Heteroscedasticity and non-monotonic efficiency effects of a stochastic frontier model

Wang, Hung-Jen

April 2002

Heteroscedasticity and Non-Monotonic Efficiency Effects of a Stochastic Frontier Model

Hung-Jen Wang*
The Institute of Economics
Academia Sinica
Taipei 115
Taiwan
hjwang@econ.sinica.edu.tw

April 2002

Abstract

We consider a model that provides flexible parameterizations of the exogenous influences on inefficiency. In particular, we demonstrate the model's unique property of accommodating non-monotonic efficiency effect. With this non-monotonicity, production efficiency no longer increases or decreases monotonically with the exogenous influence; instead, the relationship can shifts within the sample. Our empirical example shows that variables can indeed have non-monotonic effects on efficiency. Furthermore, ignoring non-monotonicity is shown to yield an inferior estimation of the model, which sometimes results in opposite predictions concerning the data.

*I would like to thank River Huang, two anonymous referees, and particularly Subal Kumbhakar (the Editor) for helpful comments and suggestions. I gratefully acknowledge the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) for granting the permission to use the data, and George Battese and Tim Coelli for providing the data.

1 Introduction

Ever since the work of Aigner, Lovell, and Schmidt (1977) and Meeusen and van den Broeck (1977), both the theoretical development and empirical application of stochastic frontier models have thrived in the literature; see Bauer (1990) and Greene (1993) for earlier reviews on the literature. One of the most important recent developments lies in the investigation of how exogenous factors influence the one-sided inefficiency effect. This effort allows researchers to understand not only the production unit's state of efficiency, but also the contributing factors of the efficiency.

Kumbhakar and Lovell (2000, Ch. 7) discuss in detail how the literature evolves from the early two-step approach, by which inefficiency and exogenous effects are identified sequentially, to the more recent one-step approach by which the exogenous effects are estimated simultaneously with the model's other parameters. The extensive Monte Carlo results presented by Schmidt and Wang (2002) give evidence in favor of the one-step approach.

Perhaps the most well-known model of the one-step approach is that of Kumbhakar, Ghosh, and McGuckin (1991), Huang and Liu (1994), and Battese and Coelli (1995) (KGMHLBC hereafter). These authors propose parameterizing the *mean* of the pre-truncated distribution as a way to study the exogenous influence on inefficiency. In a seemingly different vein of the literature, Caudill and Ford (1993), Caudill, Ford, and Gropper (1995), and Hadri (1999) (CFCFGH hereafter) seek to address the problem of heteroscedasticity by parameterizing the *variance* of the pre-truncated distribution. As pointed out by Kumbhakar and Lovell (2000), the model of CFCFGH can also be seen as another approach to study the exogenous effects on inefficiency, and its modeling strategy complements that of KGMHLBC.

In this paper, we consider a model of exogenous influence that explicitly combines KGMHLBC and CFCFGH, and we explore the property of a non-monotonic efficiency effect which is unique to this combination. For many researchers, the combined model is not new, for a similar one has been known to them for some time (though not seen in published articles). The major contribution of this paper is in demonstrating the model's unique ability to accommodate non-monotonic efficiency effects, which can be very important and useful in understanding the relationships between the inefficiency and its exogenous determinants.

Two variables having a non-monotonic relationship implies that their values can be positively related in part of the parameter space while negatively related in the rest. Such a relationship between economic variables is not new, with well-known examples including tax rates vs. government revenues, and wage rates vs. labor supply. The relationship can also take place regarding technical efficiency. For instance, while a farmer's age could represent experiences helpful in improving production efficiency, an old farmer is nevertheless likely to have a deteriorated mental and physical capacity, resulting in a negative efficiency effect. In this example, a young farmer's efficiency may improve as he matures, but the age factor eventually becomes detrimental to efficiency in the farmer's later years. Ignoring the non-monotonicity in this aspect can render estimation results imprecise at best and misleading at worst. Therefore, a model that allows for non-monotonic efficiency effects across observations can better describe the data, and the results can also be more informative for the purpose of policy analysis. Using data on Indian farmers as an example, we show that the model indeed explains the data better; more importantly, the implied marginal effects are more plausible than those offered by other models.

The rest of this paper is organized as follows. Section 2 presents the model with the property of non-monotonicity. The meaning of non-monotonicity and its economic importance are also discussed here. Section 3 provides an empirical example using production data for farmers from an Indian village. The results show that non-monotonicity is important in understanding the relationship between efficiency and its determinants. Section 4 summarizes the results and concludes the paper.

2 A Stochastic Frontier Model with Non-Monotonicity

A production frontier model that combines features of KGMHLBC and CFCFGH can be expressed as follows:

$$
\begin{aligned}
y_{it} &= \mathbf{x}_{it}\beta + (v_{it} - u_{it}), && (1)\\
v_{it} &\sim N(0, \sigma_v^2), && (2)\\
u_{it} &\sim N^+(\mu_{it}, \sigma_{it}^2), && (3)\\
\mu_{it} &= \mathbf{z}_{it}\delta, && (4)\\
\sigma_{it}^2 &= \exp(\mathbf{z}_{it}\gamma). && (5)
\end{aligned}
$$

In this setup, u_{it} is the inefficiency effect, which is a non-negative truncation of a normal random variable. The variable vector $\mathbf{z}_{it}$ includes a constant of 1 and some other exogenous variables associated with the inefficiency. The δ and γ are the corresponding coefficient vectors. All other notations and definitions follow the literature. The model encompasses KGMHLBC and CFCFGH as special cases. For KGMHLBC, this amounts to replacing $\mathbf{z}_{it}$ in (5) by a single constant of 1; for CFCFGH, it is obtained by substituting 0 for $\mathbf{z}_{it}$ in (4).

In our empirical application, we also consider the view of Caudill *et al.* (1995) and Hadri (1999), which assumes

$$\sigma_v^2 = \exp(\mathbf{z}_{i\,t}\lambda). \tag{6}$$

This is a less consequential parameterization (Kumbhakar and Lovell 2000), and our focus is on the model consisting of (1) to (5). The log-likelihood function of the above model is a straightforward extension of the KGMHLBC; we shall skip copying the function here.

