Heterogeneity in Stochastic Frontier Models

Dr. V. SEETHA LAKSHMI
Department of Mathematics
Noorul Islam University, Kanyakumari District,Tamil Nadu, India
vseetha05@ymail.com

Abstract :

Here we combined all possibilities to model the firm level heterogeneity in stochastic frontier analysis. It is found that the unobserved and observed heterogeneities cause serious biases in inefficiency. The best model is used to assess the proposed inefficiency measurements. Data collected from 100 major dairy units were used to verify the efficiency of the proposed model.

1. Introduction

Of late, the traditional stochastic frontier analysis (SFA) have been extended to account for firm specific heterogeneity estimation. When the literature started to develop it was first assumed that in the models time invariant parts are representing inefficiency whereas time variant parts can be seen as firm or unit specific heterogeneity. Recently (Greene, 2004, 2005 a, b) this interpretation has radically changed. Now it has been assumed that such parts of firm specific effects which are not changing in time are mainly due to firm specific heterogeneity while the time variant part should be seen as inefficiency. To assess the validly of this view is not an easy Job. There may be firm specific factors which may not change in time and might be beyond the managerial limits. Hence these can be interpreted as time invariant heterogeneity. It is also possible that part of the inefficiency is time variant. This might be the case with firms under monopoly for which there may not exist full incentives to minimize costs.

If firm specific heterogeneity is not accounted, then it can create bias considerably in the estimates. Many authors (Farsi, Fillippini and Greene (2005)) studied network industries and compared different SFM's in a detailed manner. They found that the true random effects model gave significantly lower inefficiency values than the other models. They have observed a short coming of that model viz : the firm specific heterogeneity terms are assumed to be uncorrelated with the explanatory variables. Farsi et.al (2006 a, b 2007, 2009) have shown the advantages of panel data stochastic models in this type of studies. In this we discuss the different ways through which the firm specific heterogeneity can be taken in to account in the SFM. Observed heterogeneity can be taken in to account by incorporating firm specific heterogeneity either in the estimated distribution of inefficiency or in the cost function itself. Unobserved heterogeneity can be taken into account by randomizing some of the model in which case it is assumed that this randomization captures all time invariant unobserved heterogeneity.

The models are estimated for 100 dairy farms in Tamil Nadu, India. The result confirms the results of Maria Kopsakansas – Savolairen and Rauli Svento (2011).

2. Methodology and data details

Pitt and Lee (1981) is the most popular model which assumes all the heterogeneities to be explained by the covariates included in the basic random effects (RE) specification. They have assumed that the firm specific inefficiency is the same for all the years. This model has the weakness that it is assumed to all unmeasured heterogeneity in the inefficiency term and that the inefficiency is assumed to be not correlated with the included variables.

To overcome these problems one has to model the observed heterogeneity in the mean and (or) variance of the distribution of inefficiency or to the variance of the distribution of the frontier error term. This reduces the effect of the assumptions in the RE model. This new model can be termed as the REH model. A positive feature of REH is that it helps in getting a more precise estimation for the frontier. Greene (2005 a) has given an extension to the RE model called as true random effects (TRE) model. This is a modification of Kumbhakar and Hjalmarsson (1993) model. The difference is in the interpretation and method of estimation. The TRE model is basically an ordinary random effects model with one exception; here the time varying error component has the asymmetric distribution. It is actually a random effects model in which the time varying error component does not have a normal distribution. A firm specific time invariant random effect is added to represent the unobserved heterogeneity among firms. The inefficiency component now varies freely across time and firms and hence it is assumed that the unobserved differences across firms that remain constant over time are driven by unobserved characteristics rather than by inefficiency. By combining TRE and REH and calling it TREH model, the unobserved and observed heterogeneity can be accounted for. Greene (2005 a) proposes a model by name "true fixed effects model" (TFE) wherein firm specific constant terms are placed in the frontier. We present below in Table 1 the main features of the existing models.

