A Post-Truncation Parameterization

of Truncated Normal Technical Inefficiency

Christine Amsler

Michigan State University

Peter Schmidt

Michigan State University

Wen-Jen Tsay

Academia Sinica

April 20, 2013

Abstract

In this paper we consider a stochastic frontier model in which the distribution of technical inefficiency is truncated normal. In standard notation, technical inefficiency u is distributed as $N^+(\mu, \sigma^2)$. This distribution is affected by some environmental variables z that may or may not affect the level of the frontier but that do affect the shortfall of output from the frontier. We will distinguish the *pre-truncation* mean (μ) and variance (σ^2) from the *post-truncation* mean $\mu_* = E(u)$ and variance $\sigma_*^2 = \mathrm{var}(u)$. Existing models parameterize the pre-truncation mean and/or variance in terms of the environmental variables and some parameters. Changes in the environmental variables cause changes in the pre-truncation mean and/or variance, and imply changes in both the post-truncation mean and variance. The expressions for the changes in the post-truncation mean and variance are quite complicated. In this paper, we suggest parameterizing the post-truncation mean and variance instead. This leads to simple expressions for the effects of changes in the environmental variables on the mean and variance of u, and it allows the environmental variables to affect the mean of u only, or the variance of u only, or both.

1. Introduction

In this paper we consider the stochastic frontier model

(1) $$y_i = \alpha + x_i'\beta + v_i - u_i \ , \ \ i = 1, \dots, n \ \ ,$$

where y_i is log output, x_i is a vector of inputs or functions of inputs, v_i is random noise distributed as $N(0, \sigma_v^2)$, and $u_i \geq 0$ represents technical inefficiency. Here *i* indexes firms and *n* is the number of firms.

We assume that the distribution of u_i is truncated normal. In standard notation, u_i is distributed as $N^+(\mu_i, \sigma_i^2)$. When μ_i and σ_i^2 are constant (do not depend on *i*), this is the truncated normal model of Stevenson (1980). In subsequent models in the literature, μ_i and/or σ_i^2 depend on some "environmental variables" z_i that may not affect the level of the frontier but that do affect the size of technical inefficiency (the shortfall of output from the frontier). For example, in the RSCFG model of Reifschneider and Stevenson (1991), Caudill and Ford (1993) and Caudill, Ford and Gropper (1995), $\mu_i = 0$ and σ_i^2 is a function of z_i and some parameters. In the KGMHLBC model of Kumbhakar, Ghosh and McGuckin (1991), Huang and Liu (1994), and Battese and Coelli (1995), σ_i^2 is constant (does not depend on *i*) and μ_i is a function of z_i and parameters. In the model of Wang (2002), both μ_i and σ_i^2 depend on z_i and parameters. Finally, in the model of Alvarez et al. (2006), there is a "scaling function" $g(z_i, \theta)$ such that $\mu_i = \mu \cdot g(z_i, \theta)$ and $\sigma_i = \sigma \cdot g(z_i, \theta)$.

We now make an important distinction. We will call μ_i and σ_i^2 the *pre-truncation* mean and variance. That is, they are the mean and variance of the random variable that is truncated to get u_i. They are not the same as the *post-truncation* mean and variance, $\mu_{*,i} = E(u_i)$ and $\sigma_{*,i}^2 =$ var(u_i), which are quantities of more direct interest. (If z_i is regarded as random, then $\mu_{*,i}$ and $\sigma_{*,i}^2$

should be regarded as $E(u_i|z_i)$ and var$(u_i|z_i)$, respectively.)

The relationship between the pre-truncation and post-truncation parameters is somewhat complicated. Specifically, letting $h_i = -\mu_i/\sigma_i$,

(2A) $$\mu_{*,i} = E(u_i) = \mu_i + \sigma_i\lambda(h_i) = \sigma_i[-h_i + \lambda(h_i)]$$

(2B) $$\sigma^2_{*,i} = \text{var}(u_i) = \sigma_i^2[1 + h_i\lambda(h_i) - \lambda^2(h_i)]\ .$$

Here $\lambda(h)$ is the normal hazard function defined by $\lambda(h) = \varphi(h)/[1 - \Phi(h)]$, where φ is the standard normal p.d.f. and Φ is the standard normal c.d.f. See, e.g., Greene (2012, p. 836).