Kumbhakar and Lovell (2000) discuss many of the properties of KGMHLBC and CFCFGH, and the discussions apply equally well to the above model. We therefore focus on the model's unique property of the non-monotonic efficiency effect, which has not been mentioned anywhere in the literature.

Non-Monotonic Efficiency Effects

By a non-monotonic efficiency effect, we mean that $\mathbf{z}_{i\,t}$ can have, within a sample, *both* positive and negative effects on the production efficiency, and that the sign of the effect depends on values of $\mathbf{z}_{i\,t}$. For instance, the kth element of $\mathbf{z}_{i\,t}$, $z[k]$, can positively (negatively) affect the efficiency when values of $z[k]$ are within a certain range, and the impacts can then turn negative (positive) for values of $z[k]$ outside the range. On an efficiency vs. $z[k]$ plot, the line representing the non-monotonic relationship would have both positive and negative slopes on the graph, indicating that $z[k]$ can be both efficiency-enhancing and efficiency-impeding within the sample, with the outcome depending on the values of $z[k]$.

In contrast to this, the effect in the model of KGMHLBC is necessarily monotonic (will be shown later), and $z[k]$ is only *either* efficiency-enhancing *or* efficiency-impeding in the sample, but not both. On an efficiency vs. $z[k]$ plot, the line would either slope upward or slope downward, but it would not have both characteristics.

The ability to accommodate non-monotonic efficiency effects is important for models seeking to understand the relationships between efficiency and the exogenous factors. The reason is obvious: many of the relationships between economic variables are indeed non-monotonic. The example of the farmer's age given in the introduction illustrates the point. While a young farmer's efficiency may benefit from an increase in age as experiences accumulate, an aged farmer, however, may likely suffer from efficiency loss, because of deteriorated mental and physical capability. In this example, efficiency increases with age in the early years of the farmer's life, but it decreases with age in the later years. Models ignoring non-monotonicity would conclude either that age is efficiency-improving or that age is efficiency-impeding; neither in fact is correct. Other examples of non-monotonicity include using financial ratios to explain firms' operating efficiency: the notion of "optimal" or "desired" ratios implies non-monotonic relationships between the ratios and the efficiency.

Non-monotonicity can also have important policy implications. For studies on farmers' efficiency, knowing the optimal farm size or the farmer's most productive age can help design policies aimed at increasing the overall efficiency. Understanding the best (efficiency-wise) R&D to sales ratio (e.x., Huang and Liu 1994) also helps firms to spend resources on R&D in the most efficient way. Such policy implications are unobtainable from models ignoring the non-monotonicity of the efficiency effect.

In what follows, we demonstrate the model's ability to accommodate $z[k]$'s non-monotonic effects on the mean and the variances of $u_{i\,t}$ measured by the unconditional statistics of $E(u_{i\,t})$ and $V(u_{i\,t})$, respectively.[1] Both statistics are consistent and observation-specific. While the mean measures the expected value of technical inefficiency, the variance measures production uncertainty (Bera and Sharma 1999). To demonstrate non-monotonicity, our strategy is to show that the marginal effect of $z[k]$ on $E(u_{i\,t})$ and/or $V(u_{i\,t})$ can be both positive and negative in the sample. If the signs can alternate in the sample, then this implies that the impacts of $z[k]$ can go in both directions.

To this end, we write down the first two moments (i.e., the mean and the variance) of $u_{i\,t}$ as follows.

$$
\begin{aligned}
m_1 &= f(\mu_{i\,t},\ \sigma_{i\,t}) = \sigma_{i\,t}\left[\Lambda + \frac{\phi(\Lambda)}{\Phi(\Lambda)}\right], && (7)\\
m_2 &= g(\mu_{i\,t},\ \sigma_{i\,t}) = \sigma_{i\,t}^2\left[1 - \Lambda\left[\frac{\phi(\Lambda)}{\Phi(\Lambda)}\right] - \left[\frac{\phi(\Lambda)}{\Phi(\Lambda)}\right]^2\right], && (8)
\end{aligned}
$$

where $\Lambda = \mu_{i\,t}/\sigma_{i\,t}$, and ϕ and Φ are the probability and cumulative density functions of a standard normal distribution, respectively.[2]

Taking into account the parameterization functions (4) and (5), the marginal effect of $z[k]$ on $E(u_{i\,t})$ is

$$
\begin{aligned}
\frac{\partial \mathrm{E}(u_{i\,t})}{\partial z[k]} &= \delta[k]\left[1 - \Lambda\left[\frac{\phi(\Lambda)}{\Phi(\Lambda)}\right] - \left[\frac{\phi(\Lambda)}{\Phi(\Lambda)}\right]^2\right] && (9)\\
&\quad + \gamma[k]\,\frac{\sigma_{i\,t}}{2}\left[\left(1+\Lambda^2\right)\left[\frac{\phi(\Lambda)}{\Phi(\Lambda)}\right] + \Lambda\left[\frac{\phi(\Lambda)}{\Phi(\Lambda)}\right]^2\right],
\end{aligned}
$$

[1] See the end of this section for a discussion on the conditional statistics' marginal effects.

[2] It is worth noting that while KGMHLBC and CFCFGH allow $\mathbf{z}_{i\,t}$ to affect the moments of $u_{i\,t}$ through only the channel of either $\mu_{i\,t}$ or $\sigma_{i\,t}$, the combined model allows an exogenous variable to exert influence through both channels. Since it is *a priorily* unjustifiable to preclude either of the two channels, the combined model is thus more complete in the modeling of exogenous influence.

where $\delta[k]$ and $\gamma[k]$ are the corresponding coefficients in (4) and (5), respectively. The equation shows that the marginal effect is the sum of the adjusted slope coefficients from the mean and the variance functions.

In the case of KGMHLBC, $\sigma_{i\,t}^2$ is constant so that $\gamma[k] = 0$ for all k, and thus the marginal effect is the slope coefficient $\delta[k]$ multiplied by an adjustment function. This adjustment function is just $m_2/\sigma_{i\,t}^2$ (see (8)), and is thus positive. Therefore, for a simple model with constant $\sigma_{i\,t}^2$, the marginal effect has the same sign as the slope coefficient, and the sign is the same for *all* of the sample's observations, indicating a monotonic efficiency effect of $z[k]$. For models with parameterized $\sigma_{i\,t}^2$, the marginal effect consists of both terms on the right-hand side of (9), and so the signs of the marginal effects do not necessarily coincide with the signs of either of the slope coefficients. As our empirical application will show, the signs can indeed alternate in the sample.

