



AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
<http://ageconsearch.umn.edu>
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

Estimation of Efficiency with the Stochastic Frontier Cost Function and Heteroscedasticity: A Monte Carlo Study

By

Taeyoon Kim
Graduate Student
Oklahoma State University
Department of Agricultural Economics

Dr. Wade Brorsen
Regents Professor
Oklahoma State University
Department of Agricultural Economics

Dr. Philip Kenkel
Professor
Oklahoma State University
Department of Agricultural Economics

**Selected Paper prepared for presentation at the American Agricultural Economics
Association Annual Meeting, Orlando, Fl, July 27-29, 2008.**

Copyright 2008 by Taeyoon Kim, Dr. Wade Brorsen, and Dr. Philip Kenkel. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.

Estimation of Efficiency with the Stochastic Frontier Cost Function and Heteroscedasticity: A Monte Carlo Study

Abstract

The objective of this article is to address heteroscedasticity in the stochastic frontier cost function using aggregated data and verify it using a Monte Carlo study. We find that when the translog form of a stochastic frontier cost function with aggregated data is estimated, all explanatory variables can inversely affect the variation of error terms. Our Monte Carlo study shows that heteroscedasticity is only significant in the random effect and the unexplained error term not in the inefficiency error term. Also, it does not cause biases, which is quite opposite of previous research. These are because our model is approximately defined by first order Taylor series around zero inefficiency area. But, disregarding heteroscedasticity causes the average inefficiency to be overestimated when the variation of inefficiency term dominates the other error terms.

Estimation of Efficiency with the Stochastic Frontier Cost Function and Heteroscedasticity: A Monte Carlo Study

Introduction

Since the advent of Farrell (1957) efficiency indexes using a deterministic frontier function, efficiency measurements have been consistently developed by researchers over all industry. Aigner, Lovell, and Schmidt (1977) brought about the possibility that deviations from the frontier may arise because of random factors and provided the disturbance term as the sum of symmetric normal and half-normal random variables. Meeusen and van den Broeck (1977) also introduced the composed error which distinguishes inefficiency from a statistical disturbance of randomness. Jondrow et al. (1982) suggested estimating inefficiency component for each observation with the stochastic frontier model so that rankings among observations were possible.

In terms of addressing heteroscedasticity, Caudill and Ford (1993) found the biases in the frontier estimation due to heteroscedasticity of a one-sided error and later Caudill, Ford, and Gropper (1995) found that the rankings of firms by efficiency measures were significantly affected by the correction for heteroscedasticity. These were followed by the suggestion from Schmidt (1986) that a one-sided error can be associated with factors

under the control of the firm while the random component can be associated with factors outside the control of the firm. Since firm-level data are used in the frontier function and firms vary widely in size, size-related heteroscedasticity is involved in a one-sided error. On the contrary, concerned with heteroscedasticity only in a one-sided error, Hadri (1999) suggested heteroscedasticity of both error terms with the same data of Caudill, Ford, and Gropper (1995).

On the other hand, Dickens (1990) showed that aggregated data caused heteroscedasticity with the size of group. This result can lead that small firms are less efficient while large firms are more efficient when the frontier (average) cost function with aggregated data is estimated because the variation is decreasing as the size of group increases. This can also induce the discussion that economies of size is also affected by heteroscedasticity.

A translog cost function and maximum likelihood estimation (MLE) are widely used in previous research. Since the joint distribution of a symmetric normal and a truncated normal had been first derived by Weinstein (1964), Aigner et al. (1977) and Greene (1980) described a method of estimating the frontier production model using a translog functional form and MLE.

This paper is about addressing heteroscedasticity using the translog form of a stochastic frontier cost function with aggregated data and examining average efficiency measurement for both cases. We specify the aggregated model from the disaggregated model and take a natural log in order to use a translog cost function. To make the equation be simplified, first order Taylor series around the frontier area is applied. Then, we can see heteroscedasticity on error terms. A Monte Carlo study enables to verify it and compare efficiency measurement in the presence of heteroskedasticity with that of disregarding heteroskedasticity.

Theory

Consider the following disaggregated model:

$$(1) \quad C_{ij} = (\mathbf{X}\boldsymbol{\beta})_{ij} + u_j + w_{ij}, \quad i = 1, \dots, n_j, \quad j = 1, \dots, J,$$

$$u_j \sim iid N(0, \sigma_u^2), \quad w_{ij} \sim iid N(0, \sigma_w^2), \quad \text{cov}(u_j, w_{ij}) = 0,$$

where C_{ij} is the cost of the i th output in the j th firm, \mathbf{X} is a vector of explanatory variables including input prices and output, $\boldsymbol{\beta}$ is a vector of unknown parameters to be estimated, u_j is the random effect of the j th firm, w_{ij} is the unexplained portion of the cost of i th output in the j th firm.