One implication of this is that the derivatives of $\mu_{*,i}$ and $\sigma^2_{*,i}$ with respect to z_i will be complicated, even when the pre-truncation mean and variance are uncomplicated functions of z_i. For example, see Wang (2002), equations (9) and (10), for the case that $\mu_i = z_i{}'\delta$ and $\sigma_i^2 = \exp\,(z_i{}'\gamma)$. The only exception is for models with the scaling property, since then $h_i = -\mu_i/\sigma_i$ does not change when z_i changes, and so the derivatives are simpler.

A more fundamental implication of these expressions is that, if a change in z_i affects either μ_i or σ_i^2 (or both), it will affect both the mean and the variance of u_i. That is, in all of the models listed above, it is impossible for z_i to affect $\sigma^2_{*,i}$ but not $\mu_{*,i}$, or vice-versa. To date, virtually all of the stochastic frontier literature has been concerned with the effects of z_i on $E(u_i)$ rather than on var(u_i). (An exception is Bera and Sharma (1999).) However, the variance of u_i may also be relevant. The enormous literature on production risk, building on the influential work of Just and Pope (e.g., Just (1975), Just and Pope (1979), Just and Pope (2003)), has emphasized the importance of risk (i.e. variance) in influencing decisions about choice of technology and choice of inputs given technology. Thus it may be important to be able to investigate the separate effects of environmental variables on the mean and variance of technical inefficiency.

In this paper, we propose a model that allows this to be done. The basic idea is simple: construct parametric models for $\mu_{*,i}$ and $\sigma_{*,i}^2$ rather than for μ_i and σ_i^2. This involves some theoretical and computational issues to be discussed below.

The plan of the paper is as follows. Section 2 describes some issues of model specification. Section 3 discusses computational issues. Section 4 gives an empirical example, and Section 5 describes the results of a small simulation. Finally, Section 6 contains our concluding remarks.

2. Model Specification

To specify an estimable model, we need to specify $\mu_{*,i}$ and $\sigma_{*,i}$ (or $\sigma_{*,i}^2$) as functions of z_i and some parameters. In generic notation, we need to specify functions μ and σ such that $\mu_{*,i} = \mu(z_i, \theta)$ and $\sigma_{*,i} = \sigma(z_i, \theta)$ for some parameters θ.

These functions are subject to some restrictions. It is obvious that we must have $\mu_{*,i} > 0$ and $\sigma_{*,i} > 0$. Less obviously, it must also be the case that

(3) $$\mu_{*,i} > \sigma_{*,i} .$$

That is, if u is the truncation from the left at zero of a normal random variable, it must be the case that $E^2(u) > \mathrm{var}(u)$; or, equivalently, the mean of u must be bigger than the standard deviation. See, for example, Horrace (2012), Lemma 1, or Bera and Sharma (1999), equation (16), or Barrow and Cohen (1954), equation (3). This restriction can be enforced through the choice of functional form or by restrictions on θ.

We will now suggest and discuss two different specifications (parameterizations).

Specification 1: $\mu_{*,i} = \exp(\delta_0 + z_i{}'\delta)$, $\sigma_{*,i} = \exp(\gamma_0 + z_i{}'\gamma)$

This specification is attractive because it satisfies the non-negativity constraints and because the interpretations of δ and γ are straightforward. Specifically, the (vector) derivatives of the mean and standard deviation of u with respect to z_i are:

(4) $$\frac{d\mu_{*,i}}{dz_i} = \mu_{*,i}\delta \ \ , \ \frac{d\sigma_{*.i}}{dz_i} = \sigma_{*,i}\gamma$$

Furthermore, this specification allows the environmental variables to affect the mean but not the variance, or vice-versa. However, the imposition of the constraint (3) is potentially troublesome. In some simple cases this constraint may be simple. For example, if there is only one variable in z_i, and if it is non-negative, then the constraint (3) will hold for all i if $\delta_0 > \gamma_0$ and $\delta > \gamma$. However, in more realistic cases we would need to restrict the parameter space in the numerical maximization of the likelihood function to the set of $\theta = \binom{\delta}{\gamma}$ such that (3) holds for all $i = 1, \dots, n$. This may be expected to cause numerical difficulties in the maximization of the likelihood and it raises issues of the statistical properties of the maximum likelihood estimates.