The marginal effect of $z[k]$ on $\mathrm{V}(u_{i\,t})$ is

$$\begin{aligned}\frac{\partial V(u_{i\,t})}{\partial z[k]} &= \frac{\delta[k]}{\sigma_{i\,t}}\left[\frac{\phi(\Lambda)}{\Phi(\Lambda)}\right](m_1^2 - m_2) \qquad (10)\\ &+\gamma[k]\sigma_{i\,t}^2\left\{1 - \frac{1}{2}\left[\frac{\phi(\Lambda)}{\Phi(\Lambda)}\right]\left(\Lambda + \Lambda^3 + (2+3\Lambda^2)\left[\frac{\phi(\Lambda)}{\Phi(\Lambda)}\right] + 2\Lambda\left[\frac{\phi(\Lambda)}{\Phi(\Lambda)}\right]^2\right)\right\},\end{aligned}$$

where m_1 and m_2 are given in (7) and (8), respectively. Again, the marginal effect is the sum of the adjusted slope coefficients. Bera and Sharma (1999, equation (16)) state that $m_1^2 - m_2 > 0$, based on a result from Barrow and Cohen (1954). Therefore, for models with constant $\sigma_{i\,t}^2$, the effect of $z[k]$ is again monotonic, and the effect can be non-monotonic when $\sigma_{i\,t}^2$ is parameterized.

We do not calculate here the marginal effects on the conditional statistics, $E(u_{i\,t}|\hat{e}_{it})$ and $V(u_{i\,t}|\hat{e}_{it})$, because the calculations are almost intractable, particularly when (5) and (6) are adopted.[3] Nevertheless, it can be shown that for a model of constant $\sigma_{i\,t}^2$ (i.e., KGMHLBC), the variables' effects on these conditional statistics are monotonic. As a sketch, the simple model's marginal effect on $E(u_{i\,t}|\hat{e}_{it})$ is the same as the first term on the right-hand side of (9), except that $\mu_{i\,t}$ and $\sigma_{i\,t}$ are replaced by μ_* and σ_*, respectively, where

$$\mu_* = \frac{\sigma_v^2\mu_{i\,t} - \sigma_{i\,t}^2(y_{i\,t} - \mathbf{x}_{i\,t}\beta)}{\sigma_v^2 + \sigma_{i\,t}^2}, \qquad (11)$$

$$\sigma_* = \frac{\sigma_v\sigma_{i\,t}}{\sqrt{\sigma_v^2 + \sigma_{i\,t}^2}}. \qquad (12)$$

Similarly, the marginal effect on $V(u_{i\,t}|\hat{e}_{it})$ is obtained by taking the first term on the right-hand side of (10), replacing $\mu_{i\,t}$ and $\sigma_{i\,t}$ by μ_* and σ_*, respectively, and then scaling the entire expression by $\sigma_*/\sigma_{i\,t}$. As can be easily seen, both imply monotonic effects of $z[k]$.

[3] On the other hand, the unconditional mean and variance do not involve σ_v^2, and the marginal effect is easier to derive.

3 Empirical Application

Annual data from 1975-76 to 1984-85 on farmers from the village of Aurepalle in India are used in this empirical illustration. This dataset and a subset of it have been used by Battese and Coelli (1995) and Coelli and Battese (1996). We adopt the model of Battese and Coelli (1995) for this empirical illustration.

Following the notation in equations (1) to (6), the model is specified as

y_{it}: $\ln(Y_{it})$;

$\mathbf{x}_{it}$: $\{\ln(Land_{it}),\ (PILand_{it}),\ \ln(Labor_{it}),\ \ln(Bullock_{it}),\ \ln[\mathrm{Max}(Cost_{it}, 1 - D_{it})],\ Year_{it}\}$;

$\mathbf{z}_{it}$: $\{Age_{it},\ Schooling_{it},\ Year_{it}\}$,

where (see also Battese and Coelli, 1995)

Y is the total value of output;

Land is the total area of irrigated and unirrigated land operated;

PILand is the proportion of operated land that is irrigated;

Labor is the total hours of family and hired labor;

Bullock is the hours of bullock labor;

Cost is the value of other inputs, including fertilizer, manure, pesticides, machinery, etc.;

D is a variable which has a value of one if *Cost* is positive, and a value of zero if otherwise;

Age is the age of the primary decision-maker in the farming operation;

Schooling is the years of formal schooling of the primary decision maker; and

Year is the year of the observations involved.

Prior to the estimation, the model's production function was subjected to the method of Hadi (1994) for identifying possible outliers in the data. Two such observations were identified and subsequently dropped from the estimation.[4] The remaining part of the data is an unbalanced panel of 34 farmers with a total of 271 observations. We have a full account of 10 observations for 16 of the farmers, and 2 minimal observations for 2 of the farmers.

Estimations of the maximum likelihood function are carried out using Stata 7.0 computer software. Its maximization routine uses a combination of the steepest ascent and the Newton-

[4]In one of the identified outliers, the family labor hours were about four times higher than in the preceding and the succeeding years, the bullock hours were doubled, and the cost of other inputs increased significantly, while at the same time the output was only about 60% of the preceding and succeeding years. The other identified outlier followed a similar pattern: while inputs increased substantially compared to the preceding and succeeding years, the output dropped significantly. These two observations belonged to two different farmers and to two different years. The same irregular pattern is not observed in other observations for the same year or same farmers.

Raphson algorithms.

3.1 Estimation Results

Four different models are estimated. All of them parameterize μ_{it}, which is the feature of KGMHLBC, and they differ on whether σ_{it}^2 and σ_v^2 are also parameterized. Model (i) parameterizes neither, and thus is the specification of KGMHLBC. Model (ii) parameterizes σ_{it}^2, resulting in a model that combines the features of KGMHLBC and CFCFGH. Model (iii) parameterizes σ_v^2 instead. Model (iv) parameterize both σ_{it}^2 and σ_v^2.