In a stochastic frontier cost function, the inefficiency is considered as the deviations from the frontier so that a one-sided error term is needed to represent that. Thus a stochastic frontier cost function can be defined as

$$(2) \quad C_{ij} = (\mathbf{X}\boldsymbol{\beta})_{ij} + u_j + w_{ij} + v_j, \quad v_j \sim iid \left| N(0, \sigma_v^2) \right|, \quad \text{cov}(u_j, v_j) = 0, \\ \text{cov}(w_{ij}, v_j) = 0,$$

where v_j is the inefficiency and a one-sided error with $E(v) = \sigma_v \sqrt{2/\pi}$ and $\text{var}(v) = \sigma_v^2 (1 - 2/\pi)$. Especially, $\sigma_v \sqrt{2/\pi}$ is known as average inefficiency measurement by Aigner et al. (1977).

When being added over all outputs within each firm, a (total) stochastic frontier cost function can be derived as

$$(3) \quad \sum_{i=1}^{n_j} C_{ij} = \sum_{i=1}^{n_j} (\mathbf{X}\boldsymbol{\beta})_{ij} + n_j u_j + \sum_{i=1}^{n_j} w_{ij} + n_j v_j.$$

where n_j is the number of output in the j th firm.

Assuming cost functions for each firm have same explanatory variables, a (total) stochastic frontier cost function can be also expressed as

$$(4) \quad TC_j = n_j (\mathbf{X}\boldsymbol{\beta})_{.j} + n_j (u_j + w_{.j} + v_j), \quad j = 1, \dots, J, \quad w_{.j} \sim N(0, \frac{\sigma_w^2}{n_j}),$$

where TC_j is the total cost for the j th firm, and the dot is the common notation to denote that the variable has been averaged over the corresponding index; outputs in this

case. Here, we can see that aggregated data cause heteroscedasticity with outputs on error terms when the (total) stochastic frontier cost function is estimated.

A translog cost function can be usually used due to several conveniences such as including multiple outputs, calculating elasticities easily, adjusting heteroscedasticity and etc. Let's take a natural log for the equation (4) and first order Talyor series of $\ln((\mathbf{X}\boldsymbol{\beta})_{.j} + u_j + w_j + v_j)$ around the mean of random errors and the frontier of inefficiency error such as $u_j = 0$, $w_j = 0$ and $v_j = 0$ gives us following model:

$$(5) \quad \ln TC_j \approx \ln(\mathbf{X}\boldsymbol{\beta})_{.j} + \ln n_j + \frac{1}{(\mathbf{X}\boldsymbol{\beta})_{.j}} (u_j + w_j + v_j).$$

The variance of error terms is $\left(\frac{1}{(\mathbf{X}\boldsymbol{\beta})_{.j}} \right)^2 \left(\sigma_u^2 + \frac{\sigma_w^2}{n_j} + \sigma_v^2 \right)$, which means that

explanatory variables also affect heteroscedasticity on error terms. In other words, the variance of all error terms is inversely affected by all squares of explanatory terms.

When letting $e = w + v$, the density function by Weinstein(1964) is known as

$$(6) \quad f(e) = \frac{2}{\sigma} f^* \left(\frac{e}{\sigma} \right) F^* \left(\frac{\lambda e}{\sigma} \right), -\infty < e < +\infty$$

where $\sigma^2 = \sigma_w^2 + \sigma_v^2$, $\lambda = \sigma_v / \sigma_w$, f^* and F^* are the standard probability density function and the standard cumulative density function, respectively. Here, λ is an indicator of the relative variability of error terms. As Aigner et al. (1977) mentioned it, $\lambda^2 \rightarrow 0$ means $\sigma_v^2 \rightarrow 0$ and/or $\sigma_w^2 \rightarrow \infty$, i.e. that inefficiency error is dominated by random error.

The log-likelihood function with heteroscedasticity in equation (5) can be expressed as

$$(7) \quad \sum_{j=1}^J \ln(f_j(e_j)) = \sum_{j=1}^J \ln \frac{2}{\sigma_j} f^*\left(\frac{e_j}{\sigma_j}\right) F^*\left(\frac{\lambda_j e_j}{\sigma_j}\right)$$

where $e_j = \frac{1}{(\mathbf{X}\boldsymbol{\beta})_{.j}}(u_j + w_{.j} + v_j)$, $\sigma_j = \left(\frac{1}{(\mathbf{X}\boldsymbol{\beta})_{.j}}\right) \sqrt{\sigma_u^2 + \frac{\sigma_w^2}{n_j} + \sigma_v^2}$, $\lambda_j = \sqrt{\frac{\sigma_v^2}{\sigma_u^2 + \sigma_w^2/n_j}}$.

