

Stochastic Error Function I – Another Composed Error: Lecture X

Charles B. Moss

¹University of Florida

September 18, 2018

- 1 Concept of the Composed Error
- 2 Models Without Random Shocks
 - Corrected Ordinary Least Squares
 - Modified Ordinary Least Squares
- 3 Stochastic Frontier Specifications

Concept of the Composed Error

- To introduce the composed error term, we will begin with a cursory discussion of technical efficiency which we develop more fully after the dual.
- We start with the standard production function

$$y_i = f(x_i, \beta) \quad (1)$$

- We begin by acknowledging that firms may not produce on the efficient frontier

$$y_i = f(x_i, \beta) TE_i \quad (2)$$

- We assume that $TE_i \leq 1$ with $TE_i = 1$ denoting a technically efficient producer

$$TE_i = \frac{y_i}{f(x_i, \beta)} \quad (3)$$

- The above model presents all the error between the firm's output and the frontier as technical inefficiency.
- The above model presents all the error between the firm's output and the frontier as technical inefficiency.
- Augmenting this model with the possibility that random shocks may affect output that do not represent inefficiency

$$y_i = f(x_i, \beta) \exp(\nu_i) TE_i \quad (4)$$

Models Without Random Shocks

- Building on the model of technical inefficiency alone, we could estimate the production function using a one-sided error specification alone.
- Mathematical Programming (Goal Programming): First we could solve two non-linear programming problems:
- First we could minimize the sum of the residuals such that we constrain the residuals to be positive:

$$\begin{aligned} & \text{Min } \sum_i u_i \\ & \text{s.t. } u_i = \beta_0 + \sum_k \beta_k \ln(x_{ik}) - \ln(y_i) \\ & u_i \geq 0 \quad i = 1, \dots, I \end{aligned} \tag{5}$$

- which approximates the distribution function for the exponential distribution with a log likelihood function

$$\ln(L) = I \ln(\sigma_u) - \frac{1}{\sigma_u} \sum_i |u_i| \quad (6)$$

- The second specification minimizes the sum of square residuals such that the residual is constrained to be positive

$$\text{Min} \sum_i u_i^2$$

$$\text{s.t. } u_i = \beta_0 + \sum_k \beta_k \ln(x_{ik}) - \ln(y_i) \quad (7)$$

$$u_i \geq 0 \quad i = 1, \dots, I$$

- Which approximate the half-normal distribution

$$\ln(L) = C - \frac{1}{2} \ln(\sigma_u^2) - \frac{1}{2\sigma_u^2} \sum_i u_i^2 \quad (8)$$

Corrected Ordinary Least Squares

- Estimate the production function using ordinary least squares, then adjust the estimated frontier by adding a sufficient constant to the estimated intercept to make all the error terms negative

$$\hat{\beta}_0^* = \hat{\beta}_0 + \max_i \{ \hat{u}_i \} \quad (9)$$

the estimated residuals are then

$$\hat{u}_i^* = \hat{u}_i - \max_i \{ \hat{u}_i \} \quad (10)$$

- This procedure simply shifts the production function estimated with OLS upward, no information on the inefficiency is used in the estimation of the slope coefficients.

Modified Ordinary Least Squares

- A related two step estimation procedure it to again estimate the constant and slope parameters using ordinary least squares, and then to fit a secondary distribution function (i.e., the half-normal, gamma, or exponential) to the residuals.
- The expected value of the residuals for this second distribution is then used to adjust the constant of the regression and the residuals:

$$\begin{aligned}\hat{\beta}_0^{**} &= \hat{\beta}_0 + E(\hat{u}_i) \\ \hat{u}_i^{**} &= \hat{u}_i - E(\hat{u}_i)\end{aligned}\tag{11}$$

- In addition to the constant shift in the production function addressed above, this specification does not necessarily guarantee that all the residuals will be negative.

Stochastic Frontier Specifications

- Adding both technical variation and stochastic effects to the production model, we get

$$\ln(y_i) = \beta_0 + \sum_k \beta_k \ln(x_{ik}) + \nu_i - u_i \quad (12)$$

- The overall error term of the regression is referred to as the composed error

$$\ln(y_i) = \beta_0 + \sum_k \beta_k \ln(x_{ik}) + \epsilon_i \quad (13)$$
$$\epsilon_i = \nu_i - u_i$$

- Assuming that the components of the random error term are independent, OLS provides consistent estimates of the slope coefficients, but not of the constant.

- Further, OLS does not provide estimates of producer-specific technical inefficiency.
- However, OLS does provide a test for the possible presence of technical inefficiency in the data
 - Specifically, if technical inefficiency is present then $u_i < 0$ so that the distribution is negatively skewed.
- Various tests for significance of skewness are available, but I typically apply the Bera and Jarque tests

$$(b_1)^{1/2} = \frac{m_3}{(m_2)^{3/2}}$$
$$\frac{m_3}{\left(6 \frac{m_2^3}{N}\right)^{1/2}} \sim N(0, 1) \quad (14)$$

Variable	Parameter
Constant	4.58582 (0.05607)
Nitrogen	0.01265 (0.01179)
Phosphorous	0.01677 (0.00732)
Potash	0.01322 (0.00629)

$$(b_1)^{1/2} = \frac{m_3}{(m_2)^{3/2}} = -2.2140057$$

$$\frac{m_3}{\left(6 \frac{m_2^3}{N}\right)^{1/2}} = -15.785307 \sim N(0, 1) \quad (15)$$