to note that the standard errors for the estimator of Hansen, McDonald and Newey are smaller (about 20% smaller) than those for C2SLS, as would have been expected.

The estimates for the LIML estimator of Section 4.3 are quite similar to the C2SLS estimates and those from the method of Hansen, McDonald and Newey. They imply slightly smaller average levels of inefficiency. Interestingly, they do not show any evidence of improved efficiency of the estimated parameters. However, they do provide a good illustration of how we can improve the prediction of u by conditioning on the reduced form errors in addition to the production function error, as discussed in Section 4.4. We have $\hat{E}[\text{var}\,(u|\varepsilon)] = 0.0052$ and $\hat{E}[\text{var}\,(u|\varepsilon, \eta)] = 0.0030$, where these are the sample averages of the theoretically calculated conditional variances. So we can reduce the variance of the prediction of u by almost a factor of two by using the reduced form errors.

The 2-Step LIML estimates are very similar to the one-step estimates just discussed.

Finally, "Copula" in Table 1 refers to the copula-based LIML estimates of Section 4.5. The parameter estimates are not very different from those for the other methods that allow for endogeneity, but we do obtain a noticeably larger estimate of σ_u^2 and correspondingly a larger value of $E(u)$. These results also illustrate the extent to which we can improve our predictions of u by using the reduced form errors when the reduced form errors are correlated with u, as discussed in Section 4.5.2. Now our estimated value of $E[\text{var}\,(u|\varepsilon, \eta)]$ is only 0.0010, compared to 0.0030 from the LIML model of Section 4.3.

6. Concluding remarks

In this paper we have attempted a systematic treatment of endogeneity in stochastic frontier models. The paper is largely a survey but it does contain some new results. We have basically concentrated on the technical details of the extension of the familiar 2SLS and LIML estimators from the standard linear simultaneous equations model to the stochastic frontier model. But there are some non-trivial issues raised by the fact that variables can be correlated with statistical noise, or with technical inefficiency, or both.

The simple COLS estimator for the standard SF model is easy to generalize to the SF model with endogeneity. The usual MLE for the standard SF model is harder to generalize. The LIML estimators that we discuss rely on a reduced form model for the endogenous variables conditional on the instruments, and this model must be correctly specified. Furthermore if the endogenous variables are correlated with inefficiency in addition to (or instead of) with noise, we need to model the joint distribution of inefficiency and the reduced form errors, and this is not simple. However, if we are willing to undertake the LIML approach, we should be able to obtain more precise predictions of the technical inefficiency terms u_i.

We have not discussed reasons why explanatory variables might be endogenous. There could be a long list of such reasons. Here are a few, phrased in the context of agricultural production. (i) The farmer may have some idea of his value of v_i, and this may affect his input choices. For example, if some farms have more favorable climates, differences in weather appear random to the econometrician but are less random to the farmer. (ii) Similarly, the farmer may have some idea of his value of u_i, and this may affect his input choices. For example, a farmer may put more (or less) fertilizer on good soil than on bad soil, or on fields that have been planted properly than on fields that have not. (iii) In some settings liquidity constraints (lack of a perfect credit market) may prevent optimal use of inputs. If a farmer has had good weather in the past, his liquidity constraints may be less binding. If weather is positively autocorrelated, input choices will be correlated with weather in a cross-section of data.

Obviously the list in the previous paragraph could go on and on. What all of these examples have in common is that if we observed more data, like climate and soil quality, and if we had

a better behavioral model of the farmer, we would have less apparent endogeneity. Perhaps there will be a "next generation" of models for which these considerations are relevant. However, in the present state of affairs, technical methods of dealing with endogeneity still seem to us to be useful.

Appendix A

Since u is independent of v and η,

$$f_{u,v,\eta}(u, v, \eta) = f_u(u) \cdot f_{v,\eta}(v, \eta) = f_u(u) f_{v|\eta}(v) f_\eta(\eta). \tag{A.1}$$

So then

$$\begin{aligned} f_{\varepsilon,\eta}(\varepsilon, \eta) &= \int_0^\infty f_{u,v,\eta}(u, \varepsilon + u, \eta)\, du \\ &= f_\eta(\eta) \int_0^\infty f_{v|\eta}(\varepsilon + u) f_u(u)\, du. \end{aligned} \tag{A.2}$$

Now $f_\eta(\eta) = \text{constant} \cdot \left|\Sigma_{\eta\eta}\right|^{-1/2} \cdot \exp(-\frac{1}{2}\eta' \Sigma_{\eta\eta}^{-1}\eta)$. Also the distribution of $v|\eta$ is $N(\mu_c, \sigma_c^2)$ where

$$\mu_c = \Sigma_{v\eta}\Sigma_{\eta\eta}^{-1}\eta, \qquad \sigma_c^2 = \sigma_v^2 - \Sigma_{v\eta}\Sigma_{\eta\eta}^{-1}\Sigma_{\eta v}. \tag{A.3}$$