Specification 2: $\sigma_{*,i} = \exp(\gamma_0 + z_i{}'\gamma)$, $\mu_{*,i} = \sigma_{*,i} + \exp(\delta_0 + z_i{}'\delta)$

In this specification, the constraint (3) is enforced by the functional form and so no restrictions on θ are needed. The interpretations of δ and γ are only slightly more complicated:

(5) $$\frac{d\mu_{*,i}}{dz_i} = \mu_{*,i}\delta \ + \sigma_{*,i}(\gamma - \delta) = \sigma_{*,i}\gamma + (\mu_{*,i} - \sigma_{*,i})\delta \ \ , \ \ \frac{d\sigma_{*.i}}{dz_i} = \sigma_{*,i}\gamma$$

In this specification, it is possible for z_i to affect $\mu_{*,i}$ but not $\sigma_{*,i}$; this corresponds to $\gamma = 0$. However, it is not possible for z_i to affect $\sigma_{*,i}$ but not $\mu_{*,i}$. If z_i does not affect $\mu_{*,i}$ (for all values of i) it must be the case that $\delta = \gamma = 0$.

An interesting special case occurs when $\delta = \gamma$. In this case, the model reduces to the "scaled Stevenson model" discussed by Alvarez et al., p. 204. Specifically, when $\delta = \gamma$, we have

(6) $$\sigma_{*,i} = \sigma_0 \exp\,(z_i'\gamma) \ \ , \ \ \mu_{*,i} = \mu_0 \exp\,(z_i'\gamma)$$

where $\sigma_0 = \exp(\gamma_0)$ and $\mu_0 = \exp(\gamma_0) + \exp(\delta_0)$. More generally, the expression in equation (6) for $\sigma_{*,i}$ always holds, whether $\gamma = \delta$ or not, whereas

(7) $$\mu_{*,i} = \exp(z_i'\gamma)\,[\exp(\gamma_0) + \exp(\delta_0)\exp\big(z_i'(\delta - \gamma)\big)]\ ,$$

which reduces to the expression for $\mu_{*,i}$ in (6) when $\gamma = \delta$. So in this case the same "scaling function" $\exp(z_i{}'\gamma)$ applies to both the mean and variance of the truncated normal distribution of u_i , and the scale but not the shape of the distribution changes when z_i changes.

3. Construction of the Likelihood

We wish to construct a (log) likelihood function of the form

(8) $$\ln L = \sum_{i=1}^{n} \ln f_i(y_i - \alpha - x_i'\beta)$$

where f_i is the density of $\varepsilon_i = v_i - u_i = y_i - \alpha - x_i{}'\beta$. Note that f_i depends on *i* because it depends on z_i (because z_i affects the distribution of u_i) as well as on σ_v^2 and θ (the parameters that determine the distribution of u_i in terms of z_i).

The numerical problem that we face is that the density f_i is most naturally written in terms of the *pre-truncation* mean and variance, μ_i and σ_i. Specifically, f_i is as given by Stevenson (1980), equation (5), p. 59, if we substitute μ_i for his μ and σ_i for his σ_u (in his equation (5) and in his expressions for "σ" and "λ" in the following line) and if we change the sign of ε_i.[1] If we define $\xi_i = \sigma_i/\sigma_v$ (which replaces his "λ") and $\omega_i^2 = \sigma_i^2 + \sigma_v^2$ (which replaces his "σ^2"), this leads to the expression

(9) $$f_i(\varepsilon_i) = \omega_i^{-1}\varphi(\tfrac{\varepsilon_i+\mu_i}{\omega_i})[1 - \Phi\left(-\tfrac{\mu_i}{\omega_i\xi_i} + \tfrac{\varepsilon_i\xi_i}{\omega_i}\right)][1 - \Phi\left(-\tfrac{\mu_i}{\sigma_i}\right)]^{-1}\ .$$

---

[1] The sign change for ε is needed because he has a cost function with error $\varepsilon = v + u$ whereas we have $\varepsilon = v - u$. Changing the sign changes u to $-u$. It also changes v to $-v$, but this does not matter when v is normal with mean zero, hence symmetric.