We present all the estimated coefficients of β, δ, γ, and λ in the working paper version of this article; the results are available from the author upon request. To focus on the non-monotonic efficiency effects, here we opt to present only the marginal effects of $\mathbf{z}_{it}$, namely *Age*, *Schooling*, and *Year*, on the two statistics of efficiency, $E(u_{it})$ and $V(u_{it})$. These marginal effects are calculated from (9) and (10) after the model's parameters are estimated. The results are presented in Table 1. The Table reports the sample means of the marginal effects, as well as the average marginal effects of the first and the last quarter of the sample (ordered by values of the effects). Bootstrapped standard errors of the corresponding statistics are also reported, along with statistical significance levels based on bias-corrected and accelerated confidence intervals. For an easier presentation of the data, we also plot the observation-wise marginal effects in Figures 1 to 6.

The remaining discussions are based on the results of Model (ii), which combines features of KGMHLBC and CFCFGH and is the main model of this paper. The selection also draws some statistical support. Since Models (i), (ii) and (iii) are obviously nested to Model (iv), we perform likelihood ratio tests on the hypotheses that the nested models are preferred to Model (iv). (This amounts to testing that the respective coefficient vectors are zero.) The results (available upon request) show that Models (i) and (iii) are rejected in favor of Model (iv) at the 5% level, while Model (ii) is rejected in favor of Model (iv) only at a marginal 10% level. As one can see from Table 1 and the associated figures, results of Model (ii) and Model (iv) are similar, both quantitatively and qualitatively.

For the variable *Age*, the marginal effects on $E(u_{it})$ and $V(u_{it})$ measure how an increase in the farmer's age changes the expected inefficiency and the production uncertainty (Bera and Sharma 1999), respectively. For Model (ii), results on $E(u_{it})$ in the upper panel of Table 1 show that an increase in age helps *reduce* (i.e., negative effects) production inefficiency for younger farmers (i.e., farmers in the first age-quartile group). The average marginal effect in the first age-quartile is -0.0113. Since $\partial E(\ln y)/\partial Age = -\partial E(u)/\partial Age$, the effect translates into an increase in the output by 1.13%. Although quantitatively small, this figure is statistically significant. On the other hand, for the older farmers in the last age-quartile group, increases in age tend to be counterproductive, leading to inefficiency increases (i.e., positive effect). The loss in the output growth is about 0.3%. The opposite marginal effects in these two quartiles indicate that *Age* affects efficiency non-monotonically in the sample. Although it is difficult to tell when the impact of *Age* turns from negative to positive, Figure 1 implicates that the turn is likely to have taken place between the ages of 50 and 60. This appears to be quite sensible.

In addition to the signs, trends in individual farmers' marginal effects are also interesting; they are visible in Figure 1.[5] Model (ii) in the Figure shows that the trend is downward sloping for any given farmer, indicating increases (decreases) in advantages (disadvantages) as one gets older. It is in this respect that experiences possibly contribute to an improvement in efficiency.

The marginal effects also differ in regard with the production uncertainty measured by $V(u_{i\,t})$, as shown in the lower panel of Table 1. For younger farmers, *Age* tends to reduce production uncertainty perhaps through making fewer mistakes, while for older farmers the uncertainty increases with age. Together with the results on $E(u_{i\,t})$, Model (ii) thus predicts that, other things being equal, a young farmer is likely to achieve higher and more stable output growth as time goes by, while an older farmer would have lower and more variable output growth.

For the variable *Schooling*, results show that education is most valuable when the number of years of schooling is relatively low (zero or one year), and the benefit disappears at the higher education level.[6] For Model (ii), the effects on $E(u_{i\,t})$ and $V(u_{i\,t})$ are both negative in the first schooling quartile, indicating improvements in efficiency and reductions in production uncertainty from additional education. The efficiency improvement amounts to a 5.05% increase in the output. On the other hand, the effect in the last quartile is positive; the non-monotonicity is evident. By contrast, Models (i) and (iii) do not seem to predict an efficiency-enhancing education effect, as the marginal effects on $E(u_{i\,t})$ are positive in both models.

Figures 2 and 5 show *Schooling*'s marginal effects graphically. Because *Schooling* has only eight distinct values and the value is fixed for a given farmer, we opt to use box plots in the

[5]Because the data have a diverse age profile and age is continuous for a given farmer, trends in an individual farmer's marginal effects are easily visible from the figure.

[6]Note that 62% of the observations have *Schooling* equal to 0 or 1.

figures to better visualize the data.[7] We also put the results of the two models in the same graph for a better comparison and to save space.

The variable *Year* is used to capture the time trend effect on technical efficiency, which could result from adoptions of new technologies. For all the four models, the marginal effects tend to be negative, indicating an improvement in technical efficiency over time. There are, however, signs that this improvement has gradually slowed down, as indicated by the zero-reverting trends in Figure 3.

The above results demonstrate the presence of non-monotonic efficiency effect in the model. A question can be raised at this point as whether similar non-linear effects could result simply from adding square terms of *Age* etc. to the model of KGMHLBC. We estimate models with these added square terms, and the results (available upon request) show that the added terms do not have the joint statistical significance, and the implied marginal effects are also less compelling. It is worth to recall that the model considered in this paper begins with flexible parameterizations of the exogenous influences, and the non-monotonicity arises naturally only as a result.

Before ending this section, it is interesting to note that the variable's marginal effects on $E(u_{i\,t})$ and $V(u_{i\,t})$ seem to be qualitatively similar. For instance, *Age* reduces both $E(u_{i\,t})$ and $V(u_{i\,t})$ in the first age-quartile group, and it increases both $E(u_{i\,t})$ and $V(u_{i\,t})$ in the fourth age-quartile group. This observation is similar to a finding from Bera and Sharma (1999) that, in their model, when a firm moves toward the production frontier by having higher efficiency (lower $E(u_{i\,t})$), it also reduces production uncertainty (lower $V(u_{i\,t})$) at the same time. We have to point out, however, that although the same result generally holds in our model, it is not necessarily true in all the cases. By inspecting values of the marginal effects on $E(u_{i\,t})$ and $V(u_{i\,t})$ of Model (ii), we find that the pair-wise marginal effects do not necessarily have the same sign for a given observation (although the majority of them do). The implication is that, for the present model, an exogenous variable may move a farmer toward the frontier, but it does not necessarily reduce production uncertainty at the same time.