This enables to use maximum likelihood estimation with heteroscedasticity for the stochastic frontier cost function.

Data and procedures

A Monte Carlo study can be used to examine heteroscedasticity of a stochastic frontier cost function on error terms. Based on equation (2), our true model is assumed as

$$(8) \quad C_{ij} = r_{ij} + u_j + w_{ij} + v_j.$$

where r_{ij} is the input price of the i th output in the j th firm, the others are the same as previously defined.

Aggregation over all outputs will derive following model:

$$(9) \quad \sum_i^{n_j} C_{ij} = TC_j = n_j r_{.j} + n_j (u_j + w_{.j} + v_j).$$

Taking a natural log and first order Taylor series around the mean of random errors and the frontier of inefficiency error result in the following model:

$$(10) \quad \ln TC_j \approx \ln(r_j) + \ln n_j + \frac{1}{r_j}(u_j + w_j + v_j).$$

So, our stochastic frontier cost function can be defined as

$$(11) \quad \ln TC_j = \beta_0 + \beta_1 \ln r_j + \beta_2 \ln n_j + \frac{1}{r_j}(u_j + w_j + v_j),$$

where heteroscedasticity is incorporated into the variances by assuming

$$\sigma_{u_j+w_j}^2 = (\delta_{1u} \sigma_u^2 + \delta_{1w} \frac{\sigma_w^2}{n_j}) \frac{1}{(r_j)^2} \quad \text{and} \quad \sigma_{v_j}^2 = \delta_{1v} \frac{\sigma_v^2}{(r_j)^2},$$
 true values of the parameters are

expected to be $\beta_0 = 0$, $\beta_1 = \beta_2 = \delta_{1u} = \delta_{1w} = \delta_{1v} = 1$. The input price is generated as

$r_{ij} \sim N(12, 4)$. The means of output is around 8.89 and the variance of that is around 112.

we assumed that there exist lots of small firms and a few of large firms.¹

In order to see the changes in a relative variability of error terms, we have two indicators ($\lambda \approx 1$ and $\lambda \approx 2$), i.e. first one has the same variability and the last has more variability in the inefficiency error, so that we can see how much the average inefficiency changes as the variability of inefficiency increases.

Using NLMIXED in SAS with 100 samples of 100 observations, the stochastic frontier cost function with heteroscedasticity and without heteroscedasticity is estimated. Outcomes are first compared with expected values to see how much the model is different from the true model and then compared each other with and without heteroscedasticity.

¹ The SAS code for output is `int(5*exp(rannor(12345)))+1`.

Results

Table 1 and Table 2 show the estimation results for the stochastic frontier cost function with 100 samples of 100 observations. We estimated $\beta's$ for the frontier function and $\delta's$ for variance equation, and variances and average inefficiencies. Second column has expected values for each parameter. Third column is the results with restricting heteroscedasticity and fourth column is the results without heteroscedasticity.

Overall, the estimated parameters are close to the expected values except the intercept and the δ_{1v} . For the intercept, it is interpreted as remainders by first order Taylor series. For the δ_{1v} , maximum value was imposed to be 0.1 because of the convergence. Later, it affects the variance of the inefficiency error to be smaller than what we expected.

First, looking at the parameters from the variance equation indicates that there exists heteroscedasticity in the translog form of the stochastic frontier cost function with aggregated data as expected.

Second, comparing the results with heteroscedasticity and without heteroscedasticity informs that there are almost no biases in the stochastic frontier cost function, which is quite opposite of previous research. This is because of first order Taylor series around zero inefficiency area so that the inefficiency is assumed to be almost zero. It is like heteroscedasticity in ordinary least square.

Third, let's focus on the average inefficiency($\sigma_v \sqrt{2/\pi}$) in table 1 and table 2. In table 1, both of the average inefficiency are almost same while in table 2, the average inefficiency without heteroscedasticity is 2 times bigger than that with heteroscedasticity. In other words, as the variability of the inefficiency error increases and dominates the random errors, the average inefficiency is overestimated when disregarding heteroscedasticity.

Conclusions

In the frontier estimation, the translog form with aggregated data is mostly used and outputs are usually included in the variance equation to see whether heteroscedasticity exists or not. Theoretically, the variation of the (total) stochastic frontier cost function with aggregated data increases as the output increases. When the translog form of the stochastic frontier cost function is estimated, all explanatory variables can inversely affect on the error terms. Our Monte Carlo study shows that heteroscedasticity is significant in the random effect and the unexplained error term while it is insignificant in the inefficiency error term. Also, it does not cause biases, which is quite opposite of previous research. This is because our model is approximately defined by first order Taylor series around zero inefficiency area. Most importantly, disregarding

heteroscedasticity causes the average inefficiency to be overestimated when the variance of inefficiency term dominates the other error terms.