Table 1. Details of existing models

Sl. No	Model Name	Observed heterogeneity	Unobserved heterogeneity
1.	Random effects model (RE)	Firm specific observed factors in the frontier	Not included in the model
2.	Heterogeneity extended random effects (REH)	Firm specific observed factors in the frontier Heterogeneity in the mean of inefficiency distribution	Not included in the model
3.	True random effects (TRE)	Firm specific observed factors in the frontier	Time invariant random component captures firm specific unobserved heterogeneity
4.	True fixed effects (TFE)	Firm specific observed factors in the frontier	Time invariant fixed component captures firm specific unobserved heterogeneity
5.	Heterogeneity extended true random effects (TREH)	Firm specific observed factors in the frontier Heterogeneity in the mean of inefficiency distribution	Time invariant random component captures firm specific unobserved heterogeneity

3. Data and analysis

The data consist of 100 major Dairy units in Tamil Nadu, India. The data covers the period of 10 years from 2001 to 2010. The details are directly collected by the investigators. The relative size of the farm varies considerably. The details of the data are presented below in Table 2.

We have used constant Rupee values by converting all money values by converting to the base year 2001 by using the retail price index. Costs (C) are expressed as average costs calculated as total annual costs per 100 litres of milk. This cost includes the expenses on the delivery constant up to the consumer networks. Annual output Y is measured in thousands. It varies significantly. CU is the total number of customers per farm. Annual Labour expense (AL) is calculated by averaging over all the farms. The Capital Investment (CI) is also computed accordingly. It includes the price of animals, building and other fixed asset costs.

Table 2. Descriptive statistics (100 dairy farms)

Sl. No	Variables	Mean in Rupees in thousand	Standard deviation in Thousands	Minimum in thousand	Maximum in thousand
1.	Total annual cost (C) per 100 litres of out put	1.272	.324	1.197	1.363
2.	Annual output (Y)	435.50	123.87	326.31	543.36
3.	Number of customers (CU)*	465	83	436	584
4.	By product income (BI)	12.63	3.21	14.82	10.64
5.	Annual Labour expense (AL)	42.43	2.75	38.79	43.56
6.	Capital Investment / farm (C.I)	8422.5	167.81	7541.8	9633.4
7.	Price of output / 100 litres (OP)	1.74	0.40	0.77	2.25

4. Model Specification and Estimation

We estimate the five modifications as discussed, using the Cobb – Douglas specifications for the frontier. Taking log, the log-linear SF for costs with random effects (RE) model can be written as

$$\ln C_{ij} = \sim + {}_{Y} \ln Y_{it} + {}_{BI} \ln BI_{it} + {}_{CU} \ln CU_{it} + {}_{AL} \ln AL_{it} + {}_{CI} \ln CI_{it} + {}_{T} T + v_{it} + u_{i}$$

Here $v_{it} = N\left(0, \pm^2_v\right), u_t = N\left(0, \pm^2_u\right) \quad i = 1,2.....,100$ .... (1)

T is for time to take care of the Technical changes due to time. The two factors by product income (BI) and the number of customers (CU) as the indicators of observable heterogeneity in this model. Even though these variables are time variant, the actual variation within one form is very small. In this case since the explained variable is the average costs, these variables will capture well the observed farm specific heterogeneity.

The next RE model is extended by the inclusion of a heterogeneity component into the mean of the distribution of u_i (REH)

$$\ln C_{it} = \sim + {}_{Y}\ln Y_{it} + {}_{BI}\ln BI_{it} + {}_{CU}\ln CU_{it} + {}_{AL}\ln AL_{it} + {}_{CI}\ln CI_{it} + {}_{T}T + v_{it} + u_{i}$$

Here

$$v_{it} = N\left(0, \pm_{v}^{2}\right), u_{i} = N\left(«_{i}, \pm_{u}^{2}\right), \quad «_{i} = ¢_{0} + ¢_{1}\ln BI_{i}, \quad i = 1,2.....,100$$

............(2)

The heterogeneity variable we use in the byproduct factor is defined as the average when it is optimum. It is clearly, higher in urban areas than in rural areas.