[7] The box extends from the group data's 25th percentile to the 75th percentile, or the interquartile range. The line in the middle indicates the median. Lines emerging from the box extend in both directions to observations within, but most close to, the boundaries defined by 1.5 times the 75th and the 25th percentiles.

4 Conclusions

This paper investigates the properties of a model that combines the features of KGMHLBC and CFCFGH. The combined model allows exogenous variables to affect inefficiency through two different channels, and a result of this flexibility is the model's ability to accommodate non-monotonic efficiency effects.

Using the data on farmers from an Indian village, we compare the empirical performances of four models, each differing in the assumption regarding channels through which exogenous variables affect u_{it}. The results show that the marginal effect estimates are sensitive to the different assumptions, and sometimes opposite predictions can result when different assumptions are adopted. In general, models that allow exogenous variables to work through both the mean and the variance of the pre-truncated distribution yield the most plausible estimates.

Our preferred model indicates that the non-monotonic relationship between variables and inefficiency is important. For instance, we find that an increase in age is efficiency-improving for young farmers, but is efficiency-impeding for senior farmers. Models that ignore the non-monotonicity greatly compromise the estimation.

Table 1: Marginal Effects on Inefficiency

		(i)	(ii)	(iii)	(iv)
marginal effects on $E(u_{it})$					
Age	sample avg.	-0.0006 (0.0021)	-0.0026 (0.0027)	0.0012 (0.0022)	-0.0017 (0.0027)
	1st quarter avg.	-0.0010 (0.0026)	-0.0113 (0.0050)***	0.0005 (0.0020)	-0.0107 (0.0048)***
	4th quarter avg.	-0.0003 (0.0024)	0.0030 (0.0073)	0.0021 (0.0032)	0.0044 (0.0065)
Schooling	sample avg.	0.0041 (0.0097)	-0.0120 (0.0162)	0.0112 (0.0112)	-0.0073 (0.0167)
	1st quarter avg.	0.0022 (0.0090)	-0.0505 (0.0265)***	0.0050 (0.0093)	-0.0462 (0.0280)***
	4th quarter avg.	0.0066 (0.0143)	0.0134 (0.0366)	0.0205 (0.0172)	0.0194 (0.0344)
Year	sample avg.	-0.0300 (0.0138)*	-0.0265 (0.0153)	-0.0367 (0.0205)	-0.0242 (0.0270)
	1st quarter avg.	-0.0487 (0.0295)**	-0.0583 (0.0317)***	-0.0669 (0.0383)*	-0.0564 (0.0372)**
	4th quarter avg.	-0.0160 (0.0066)**	0.0107 (0.0219)	-0.0163 (0.0125)	0.0130 (0.0316)
marginal effects on $V(u_{it})$					
Age	sample avg.	-0.0003 (0.0008)	0.0001 (0.0015)	0.0005 (0.0008)	0.0006 (0.0017)
	1st quarter avg.	-0.0005 (0.0010)	-0.0021 (0.0030)	0.0002 (0.0006)	-0.0017 (0.0026)
	4th quarter avg.	-0.0001 (0.0011)	0.0024 (0.0022)*	0.0010 (0.0015)	0.0033 (0.0028)*
Schooling	sample avg.	0.0018 (0.0038)	0.0002 (0.0064)	0.0050 (0.0038)*	0.0025 (0.0069)
	1st quarter avg.	0.0008 (0.0037)	-0.0094 (0.0128)	0.0017 (0.0033)	-0.0072 (0.0125)
	4th quarter avg.	0.0032 (0.0064)	0.0108 (0.0091)*	0.0101 (0.0074)**	0.0146 (0.0110)**
Year	sample avg.	-0.0133 (0.0063)**	-0.0181 (0.0072)***	-0.0163 (0.0078)**	-0.0166 (0.0089)**
	1st quarter avg.	-0.0236 (0.0166)**	-0.0411 (0.0170)***	-0.0329 (0.0201)***	-0.0387 (0.0192)***
	4th quarter avg.	-0.0059 (0.0021)***	-0.0028 (0.0044)	-0.0055 (0.0030)	-0.0021 (0.0057)

[1] Standard errors and significance tests are based on bootstrapped results of 1,000 replications (bias-corrected and accelerated). The results show that Model (ii)'s and (iv)'s marginal effects can take different signs in the samples.