Therefore $\int_0^\infty f_{v|\eta}(\varepsilon + u) f_u(u)\, du$ is the convolution of $N(\mu_c, \sigma_c^2)$ and $N^+(0, \sigma_u^2)$. If we make a change of variables, $\tilde{\varepsilon} = \varepsilon - \mu_c$, which corresponds to $\tilde{v} = v - \mu_c$, then the distribution of $\tilde{\varepsilon}$ conditional on η is the convolution of $N(0, \sigma_c^2)$ and $N^+(0, \sigma_u^2)$. By the result in Aigner et al. (1977), the density of $\tilde{\varepsilon}|\eta$ is $\frac{2}{\sigma}\varphi\left(\frac{\tilde{\varepsilon}}{\sigma}\right)\Phi\left(\frac{-\lambda\tilde{\varepsilon}}{\sigma}\right)$. Here $\sigma^2 = \sigma_u^2 + \sigma_c^2 = \sigma_u^2 + \sigma_v^2 - \Sigma_{v\eta}\Sigma_{\eta\eta}^{-1}\Sigma_{\eta v}$, $\lambda = \frac{\sigma_u}{\sigma_c} = \frac{\sigma_u}{\sqrt{\sigma_v^2 - \Sigma_{v\eta}\Sigma_{\eta\eta}^{-1}\Sigma_{\eta v}}}$, and φ and Φ are respectively the standard normal density and cdf. Then we simply change the variables back to their original form, so that $\varepsilon = \tilde{\varepsilon} + \mu_c$, and we obtain the desired result:

$$\begin{aligned} f_{\varepsilon,\eta}(\epsilon, \eta) = {} & \text{constant} \cdot \left|\Sigma_{\eta\eta}\right|^{-\frac{1}{2}} \\ & \cdot \exp\left(-\frac{1}{2}\eta' \Sigma_{\eta\eta}^{-1}\eta\right) \cdot \sigma^{-1} \cdot \varphi\left(\frac{\varepsilon - \mu_c}{\sigma}\right) \\ & \cdot \Phi\left(\frac{-\lambda(\varepsilon - \mu_c)}{\sigma}\right). \end{aligned} \tag{A.4}$$

Now we add subscripts i where appropriate; substitute $\eta_i = x_{2i} - \Pi' z_i$, $\varepsilon_i = y_i - \alpha - x_i'\beta$ and $\mu_{ci} = \Sigma_{v\eta}\Sigma_{\eta\eta}^{-1}\eta_i$; and take logs and sum over i; and we arrive at the log likelihood of Eq. (13).

Appendix B

We have $E(u_i|\varepsilon_i, \eta_i) = E(u_i|\tilde{\varepsilon}_i, \eta_i) = E(u_i|\tilde{\varepsilon}_i)$, where the first equality follows from the fact that there is a one-to-one relationship between (ε_i, η_i) and $(\tilde{\varepsilon}_i, \eta_i)$ and the second equality follows from the fact that η_i is independent of u_i and $\tilde{\varepsilon}_i$. Also $\tilde{\varepsilon}_i = \tilde{v}_i - u_i$ is a "composed error" just as in the usual stochastic frontier model (it is the convolution of a normal and a half-normal), so the result of Jondrow et al. (1982) applies. Since $\text{var}(\tilde{v}_i) = \sigma_c^2 = \sigma_v^2 - \Sigma_{v\eta}\Sigma_{\eta\eta}^{-1}\Sigma_{\eta v}$, we simply have to make a few substitutions in the Jondrow et al. formula: their σ_v^2 now becomes σ_c^2; their $\lambda = \frac{\sigma_u}{\sigma_v}$ now becomes $\lambda = \frac{\sigma_u}{\sigma_c}$; and their $\sigma^2 = \sigma_u^2 + \sigma_v^2$ now becomes $\sigma^2 = \sigma_u^2 + \sigma_c^2$. With these changes we can show, by the same algebra as in Jondrow et al., that the distribution of $u_i|\tilde{\varepsilon}_i$ is $N^+(\mu_{*i}, \sigma_*^2)$, where $\mu_{*i} = -\frac{\sigma_u^2}{\sigma^2}\tilde{\varepsilon}_i$ and $\sigma_*^2 = \frac{\sigma_u^2\sigma_c^2}{\sigma^2}$. The expression in (15) then follows from a standard result for the mean of a truncated normal distribution.

