Economics Letters 147 (2016) 19–22

Contents lists available at ScienceDirect

Economics Letters

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

On the estimation of zero-inefficiency stochastic frontier models with endogenous regressors

Kien C. Tran [a,*], Mike G. Tsionas [b,c]

[a] Department of Economics, University of Lethbridge, 4401 University Drive W, Lethbridge, AB, T1K 3M4, Canada
[b] Department of Economics, Lancaster University Management School, LA1 4YX UK
[c] Athens University of Economics and Business, Greece

HIGHLIGHTS

- We investigate endogeneity issues in the zero-inefficiency stochastic frontier.
- Parameters of the model are estimated using a modified LIML approach.
- Prediction of firm specific inefficiency score is also provided.
- The model is extended to allow for all errors to be potentially correlated.

ARTICLE INFO

Article history:
Received 23 May 2016
Received in revised form
10 August 2016
Accepted 13 August 2016
Available online 17 August 2016

JEL classification:
C3
C5
C13

Keywords:
Endogeneity
Fully efficient firm
Limited information maximum likelihood
Firm specific inefficiency score
Zero-inefficiency stochastic frontier

ABSTRACT

In this paper, we investigate endogeneity issues in the zero-inefficiency stochastic frontier (ZISF) models by mean of simultaneous equation setting. Specifically, we allow for one or more regressors to be correlated with the statistical noise. A modified limited information maximum likelihood (LIML) approach is used to estimate the parameters of the model. Moreover, the firm specific inefficiency score is also provided. Limited Monte Carlo simulations show that the proposed estimators perform well in finite sample.

1. Introduction

The so-called "zero-inefficiency stochastic frontier" (ZISF) model proposed by Kumbhakar et al. (2013) and Rho and Schmidt (2015) allows for some firms in a typical sample to be fully efficient with a certain probability, a fact that we cannot preclude priori. Under standard assumptions on the composed errors, they suggest a maximum likelihood (ML) estimation procedure of the model's parameters as well as how to predict firm specific inefficiency. However, in their models, they assumed that all the regressors (or inputs, in the production frontier setting) are exogenous with respect to the statistical noise and inefficiency. In practice, this assumption might not be valid in some situations and consequently, invalidate the consistency of ML estimator and the estimates of firm specific inefficiency can be misleading. In this paper, we will relax this assumption and allow for one or more regressors to be correlated with the statistical noise in the composed error term; that is, we will investigate the case that one more of the regressors is endogenous, in the sense of simultaneous equation context.

In the standard stochastic frontier setting, the issues of endogeneity have recently been addressed by Amsler et al. (2016a,b), Tran and Tsionas (2013, 2015) and Kutlu (2010). However, to the best of our knowledge, it does not appear that the endogeneity problem has been considered in the ZISF setting. The plan of the paper is as follows. Section 2 introduces the ZISF model with

* Corresponding author. Fax: +1 403 329 2519.
E-mail addresses: kien.tran@uleth.ca (K.C. Tran), tsionas@oternet.gr (M.G. Tsionas).

http://dx.doi.org/10.1016/j.econlet.2016.08.014

endogeneity and discusses various assumptions as well as identification issues. Section 3 derives the limited information maximum likelihood (LIML) procedure as well as firm specific inefficiency predictor. Limited Monte Carlo simulations are presented in Section 4 to examine the finite sample performance of the proposed methods. Section 5 extends the model to allow for one or more inputs to be correlated with both statistical noise and inefficiency. Section 6 concludes the paper.

2. The model

Consider the following Zero-Inefficiency Stochastic Frontier model with endogenous inputs:

$$y_i = \begin{cases} z_{1i}'\alpha + x_i'\beta + v_i & \text{with probability } p(w_i) \\ z_{1i}'\alpha + x_i'\beta + v_i - u_i & \text{with probability } 1 - p(w_i) \end{cases} \quad (1)$$