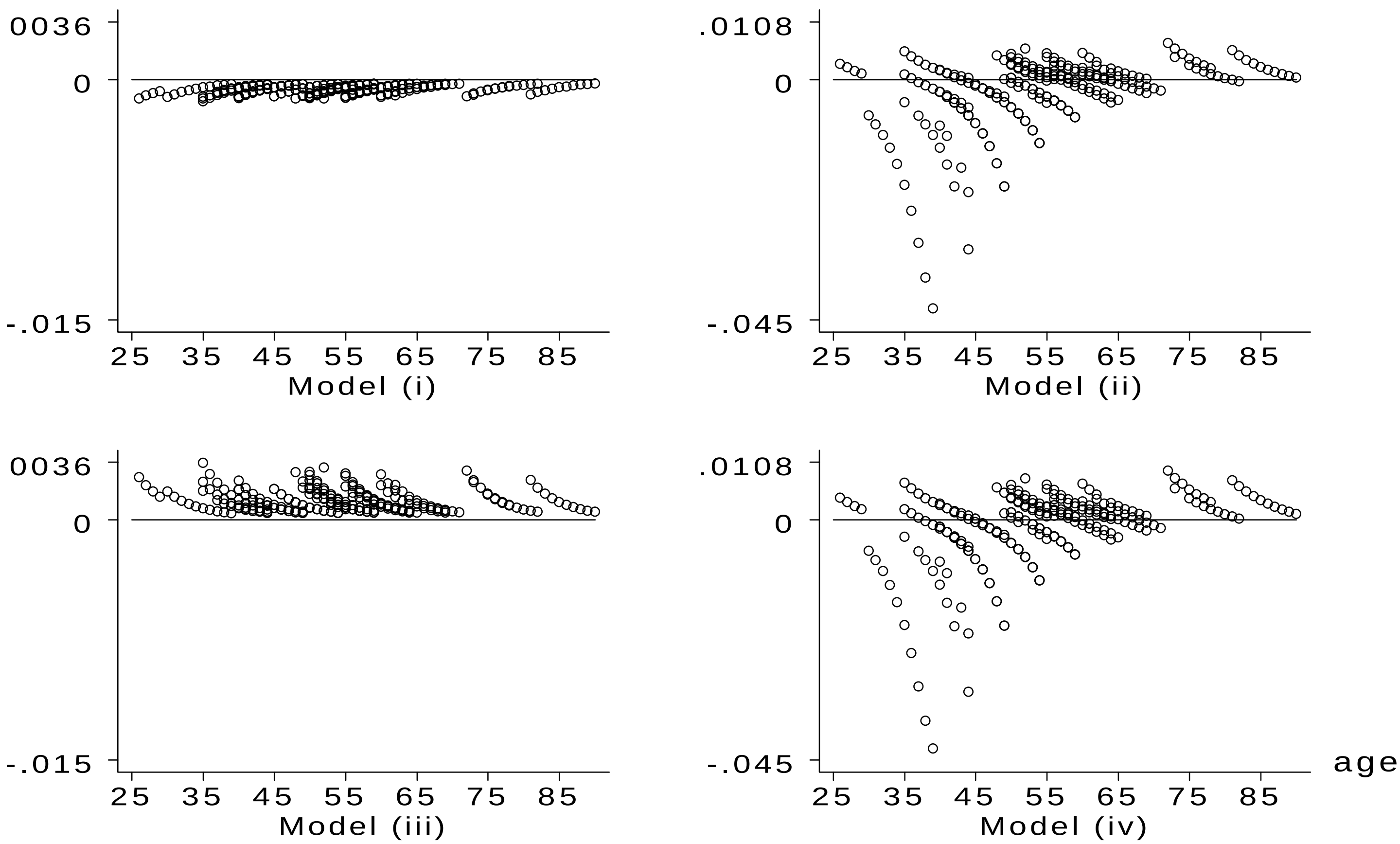

Figure 1: Marginal Effects of Age on $E(u_{it})$. The horizontal axis is Age and the vertical axis is the marginal effect. Note that the two graphs on the left have vertical scales three times smaller than the graphs on the right, so the plotted effects are visually magnified compared to the graphs on the right. The results show that Models (ii) and (iv) imply non-monotonic efficiency effects of Age. Younger farmers are more likely to show technical efficiency than older farmers. On the other hand, given a starting age, the accumulation of experience can always enhance the benefit or alleviate the impediment.

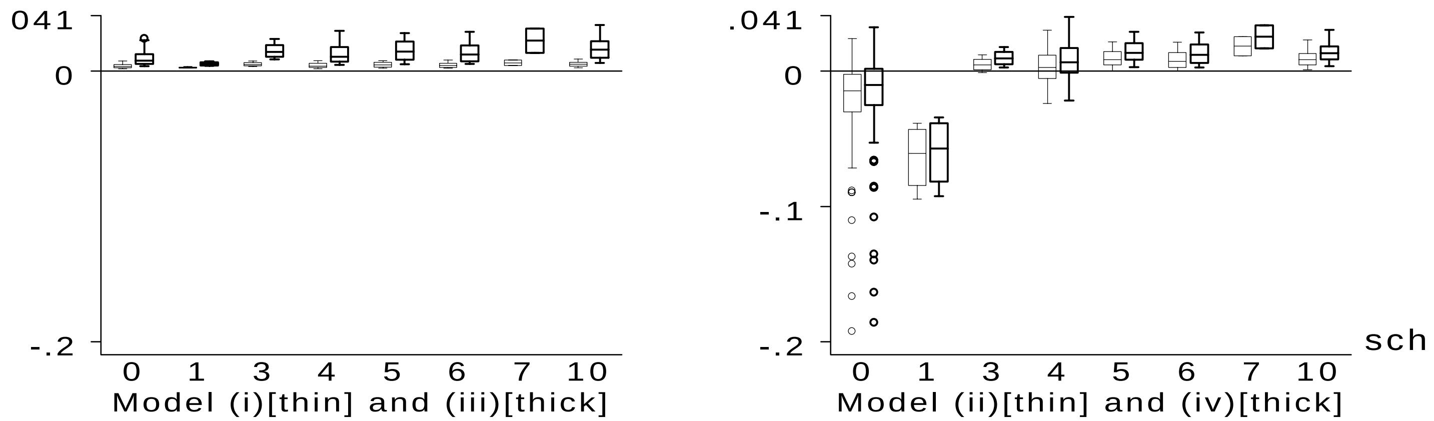

Figure 2: Marginal Effects of *Schooling* on $E(u_{it})$. The horizontal axis is *Schooling* and the vertical axis is the marginal effect. See footnote 7 for explanations of the box plots. Note that 62% of the observations have *Schooling* equal to 0 or 1, so the results of *Schooling* that are greater than 1 should be weighted more lightly. The results of Models (ii) and (iv) indicate that schooling helps reduce inefficiency when the education level is low.

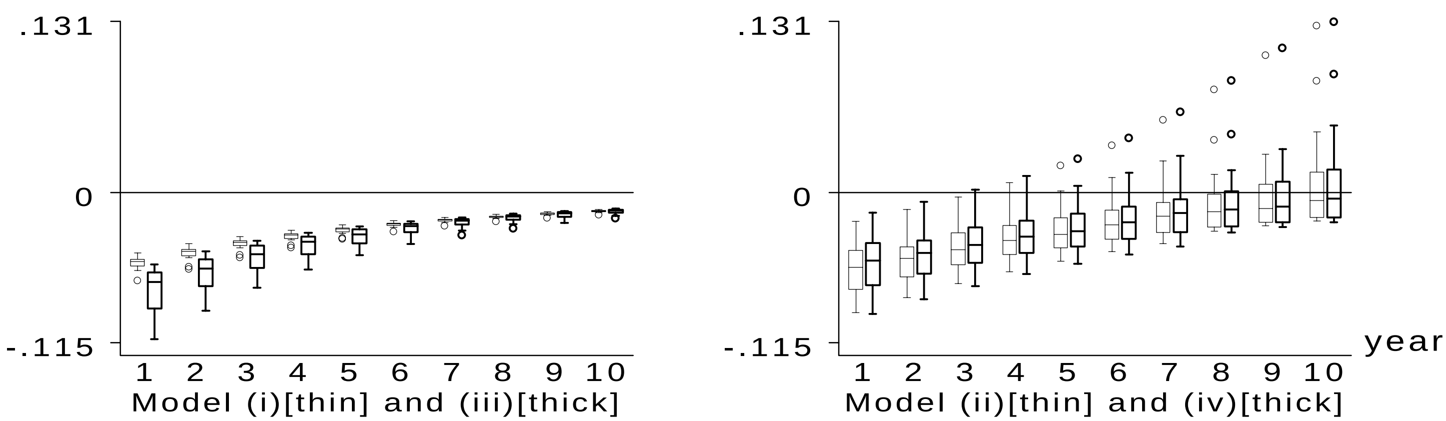

Figure 3: Marginal Effects of *Year* on $\mathrm{E}(u_{it})$. The horizontal axis is *Year* and the vertical axis is the marginal effect. The results are consistent among the four models. Efficiency improves over time, but the improvement slows down gradually.