The third estimated model is, TRE, is the random parameter version of the RE model, here also the inefficiency term (u) is time variant. A firm specific random constant term is used here.

$$\ln C_{it} = \left(\sim + w_{i}\right) + {}_{Y}\ln Y_{it} + {}_{BI}\ln BI_{it} + {}_{CU}\ln CU_{it} + {}_{AL}\ln AL_{it} + {}_{CI}\ln CI_{it} + {}_{T}T + v_{it} + u_{it}$$

Here $v_{it} = N\left(0, \pm_{v}^{2}\right), u_{it} = N\left(0, \pm_{u}^{2}\right), \quad i = 1,2.....,100$ .... (3)

We estimate the true fixed effect model (TFE).

$$\ln C_{it} = \sim_{i} + {}_{Y}\ln Y_{it} + {}_{BI}\ln BI_{it} + {}_{CU}\ln CU_{it} + {}_{AL}\ln AL_{it} + {}_{CI}\ln CI_{it} + {}_{T}T + v_{it} + u_{it}$$

Here $v_{it} = N\left(0, \pm_{v}^{2}\right), u_{it} = N\left(0, \pm_{it}^{2}\right) \quad i = 1,2.....,100$ .... (4)

The fifth estimated model is the TREH model which is a combination of models REH and TRE. The reason being that this combination can take both the observed and unobserved heterogeneities into account at the same time. The mathematical structure of this combinationis

$$\ln C_{it} = \left(\sim + w_{i}\right) + {}_{Y}\ln Y_{it} + {}_{BI}\ln BI_{it} + {}_{CU}\ln CU_{it} + {}_{AL}\ln AL_{it} + {}_{CI}\ln CI_{it} + {}_{T}T + v_{it} + u_{it}$$

Here $u_{it} = N\left(«_{it}, \pm_{it}^{2}\right), \quad «_{it} = ¢_{0} + ¢_{1}\ln BI_{it}, \quad i = 1,2.....,100$ .... 5)

5. Results and discussion :

All the estimated informations from all the 5 equations are presented in Table 3 below: Results presented reveals that.

(i) All the covariate coefficients of the frontier are highly significant except the BI in the RE model, the coefficients have expected signs.

(ii) Both the price effects have positive signs in all the models specified and the CI (capital investment) effect is larger in absolute terms in all the 4 models except the TFE model. This might he due to the capital intensity of distribution networks.

(iii) The sign of output estimator (Y) is positive indicating it's value greater than one which is the limit price of the output.

(iv) As the distributed quantity increases the unit costs decreases up to a point of minimum efficient scale.

(v) The time estimate has a negative sign. This might be due to the fact that the technical efficiency might have reduced the unit wise production cost.

Table 3 : Cost Frontier Parameters of Models 1 - 5

Variable Name	Name of Model									
	RE		REH		TRE		TFE		TREH	
	Statistical Details									
	Coeff	Std.er	Coeff	Std.er	Coeff	Std.er	Coeff	Std.er	Coeff	Std.er
Constant	1.93752	0.326	-2.7912	0.1512+0.07	-1.780	0.059	-	-	-1.295	0.0828
ln Y	.776	0.064	-0.788	0.060	-0.844	0.012	-0.714	0.0216	0.804	0.016
ln BI	+0.684	0.06	-0.024	0.0576	-0.0384	0.017	-0.422	0.039	0.041	0.023
ln CU	0.701	0.065	0.724	0.062	0.773	0.012	0.683	0.0192	0.698	0.0168
ln AL	0.356	0.010	0.346	0.009	0.360	0.005	0.534	0.017	0.385	0.007
ln CI	0.463	0.011	0.473	0.011	0.482	0.004	0.333	0.015	0.476	0.006
ln BI^2	-	-	-2.980	0.810	-	-	-	-	-	-
T	-0.017	0.003	-0.017	0.002	-0.018	0.001	-0.020	0.006	-0.017	0.001
Scale Par. For distr	-	-	-	-	0.232	0.003	-	-	0.180	0.003
Log Likelihood	392.35		416.71		420.45		313.38		429.62	
N – Size	100		100		100		100		100	
BIC - Critieria	-742		-785		-801		-573		-813	

In the 7th row ln BI^2 stand for the estimate from the model $«_i = ¢_0 + ¢_1 \ln BI_i$ in equation (2). BIC = -2 log L + Q log N, Where Q is the number of parameters.