and

$$x_i = Z_{2i}\delta + e_i, \quad (2)$$

where y_i is a scalar representing output of firm i, z_i is a $q_1 \times 1$ vector of exogenous inputs, x_i is a $d \times 1$ vector of endogenous inputs, v_i is random noise, u_i is one-sided random variable representing technical inefficiency, $p(.)$ is a known function representing the proportion of firms that are fully efficient and w_i is a $k \times 1$ vector of covariates which influence whether a firm is inefficient or not; $Z_{2i} = I_d \otimes z_{2i}'$ where z_{2i} is a $q_2 \times 1$ vector of exogenous instrument variables, and e_i is a $d \times 1$ vector of two-sided random error terms which we assume that $e_i \sim N(0, \Omega_{ee})$ where Ω_{ee} is a $d \times d$ covariance matrix. Following standard practice, we assume $v_i \sim N(0, \sigma_v^2)$, $u_i \sim \left|N(0, \sigma_u^2)\right|$, u_i is independent of $\eta_i = (v_i, e_i)'$ and condition on $Z_i = (z_{1i}, z_{2i})'$, $\eta_i \sim N(0, \Omega)$ where $\Omega = \begin{bmatrix} \sigma_v^2 & \Omega_{ve} \\ \Omega_{ev} & \Omega_{ee} \end{bmatrix}$, so that the endogeneity is due to the correlation between v_i and e_i. For more general case where e_i is allowed to be correlated with both v_i and u_i, see Section 4 below. Also, to ensure that $p(w_i) \in [0, 1]$, we assume $p(w_i)$ takes a logistic function, $p(w_i) = \frac{\exp(w_i'\gamma)}{1+\exp(w_i'\gamma)}$. Finally, for identification purpose, we assume that $\sigma_u^2 > 0$ and $q_2 \geq d$ (so that there are at least as many instruments as x's).

3. LIML procedure

To obtain the likelihood function, we follow density decomposition of Amsler et al. (2016a) approach, albeit one can also use Cholesky's decomposition approach as suggested in Kutlu (2010) and Tran and Tsionas (2013). Let $\varepsilon_i = v_i - u_i = y_i - z_{1i}'\alpha - x_i'\beta$, $\tilde{\varepsilon}_i = \varepsilon_i - \mu_{ci}$ where $\mu_{ci} = \Omega_{ve}\Omega_{ee}^{-1}e_i$ with $e_i = x_i - Z_i\delta$ and $\sigma_c^2 = \sigma_v^2 - \Omega_{ve}\Omega_{ee}^{-1}\Omega_{ev}$. Next, since u_i is independent of v_i and e_i, we have:

$$\begin{aligned} f_{u,v,e}(u, v, e) &= p(w)f_{v,e}(v, e) + \{1 - p(w)\}f_u(u)f_{v,e}(v, e) \\ &= p(w)f_{v|e}(v)f_e(e) + \{1 - p(w)\}f_u(u)f_{v|e}(v)f_e(e) \\ &= f_e(e)\{p(w)f_{v|e}(v) + (1 - p(w))f_u(u)f_{v|e}(v)\}, \end{aligned} \quad (3)$$

where $f_e(e) = const \times |\Omega_{ee}|^{-1} \times \exp(-\frac{1}{2}e'\Omega_{ee}^{-1}e)$ and the distribution of $v \mid e$ is $N(\mu_c, \sigma_c^2)$. Consequently,

$$f_{\varepsilon,e}(\varepsilon, e) = \int_0^\infty f_{u,v,e}(u, \varepsilon + u, e)du = f_e(e)\int_0^\infty f_{v|e}(\varepsilon + u)du. \quad (4)$$

By making use of change in variables, $\tilde{\varepsilon} = \varepsilon - \mu_c$ and using the result in Aigner et al. (1977), the density of $\tilde{\varepsilon} \mid e$ can be shown to be

$$f_{\tilde{\varepsilon}|e}(\tilde{\varepsilon}) = \frac{2}{\sigma}\phi\left(\frac{\tilde{\varepsilon}}{\sigma}\right)\Phi\left(\frac{-\lambda\tilde{\varepsilon}}{\sigma}\right), \quad (5)$$

where $\sigma^2 = \sigma_u^2 + \sigma_c^2 = \sigma_u^2 + \sigma_v^2 - \Omega_{ve}\Omega_{ee}^{-1}\Omega_{ev}$, $\lambda = \frac{\sigma_u}{\sigma_c}$ and $\phi(.)$ and $\Phi(.)$ are respectively the standard normal density and cdf. Thus, by writing $\varepsilon = \tilde{\varepsilon} + \mu_c$, we obtain:

