University of Texas Rio Grande Valley

ScholarWorks @ UTRGV

Economics and Finance Faculty Publications and Presentations

Robert C. Vackar College of Business & Entrepreneurship

2-2018

A distribution-free stochastic frontier model with endogenous regressors

Levent Kutlu
The University of Texas Rio Grande Valley, levent.kutlu@utrgv.edu

Follow this and additional works at: https://scholarworks.utrgv.edu/ef_fac

Part of the Finance Commons

Recommended Citation

Kutlu, Levent. "A Distribution-Free Stochastic Frontier Model with Endogenous Regressors." Economics Letters 163 (February 1, 2018): 152–54. https://doi.org/10.1016/j.econlet.2017.12.026.

This Article is brought to you for free and open access by the Robert C. Vackar College of Business & Entrepreneurship at ScholarWorks @ UTRGV. It has been accepted for inclusion in Economics and Finance Faculty Publications and Presentations by an authorized administrator of ScholarWorks @ UTRGV. For more information, please contact justin.white@utrgv.edu, william.flores01@utrgv.edu.

A Distribution-Free Stochastic Frontier Model with Endogenous Regressors

Levent Kutlu
School of Economics
Georgia Institute of Technology

December 11, 2017

Abstract

We provide a guideline for estimating a distribution-free panel data stochastic frontier model in the presence of endogenous variables. In particular, we consider variations of the within estimator of Cornwell et al. (1990) to allow endogenous regressors.

Keywords: Endogeneity; time-varying efficiency; panel data; stochastic frontier

JEL classification numbers: C13, C23

1 Introduction

The stochastic frontier literature started with the cross-sectional works of Meeusen and van den Broeck (1977) and Aigner et al. (1977). Pitt and Lee (1981) and Schmidt and Sickles (1984) provided panel data models with time-invariant inefficiency. Cornwell et al. (1990), Kumbhakar (1990), and Battese and Coelli (1992) exemplify earlier panel data models that relaxed time-invariance assumption.[1] Starting with Guan et al. (2009) and Kutlu (2010), there is a recent trend in the stochastic frontier literature that aims to handle endogeneity issues. Both Guan et al. (2009) and Kutlu (2010) present models that allow regressors in the frontier to be correlated with the two-sided error term. Among others, Karakaplan and Kutlu (2017a,b) and Amsler et al. (2017) further developed models where the environmental variables can also be correlated with the two-sided error term. However, many of the studies that solve endogeneity problems rely on distributional assumptions and are considerably harder to estimate compared to the within estimator of Cornwell et al. (1990) (CSSW).[2] Hence, it is our interest to extend the CSSW estimator to allow endogeneity. However, the

[1] See Duygun et al. (2016) for a dynamic counterpart of Cornwell et al. (1990).

[2] For more details about cons and pros of CSSW estimator, see Kutlu (2017) and Kutlu et al. (2017). Also, see Adams and Sickles (2007) for a semi-parametric distribution-free estimator.

purpose of this study is beyond such an extension of this well-known estimator. In particular, we aim to provide a guideline for empirical researchers about how endogeneity issues can be solved in a distribution-free stochastic frontier framework. For this purpose, we provide solutions where the parameters and inefficiency can be estimated consistently when frontier or environmental variables are correlated with the two-sided error term.

2 Distribution-Free Estimators and Endogeneity

Consider a panel of N productive units observed over T_i periods for panel unit i. For the sake of fixing ideas, we consider stochastic frontier production function estimation. A commonly used stochastic frontier model for production function is given by:

$$y_{it} = \alpha + x'_{it}\beta - u_{it} + v_{it} \tag{1}$$

where y_{it} is the logarithm of the output and x_{it} is a vector of frontier variables, $u_{it} \geq 0$ is the panel unit effects representing technical inefficiency, v_{it} is the usual two-sided error term, and α and β are parameters.