References

Aigner, D.J., Lovell, C.A.K., Schmidt, P., 1977. Formulation and estimation of stochastic frontier production function models. J. Econometrics 6, 21–37.

Alvarez, A., Arias, C., 2004. Technical efficiency and farm size: a conditional analysis. Agric. Econom. 30, 241–250.

Amsler, C., Prokhorov, A., Schmidt, P., 2014. Using copulas to model time dependence in stochastic frontier models. Econom. Rev. 33, 497–522.

Atkinson, S.E., Cornwell, C., 1994. Estimation of output and input technical efficiency using a flexible functional form and panel data. Internat. Econom. Rev. 35, 245–255.

Azzalini, A., 2005. The skew-normal distribution and related multivariate families. Scand. J. Stat. 32, 159–188.

Blundell, R., Powell, J., 2004. Endogeneity in semiparametric binary response models. Rev. Econom. Stud. 71, 655–679.

Gourieroux, C., Monfort, A., 1993. Simulation-based inference: A survey with special reference to panel data models. J. Econometrics 59, 5–33.

Greene, W.H., 1980. On the estimation of a flexible frontier production model. J. Econometrics 13, 101–115.

Greene, W.H., 2005. Fixed and random effects in stochastic frontier models. J. Productivity Anal. 23, 7–32.

Guan, Z., Kumbhakar, S.C., Myers, R.J., Lansink, A.O., 2009. Measuring excess capital capacity in agricultural production. Amer. J. Agric. Econom. 91, 765–776.

Hansen, C., McDonald, J.B., Newey, W.K., 2010. Instrumental variables estimation with flexible distributions. J. Bus. Econom. Statist. 28, 13–25.

Horrace, W.C., 2005. Some results on the multivariate truncated normal distribution. J. Multivariate Anal. 94, 209–221.

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

Karakaplan, M.U., Kutlu, L., 2013. Handling endogeneity in stochastic frontier analysis, unpublished manuscript.

Kumbhakar, S.C., 1987. The specification of technical and allocative inefficiency in stochastic production and profit frontiers. J. Econometrics 335–348.

Kumbhakar, S.C., 1997. Modeling allocative inefficiency in a translog cost function and cost share equations: An exact relationship. J. Econometrics 76, 351–356.

Kumbhakar, S.C., Tsionas, E.G., 2005. Measuring technical and allocative inefficiency in the translog cost system: A Bayesian approach. J. Econometrics 126, 355–384.

Kumbhakar, S.C., Tsionas, E.G., 2009. Stochastic error specification in primal and dual production systems. J. Appl. Econometrics 26, 270–297.

Kutlu, L., 2010. Battese–Coelli estimator with endogenous regressors. Econom. Lett. 109, 79–81.

Lewbel, A., 1997. Constructing instruments for regressions with measurement error when no additional instruments are available, with an application to patents and R&D. Econometrica 65, 1201–1213.

Lewbel, A., 2012. Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models. J. Bus. Econom. Statist. 30, 67–81.

Murphy, K.M., Topel, R.H., 2002. Estimation and inference in two-step econometric models. J. Bus. Econom. Statist. 20, 88–97.

Nelsen, R.B., 2006. An Introduction to Copulas. In: Springer Series in Statistics, vol. 139. Springer.

Olson, J.A., Schmidt, P., Waldman, D.M., 1980. A Monte Carlo study of estimators of stochastic frontier production functions. J. Econometrics 13, 67–82.

Park, S., Gupta, S., 2012. Handling endogenous regressors by joint estimation using copulas. Marketing Scie. 31, 567–586.

Schmidt, P., 1976. Econometrics. Marcel Dekker.

Schmidt, P., Lovell, C.A.K., 1979. Estimating technical and allocative inefficiency relative to stochastic production and cost frontiers. J. Econometrics 9, 343–366.

Schmidt, P., Lovell, C.A.K., 1980. Estimating stochastic production and cost frontiers when technical and allocative inefficiency are correlated. J. Econometrics 13, 83–100.

Terza, J.V., Basu, A., Rathouz, P.J., 2008. Two-stage residual inclusion estimation: Addressing endogeneity in health econometric modelling. J. Health Econom. 27, 531–543.

Tran, K.C., Tsionas, E.G., 2013. GMM estimation of stochastic frontier models with endogenous regressors. Econom. Lett. 118, 233–236.

Tsionas, E.G., Atkinson, S.E., Assaf, A.G., 2013. Limited information analysis of multiple output production, unpublished manuscript.

Waldman, D.M., 1982. A stationary point for the stochastic frontier likelihood. J. Econometrics 18, 275–279.

Wooldridge, J.M., 2010. Econometric Analysis of Cross Section and Panel Data. MIT Press.