$$\begin{aligned} f_{\varepsilon,e}(\varepsilon, e) = &\left\{p(w)(2\pi\sigma_c^2)^{-1/2}\exp\left(-\frac{1}{2\sigma_c^2}(v - \mu_c)^2\right)\right. \\ &\left. + (1 - p(w))\sigma^{-1}\phi\left(\frac{\varepsilon - \mu_c}{\sigma_c}\right)\Phi\left(\frac{-\lambda(\varepsilon - \mu_c)}{\sigma_c}\right)\right\} \\ &\times (2\pi)^{-1}\Omega_{ee}^{-1/2}\exp\left(-\frac{1}{2}e'\Omega_{ee}^{-1}e\right). \end{aligned} \quad (6)$$

Then the log-likelihood function is:

$$\ln L = \ln L_1 + \ln L_2 \quad (7a)$$

where

$$\begin{aligned} \ln L_1 = &\sum_{i=1}^n \ln\left\{p(w_i)(2\pi\sigma_c^2)^{-1/2}\right. \\ &\times \exp\left(-\frac{1}{2\sigma_c^2}(y_i - z_{1i}'\alpha - x_i'\beta - \mu_{ic})^2\right) \\ &+ [1 - p(w_i)]\sigma^{-1}\phi\left(\frac{y_i - z_{1i}'\alpha - x_i'\beta - \mu_{ic}}{\sigma_c}\right) \\ &\left.\times \Phi\left(\frac{-\lambda(y_i - z_{1i}'\alpha - x_i'\beta - \mu_{ic})}{\sigma_c}\right)\right\} \end{aligned} \quad (7b)$$

and

$$\ln L_2 = -\frac{n}{2}\ln|\Omega_{ee}| - \frac{1}{2}\sum_{i=1}^n(x_i - Z_{2i}\delta)'\Omega_{ee}^{-1}(x_i - Z_{2i}\delta). \quad (7c)$$

Note that, in a special case where $p(.) = 0$, (7a) reduces to the log-likelihood function of the standard SF models with endogenous regressors (e.g., Kutlu, 2010; Tran and Tsionas, 2013; and Amsler et al., 2016a). On the other hand, when $p(.) = 1$, it reduces to the log-likelihood function of simultaneous regression models. Finally, when there are no endogenous regressors, (i.e., $\Omega_{ve} = 0$) it reduces to the case of ZISF models of Kumbhakar et al. (2013) and Rho and Schmidt (2015).

Now by maximizing the log-likelihood function in (7a) directly with respect to the parameters $\theta = (\alpha, \beta, \sigma_v^2, \sigma_u^2, \delta, \Omega_{ve}, \Omega_{ee}, \gamma)$, we can obtain the LIML estimates. Or alternatively, we can use Generalized Method of Moment (GMM) approach of Tran and Tsionas (2013) which uses the moment conditions that are based on the score of the log-likelihood function. This GMM procedure is similar to the direct MLE. Finally, a control function type two-step procedure suggested by Kutlu (2010) can also be used. To construct a two-step procedure, let $\theta_1 = (\alpha, \beta, \sigma_v^2, \sigma_u^2, \Omega_{ve}, \gamma)$ and $\theta_2 = (\delta, \Omega_{ee})$. Then in the first step, we maximize (7c) with respect to θ_2, and this is essentially the least square estimation of x on Z from the reduced form equations to obtain $\hat{\delta}$ and $\hat{\Omega}_{ee} = n^{-1}\sum_{i=1}^n(X_i - Z_i'\hat{\delta})(X_i - Z_i'\hat{\delta})'$. In the second step, given the estimates of $\hat{\theta}_2$, we maximize (7b) to obtain the remaining parameters θ_1. Note that, unlike the direct MLE or GMM procedure, this two-step procedure is generally inefficient because it ignores the information about θ_2 in (7b) and treating as though it is known. Consequently, a practical implication is that the conventional estimated standard errors from step 2 are not correct, and they need to be adjusted to reflect the fact that θ_2 have been estimated. One simple way to mitigate this problem is to use bootstrapping procedure. Or alternatively, one could follow Wooldridge (2010, Section 12.4.2) to construct the correct standard errors analytically.

Prediction of firm specific inefficiency:

Once the parameters of the model have been estimated, we can use (Jondrow et al., 1982) procedure to construct the estimate for