In the panel data context, many researchers assume that $v_{it} \sim N\left(0, \sigma_v^2\right)$ and $u_{it} = h\left(w'_{it}\gamma\right) u_i^*$ where $h > 0$ is a function, w_{it} is a vector of environmental variables and constant that effect technical inefficiency, and $u_i^* \geq 0$ is drawn from a one-sided distribution such as half-normal, exponential, truncated normal, and gamma distribution. The conventional assumption of these models is that u_i^*, v_{it}, and (x'_{it}, w'_{it}) are independent of each other. Guan et al. (2009) and Kutlu (2010) relax the independence assumption of x_{it} and v_{it}. Karakaplan and Kutlu (2017b) relax the independence assumption of (x'_{it}, w'_{it}) and v_{it}.

Unlike these models, Cornwell et al. (1990) consider a distribution-free stochastic frontier model:

$$y_{it} = \alpha_{it} + x'_{it}\beta + v_{it} \tag{2}$$

where $\alpha_{it} = \alpha - u_{it}$. Cornwell et al. (1990) assume that $\alpha_{it} = w'_{it}\delta_i$ where w_{it} is a vector of environmental variables and constant that determine inefficiency and δ_i is a panel unit specific parameter vector.[3] The model becomes:

$$y_{it} = w'_{it}\delta_i + x'_{it}\beta + v_{it}. \tag{3}$$

In matrix notation, the model is:

$$y = w\delta + x\beta + v \tag{4}$$

[3] Schmidt and Sickles (1984) and Cornwell et al. (1990) use $w_{it} = 1$ and $w_{it} = \left(1, t, t^2\right)'$, respectively.

where $w = I_N \otimes w_{i.}$ is a block-diagonal matrix and $w_{i.}$ is a matrix with rows w'_{it}.

We denote the projection matrix onto the column space of a matrix A by $P_A = A\left(A'A\right)^{-1} A'$ and the projection matrix onto the null space of A by $M_A = I - P_A$. Hence, the subscripts refer to the matrix on which the projections are made. Using a model transformation by M_w, Cornwell et al. (1990) eliminate the $w\delta$ term and obtain:

$$\tilde{y} = \tilde{x}\beta + \tilde{v} \tag{5}$$

where $\tilde{y} = M_w y$, $\tilde{x} = M_w x$, and $\tilde{v} = M_w v$. The CSSW estimator of β is given by:

$$\hat{\beta} = \left(\tilde{x}'\tilde{x}\right)^{-1} \tilde{x}'\tilde{y}. \tag{6}$$

Then, δ_i can be estimated by regressing residuals, $y_{it} - x'_{it}\hat{\beta}$, for panel unit i on w_{it}. The fitted values from this regression gives an estimate of α_{it} that is consistent as $T_i \to \infty$. The frontier intercept at time t, α_t, and the panel unit-specific level of inefficiency, u_{it}, for panel unit i at time t are estimated, respectively, as:

$$\begin{aligned} \hat{\alpha}_t &= \max_j \left\{\hat{\alpha}_{jt}\right\} \\ \hat{u}_{it} &= \hat{\alpha}_t - \hat{\alpha}_{it}. \end{aligned} \tag{7}$$

This model allows the inefficiency to be correlated with frontier variables. Unlike the models that we mentioned above, this model does not need to worry about the u_i^* term and its correlation with the regressors, as it is not present in the model. However, when the two-sided error term is correlated with the frontier or environmental variables, the CSSW estimator, $\hat{\beta}$, would be inconsistent. Below, we discuss endogeneity problems and their solutions in this framework.