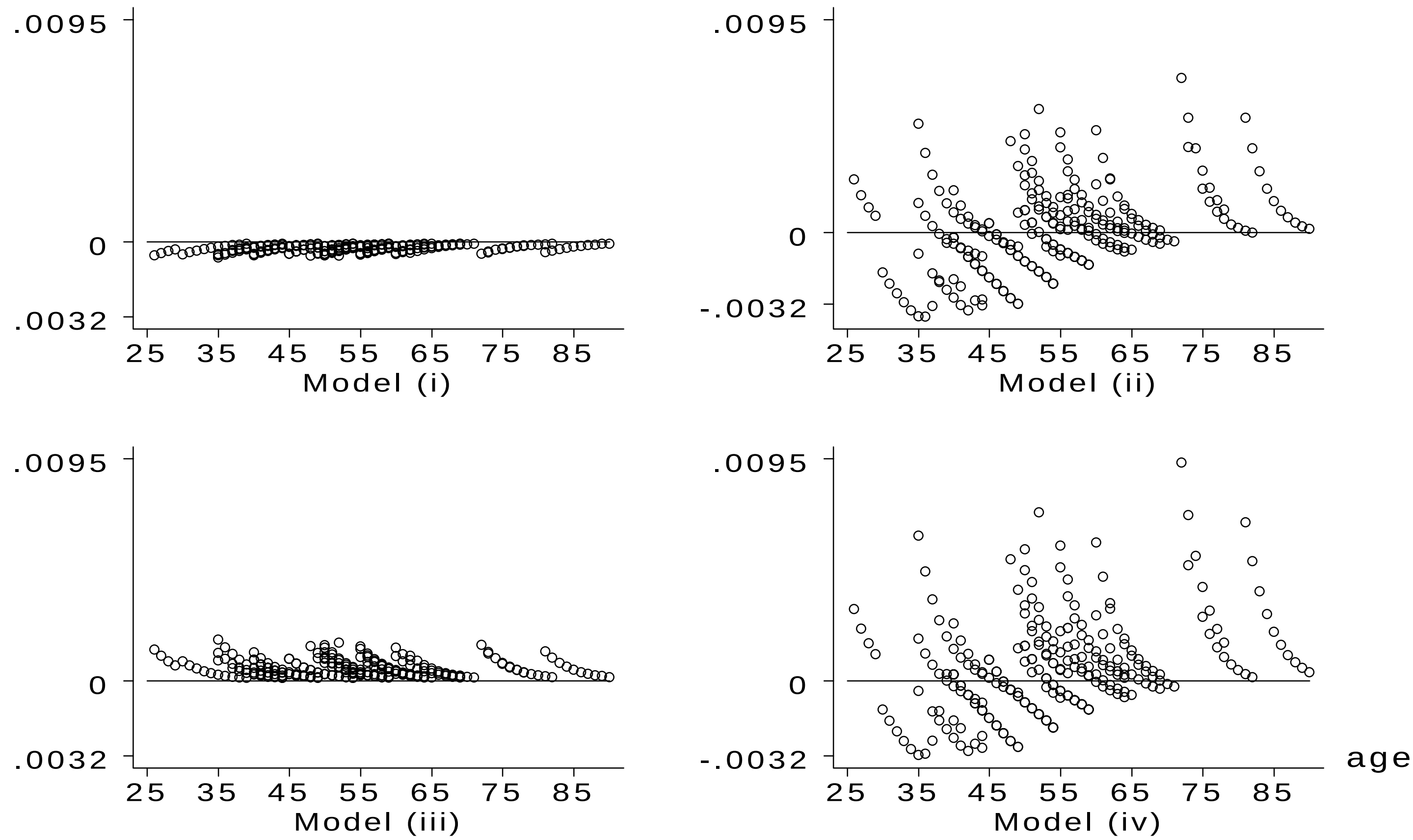

Figure 4: Marginal Effects of *Age* on $V(u_{it})$. The horizontal axis is *Age* and the vertical axis is the marginal effect. Models (ii) and (iv) imply non-monotonic marginal effects of *Age* on $\mathrm{V}(u_{it})$.

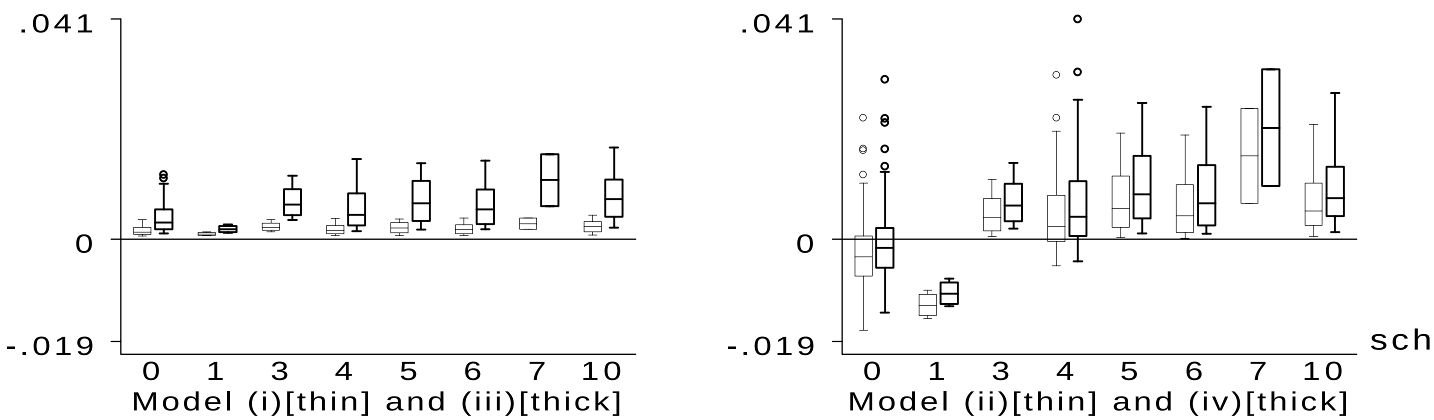

Figure 5: Marginal Effects of *Schooling* on $V(u_{it})$. The horizontal axis is *Schooling* and the vertical axis is the marginal effect.