Using first order Taylor series around the mean of inefficiency error might be more close to the previous research. Then, the relationship between the inefficiency measurement for each observation and output level with heteroscedasticity and without heteroscedasticity will give us more interesting findings. These will be for the future research.

Table 1. Estimation Results with 100 Samples of 100 Observations ($\lambda \approx 1$)

Parameters	Expected Value	MLE w/ Heteroscedasticity	MLE w/o Heteroscedasticity
β_0	0	0.41984 (0.04464)	0.41584 (0.04804)
β_1	1	0.87401*** (0.01783)	0.87530*** (0.01927)
β_2	1	1.00068*** (0.00152)	1.00168*** (0.00160)
δ_{1u}	1	1.76610*** (0.05038)	
δ_{1w}	1	0.83868* (0.25499)	
δ_{1v}	1	0.02115 ^a (0.00426)	
σ_{u+w}^2	0.01	0.01860 (0.00036)	0.01848 (0.00035)
σ_v^2	0.01	0.00021 (0.00003)	0.00020 (0.00004)
$\sigma_v \sqrt{2/\pi}$	0.079	0.01179	0.01139

Note: Standard deviations are reported in parentheses. Asterisk(*), double asterisk(**), and triple asterisk(***) denote significance on average at 10%, 5%, and 1%, respectively.

^a imposed between 0 and 0.1 because of convergence.

Table 2. Estimation Results with 100 Samples of 100 Observations ($\lambda \approx 2$)

Parameters	Expected Value	MLE w/ Heteroscedasticity	MLE w/o Heteroscedasticity
β_0	0	0.75872 (0.05613)	0.77062 (0.06032)
β_1	1	0.77906*** (0.02247)	0.77073*** (0.02403)
β_2	1	1.00010*** (0.00198)	1.00070*** (0.00205)
δ_{1u}	1	3.78227*** (0.08685)	
δ_{1w}	1	0.75843* (0.31322)	
δ_{1v}	1	0.00726 ^a (0.00483)	
σ_{u+w}^2	0.01	0.03233 (0.00051)	0.03195 (0.00052)
σ_v^2	0.04	0.00030 (0.00003)	0.00064 (0.00005)
$\sigma_v \sqrt{2/\pi}$	0.159	0.01380	0.02015

Note: Standard deviations are reported in parentheses. Asterisk(*), double asterisk(**), and triple asterisk(***) denote significance on average at 10%, 5%, and 1%, respectively.

^a imposed between 0 and 0.1 because of convergence.

References

- Aigner, D., C.A.K. Lovell, and P. Schmidt. 1977. "Formulation and Estimation of Stochastic Frontier Production Models." *Journal of Econometrics* 6: 21-37.
- Caudill, S.B., and J.M. Ford. 1993. "Biased in frontier estimation due to heteroscedasticity." *Economics Letters* 41: 17-20.
- Caudill, S.B., J.M. Ford, and D.M. Gropper. 1995. "Frontier Estimation and Firm-Specific Inefficiency Measures in the Presence of Heteroscedasticity." *Journal of Business & Economic Statistics* 13(1): 105-111.
- Dickens, W.T. 1990. "Error Components in Grouped Data: Is It Ever Worth Weighting." *The Review of Economics and Statistics* 72(2): 328-333.
- Farrell, M.J. 1957. "The Measurement of Productive Efficiency." *Journal of the Royal Statistical Society. Series A(General)* 120: 253-90.
- Greene, W.H. 1980. "On the Estimation of a Flexible Frontier Production Model." *Journal of Econometrics* 13: 101-115.
- Hadri, K. 1999. "Estimation of a Doubly Heteroscedastic Stochastic Frontier Cost Function." *Journal of Business & Economic Statistics* 17(3): 359-363
- Jondrow, J., C.A.K. Lovell, I.S. Materov, and P. Schmidt. 1982. "On the Estimation of Technical Inefficiency in the Stochastic Frontier Production Function Model." *Journal of Econometrics* 19: 233-238.

- Meeusen, W., and J. van den Broeck. 1977. "Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error." *International Economic Review* 18: 435-444.
- Schmidt, P. 1986. "Frontier Production Functions." *Econometric Reviews* 4: 289-328.
- Weinstein, M.A. 1964. "Query 2: The Sum of Values from a Normal and a Truncated Normal Distribution." *Technometrics* 6(1): 104-105.