For now, we assume that w is independent of v but x has endogenous variables. In this case, the CSSW estimator, $\hat{\beta}$, would be inconsistent. However, the following instrumental variables estimator (CSSWIV) of β would be consistent:

$$\hat{\beta}_{IV} = \left(\tilde{x}'P_{\tilde{z}}\tilde{x}\right)^{-1} \tilde{x}'P_{\tilde{z}}\tilde{y} \tag{8}$$

where $\tilde{z}$ is a vector of instrumental variables for $\tilde{x}$ so that $E\left[\tilde{v} \mid \tilde{z}\right] = 0$. In our case, a particular choice for $\tilde{z}$ would be so that $\tilde{z} = M_w z$ where z satisfies $E\left[v \mid z\right] = 0$. For this estimator, since M_w annihilates w, we do not include w in z. The consistency of $\hat{\beta}_{IV}$ follows as $E\left[v \mid z\right] = 0$ and independence of w and v implies that $E\left[\tilde{v} \mid \tilde{z}\right] = 0$. Basically, we estimate Equation (5) by the two-stage least squares (2SLS) using the transformed instruments, i.e., $\tilde{z}$. As earlier, δ_i can be estimated by regressing residuals, $y_{it} - x'_{it}\hat{\beta}_{IV}$, for panel unit i on w_{it} and the inefficiency is estimated by Equation (7).

When both w and x have endogenous variables, a solution is estimating Equation (4) by 2SLS without transformation (WTIV) using z as instruments for w and x so that $E\left[v \mid z\right] = 0$. Here, z includes instruments q that are specifically designed for w, i.e., $q = I_N \otimes q_{i.}$ and $q_{i.}$ is a matrix with rows q'_{it}.

An alternative solution would be handling the endogeneity issue in two stages as it is done by Cornwell et al. (1990). Therefore, we consider the following transformation of the original model:

$$\bar{y} = \bar{x}\beta + \bar{v} \tag{9}$$

where $\bar{y} = P_z M_w y$, $\bar{x} = P_z M_w x$, $\bar{v} = P_z M_w v$, and z is the matrix of instrumental variables for w and x so that $E[v \mid z] = 0$ and $P_z M_w$ is independent from v. The following instrumental variables estimator (CSSWIV2) of β would be consistent:

$$\begin{aligned} \hat{\beta}_{IV2} &= \left(\bar{x}' \, P_{\bar{z}} \bar{x} \right)^{-1} \, \bar{x}' P_{\bar{z}} \, \bar{y} \\ &= \left(\tilde{x}' P_z \, \tilde{x} \right)^{-1} \, \tilde{x}' P_z \tilde{y} \end{aligned} \tag{10}$$

where $\bar{z} = P_z M_w z$. To see this, note that when $P_z M_w$ is independent of v and $E[v \mid z] = 0$, we have $E[\bar{v} \mid \bar{z}] = 0$. Then, δ_i can be estimated by regressing residuals, $y_{it} - x_{it}'\hat{\beta}_{IV2}$, for panel unit i on w_{it} using the 2SLS method with z_{it} being the instruments.

We note that when x is endogenous and w is exogenous, we have $\hat{\beta}_{IV} = \hat{\beta}_{IV2}$. However, the estimates differ when w is endogenous. Moreover, when x is exogenous and w is endogenous, we have $\hat{\beta} = \hat{\beta}_{IV}$ although the efficiency estimates differ.

3 Monte Carlo Experiments

We conduct the Monte Carlo experiments with $1,000$ replications for two different scenarios. The estimators that we consider are: CSSW, CSSWIV, CSSWIV2, and WTIV. For each scenario, we assume:

$$y_{it} = w_{it}\delta_i + x_{1it}\beta_1 + x_{2it}\beta_2 + v_{it}.$$

We summarize the data generating processes for Monte Carlo experiments below:

Scenario 1 (Endogenous x_2): $(x_{1it}, d_{it}, w_{it})' \sim \mathbf{N}(\mu, \Sigma)$, $x_{2it} = d_{it} + e_{it}$, $(e_{it}, v_{it})' \sim \mathbf{N}(0, \Omega)$, $\delta_i \sim \mathbf{N}\left(0, h^2\right)$, and $(\beta_1, \beta_2) = (0.5, 0.5)$.

Scenario 2 (Endogenous w): $(x_{1it}, x_{2it}, d_{it})' \sim N(\mu, \Sigma)$, $w_{it} = d_{it} + e_{it}$, $(e_{it}, v_{it})' \sim \mathbf{N}(0, \Omega)$, $\delta_i \sim \mathbf{N}\left(0, h^2\right)$, and $(\beta_1, \beta_2) = (0.5, 0.5)$.