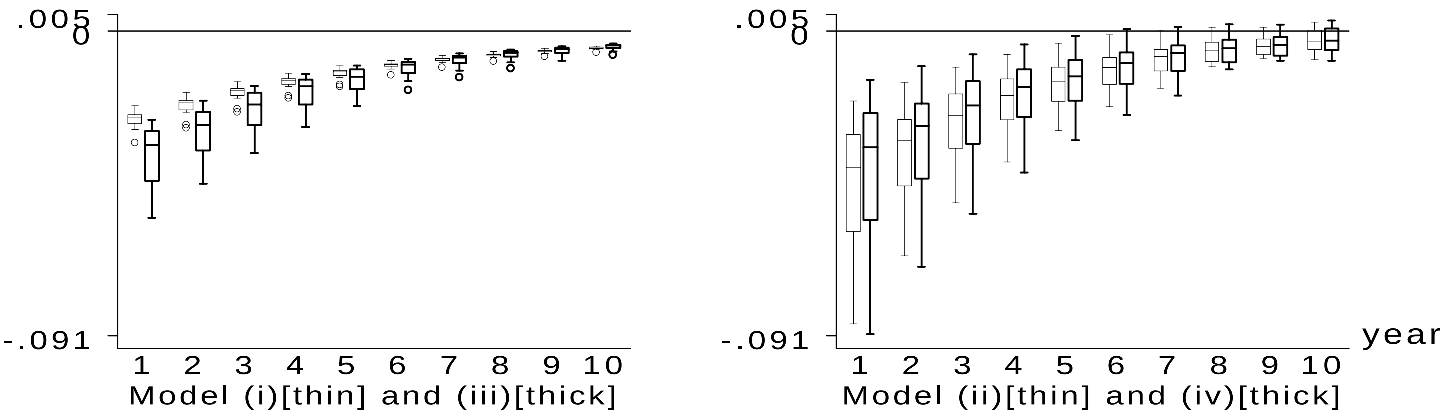

Figure 6: Marginal Effects of *Year* on $V(u_{it})$. The horizontal axis is *Year* and the vertical axis is the marginal effect. See footnote 7 for explanations of the box plots. Note that 62% of the observations have *Schooling* equal to 0 or 1, so results of *Schooling* greater than 1 should be weighted more lightly.

References

[1] Aigner, Dennis, C. A. Knox Lovell, and Peter Schmidt. (1977) "Formulation and Estimation of Stochastic Frontier Production Function Models," *Journal of Econometrics* **6**, pp. 21-37.

[2] Barrow, D. F., and A. C. Cohen. (1954) "On Some Functions Involving Mill's Ratio," *Annals of Mathematical Statistics* **25**, pp. 405-408.

[3] Battese, George. E., and Tim. J. Coelli. (1995) "A Model for Technical Inefficiency Effects in a Stochastic Frontier Production Function for Panel Data," *Empirical Economics* **20**, pp. 325-32.

[4] Bauer, Paul W. (1990) "Recent Developments in the Econometric Estimation of Stochastic Frontiers," *Journal of Econometrics* **46**, pp. 39-56.

[5] Bera, Anil K., and Subhash C. Sharma. (1999) "Estimating Production Uncertainty in Stochastic Frontier Production Function Models," *Journal of Productivity Analysis* **12**, pp. 187-210.

[6] Caudill, Steven B., and Jon M. Ford. (1993) "Biases in Frontier Estimation Due to Heteroscedasticity," *Economics Letters* **41**, pp. 17-20.

[7] Caudill, Steven B., Jon M. Ford, and Daniel M. Gropper. (1995) "Frontier Estimation and Firm-Specific Inefficiency Measures in the Presence of Heteroscedasticity," *Journal of Business and Economic Statistics* **13**, pp. 105-11.

[8] Coelli, Tim J., and George E. Battese. (1996) "Identification of Factors Which Influence the Technical Inefficiency of Indian Farmers," *Australian Journal of Agricultural Economics* **40**, pp. 103-28.

[9] Greene, William H. (1993) "The Econometric Approach to Efficiency Analysis." pp. 68-119 in *The Measurement of Productive Efficiency*, edited by Harold O. Fride, C.A. Knox Lovell, and Shelton S. Schmidt. New York Oxford: Oxford University.

[10] Hadi, Ali S. (1994) "A modification of a method for the detection of outliers in multivariate samples," *Journal of the Royal Statistical Society, Series B 56*, pp. 393-396.

[11] Hadri, Kaddour. (1999) "Estimation of a Doubly Heteroscedastic Stochastic Frontier Cost Function," *Journal of Business and Economic Statistics* **17**, pp. 359-63.

[12] Huang, Cliff J., and Jin Tan Liu. (1994) "Estimation of a Non-neutral Stochastic Frontier Production Function," *Journal of Productivity Analysis* **5**, pp. 171-80.

[13] Kumbhakar, Subal C., and C. A. Knox Lovell. (2000) Stochastic Frontier Analysis. Cambridge [England]; New York: Cambridge University Press.

[14] Kumbhakar, Subal C., Soumendra Ghosh, and J. Thomas McGuckin. (1991) "A Generalized Production Frontier Approach for Estimating Determinants of Inefficiency in U.S. Dairy Farms," *Journal of Business and Economic Statistics* **9**, pp. 279-86.

[15] Meeusen, Wim, and J. van den Broeck. (1977) "Technical Efficiency and Dimension of the Firm: Some Results on the Use of Frontier Production Functions," *Empirical Economics* **2**, pp. 109-22.