Table 4 : Statistics of inefficiency scores. (Jondrow et. Al (1989)

	RE	REH	TRE	TFE	TREH
Minimum	0.0987	0.0426	0.0113	0.0584	0.0102
Maximum	0.795	0.503	0.465	0.148	0.449
Mean	0.343	0.154	0.074	0.078	0.081
Std. Deviation $E(u_i/£_i)$ E[Ui/Ei) of	0.131	0.074	0.048	0.0095	0.061
σ (v)	0.067	0.068	0.033	0.159	0.024
σ (u)	0.354	0.152	0.094	0.102	0.107

The variance parameter of the underlying distribution of $u_i, ±_u$ is estimated as 0.354 (Table 4) in RE model. The value of their counter parts in REH, TRE, TFE and TRCH are respectively 0.152, 0.094, 0.102 and 0.107. These imply that some of the variation in the inefficiency in the original RE model can be explained as heterogeneity. Based on this we can expect a decrease in inefficiencies. But according to BIC – Criteria it implies that the model accounts both the observed and unobserved heterogeneities at the same time. That is the combined TREH fits the data best.

The inefficiency scores presented in Table 4 represents the expected percentage deviation from a minimum level that would have been incurred if the farm had operated as best practice (or cost efficient) based on the data.

Thus the study indicates that firm / farm specific inefficiency scores based on the true fixed effects model are very close to each other and if we look at the variance of the frontier, it is high showing that the model does not produce robust estimates for the frontier. This might be due to the insufficiency in the size of the sample. The result by BIC – Criteria, the model which combines the characteristics of unobserved and observed heterogeneity fits the data best and this point to the importance of the taken unobserved heterogeneity into account.

References

1. Farsi, M., Filippimi, M., 2009. An analysis of cost efficiency in Swiss multi – utilities. Energy Economics 31, 306 – 315.
2. Farsi, M., Filippini, M., Greene, W., 2005. Efficiency measurement in network industries: application to the Swiss railway companies. The Journal of Regulatory Economics 28 (1), 69-60.
3. Farsi, M., Greene, W., 2006 a. Application of panel data models in benchmarking analysis of the electricity distribution sector. Annals of public and co-operative Economics 77(3), 271-290.
4. Farsi, M., Filippimi, M., Kuenzle, M., 2006 b. Cost efficiency in regional bus companies: an application of alternative stochastic frontier models. Journal of Transport Economics and Policy 40 (1), 95-118.
5. Farsi, M., Kuenzle, M., 2007. Cost efficiency in the Swiss gas distribution sector. Energy Economics 29, 64-78.
6. Greene, W., 2004. Distinguishing between heterogeneity and inefficiency: stochastic frontier analysis of the World Health Organization's panel data on national health care systems. Health Economics 30, 959 – 980.
7. Greene, W., 2005a. Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. Journal of Econometrics 126, 269-303.
8. Greene, W., 2005b. Fixed and random effects in stochastic frontier models. Journal of Productivity Analysis 23, 7-32
9. Kumbhakar, S.S., Hjalmarsson, L., 1993. Technical efficiency and technical progress in Swedish dairy farms.In: Fired, H., Lovell, K. Schmidt, S., (Eds), The measurement of productive efficiency: Techniques and Applications. Oxford University Press, Now York.
10. Kumbhakar, S., Lovell, K., 2000. Stochastic Frontier Analysis. Cambridge University Press, Cambridge.
11. Pitt, M., Lee, L., 1981. The measurement and sources of technical inefficiency in Indonesian Weaving Industry, Journal of Development Economics 9, 43-64.
12. Schmidt, P., Sickles, R.E., 1984. Production frontiers and panel data. Journal of Business and Economic Statistics 2, 367 – 374.