We assume that $\mu = (1, 1, 0)'$, $\Sigma = \begin{bmatrix} 1 & 0.4 & 0.8 \\ 0.4 & 1 & 0.8 \\ 0.8 & 0.8 & 1 \end{bmatrix}$, and $\Omega = \begin{bmatrix} 0.25 & 0.2 \\ 0.2 & 0.25 \end{bmatrix}$ so that correlation of e_{it} and v_{it} equals 0.8.

We present the results of Monte Carlo experiments in Table 1 and Table 2.[4] As we mentioned earlier, when w is exogenous it does not matter which one of

[4] We also considered a scenario where all variables are exogenous. All estimators performed well in this scenario.

the instrumental variables estimators that we use. However, the results change drastically when w is endogenous. CSSWIV2 and WTIV perform similarly in terms of estimating efficiency. However, WTIV outperforms CSSWIV2 in terms of estimating β parameters. Finally, as expected, the estimators performs better when the sample size increases.

Table 1-2 here

4 References

Adams, R.M. and Sickles, R.C. (2007), Semiparametric Efficient Distribution Free Estimation of Panel Models, *Communications in Statistics-Theory and Methods*, 36, 1-18.

Aigner, D.J., Lovell, C.A.K., and Schmidt, P. (1977), Formulation and Estimation of Stochastic Frontier Production Functions, *Journal of Econometrics*, 6, 21-37.

Amsler, C., Prokhorov, A., and Schmidt, P. (2017), Endogenous Environmental Variables in Stochastic Frontier Models, *Journal of Econometrics*, 199, 131-140.

Battese, G.E. and Coelli, T.J. (1992), Frontier Production Functions, Technical Efficiency and Panel Data with Application to Paddy Farmers in India, *Journal of Productivity Analysis*, 3, 153-169.

Cornwell, C., Schmidt, P., and Sickles, R.C. (1990), Production Frontiers with Time-Series Variation in Efficiency Levels, *Journal of Econometrics*, 46, 185-200.

Duygun, M., Kutlu, L., and Sickles, R.C. (2016), Measuring Productivity and Efficiency: A Kalman Filter Approach, *Journal of Productivity Analysis*, 46, 155-167.

Guan Z., Kumbhakar S.C., Myers R.J., and Lansink A.O. (2009), Measuring Excess Capital Capacity in Agricultural Production, *American Journal of Agricultural Economics*, 91, 765-776.

Karakaplan, M.U. and Kutlu, L. (2017a), Handling Endogeneity in Stochastic Frontier Analysis, *Economics Bulletin*, 37, 889-901.

Karakaplan, M.U. and Kutlu, L. (2017b), Endogeneity in Panel Stochastic Frontier Models: An Application to the Japanese Cotton Spinning Industry, *Applied Economics*, 49, 5935-5939.

Kumbhakar, S.C. (1990), Production Frontiers, Panel Data, and Time-Varying Technical Inefficiency, *Journal of Econometrics*, 46, 201-211.

Kutlu, L. (2010), Battese-Coelli Estimator with Endogenous Regressors, *Economics Letters*, 109, 79-81.

Kutlu, L. (2017), A Constrained State Space Approach for Estimating Firm Efficiency, *Economics Letters*, 152, 54-56.

Kutlu, L., Tran, K., and Tsionas, E.G. (2017), A Time-Varying True Individual Effects Model with Endogenous Regressors, Working paper (Available at SSRN: https://ssrn.com/abstract=2721864)

Meeusen, W. and van den Broeck, J. (1977), Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error, *International Economic Review*, 2, 435-444.

Pitt, M.M. and Lee, L.F. (1981), The Measurement and Sources of Technical Inefficiency in the Indonesian Weaving Industry, *Journal of Development Economics*, 9, 43-64.

Schmidt, P. and Sickles, R.C. (1984), Production Frontiers and Panel Data, *Journal of Business and Economic Statistics*, 2, 367-374.