As noted previously, the model of Wang corresponds to a particular parameterization of μ_i and σ_i. Our model corresponds to using the values of μ_i and σ_i that are implied by values of $\mu_{*,i}$ and $\sigma_{*,i}$, which in turn are implied by our chosen parameterization.

The formation of the likelihood (8) for our model as a function of θ (as well as α, β and σ_v^2) therefore involves the following logical steps. First, for a given value of θ, calculate $\mu_{*,i}$ and $\sigma_{*,i}$. This will depend on the parameterization (e.g. Specification 1) chosen. Second, for these values of $\mu_{*,i}$ and $\sigma_{*,i}$, calculate the corresponding values of μ_i and σ_i that satisfy equations (2A) and (2B). Third, use these values of μ_i and σ_i in equation (9) to calculate $f_i(\varepsilon_i)$ and insert that value in equation (8).

The only difficult step is the second step, calculating the values of μ_i and σ_i that correspond to given values of $\mu_{*,i}$ and $\sigma_{*,i}$. This amounts to inverting the functions given in (2A) and (2B). To pursue this solution, define $h_i = -\mu_i/\sigma_i$ (as above), and define the function *H(h)*:

(10) $$H(h) = \frac{\mu_*^2}{\mu_*^2+\sigma_*^2} = \frac{h^2-2h\lambda(h)+\lambda^2(h)}{1+h^2-h\lambda(h)}$$

H(h) is a monotonically decreasing function that has a limit of one as $h \rightarrow -\infty$ and a limit of 0.5 as $h \rightarrow \infty$. (The fact that H is less than one is obvious from the definition. The fact that it is greater than 0.5 is due to the constraint (3).) Therefore it has an inverse, and we can calculate

(11) $$h_i = H^{-1}\left(\frac{\mu_{*,i}^2}{\mu_{*,i}^2+\sigma_{*,i}^2}\right).$$

The value of h_i that corresponds to a specific value of $\frac{\mu_{*,i}^2}{\mu_{*,i}^2+\sigma_{*,i}^2}$ can be calculated by solving (10) numerically. Alternatively, we have constructed a large table, available on request, of values of [*h*, *H*(*h*)] for *h* between -15 and 15, with increments of 0.001. The value of *h* corresponding to a value of *H(h)* can be found by interpolating in this table.

Figure 1 is a graph of the function *H(h).* Table 1 gives a few values.

Once we have obtained h_i, we can solve equation (2B) for σ_i: $\sigma_i^2 = \frac{\sigma_{*,i}^2}{1+h_i\lambda(h_i)-\lambda^2(h_i)}$, and then $\mu_i = -h_i\sigma_i$. We then insert these values of μ_i and σ_i into the expressions for ξ_i and ω_i^2 (see the discussion preceeding equation (9)) and into equation (9) to evaluate the likelihood.

4. Empirical Example

We apply the models given above to the Philippine rice data used in the empirical examples of Coelli *et al.* (2005), chapters 8 and 9. These are annual data on 43 farmers over eight years, for a total of 344 observations. Coelli *et al.* estimate a variety of stochastic frontier models, ignoring the panel nature of the observations, which we will also do. The output variable is tons of freshly threshed rice, and the input variables are planted area in hectares (*area*), labor (*labor*), and fertilizer used in kilograms (*fert*). These variables are scaled to have unit means so the first-order coefficients of the translog function can be interpreted as elasticities of output with respect to inputs, evaluated at the sample means. Data on the age of the household head (*age*), education of the household head (*edyrs*), household size (*hhsize*), number of adults in the household (*nadult*), and the percentage of planted area classified as bantog (upland) fields (*banrat*) are used as farm characteristics (z_i) that affect the distribution of technical inefficiency. See Coelli et al., Appendix 2, for a detailed description of the data.

We specify a translog production function with a time trend, of the form

$$\text{(12)} \quad \ln y_i = \beta_0 + \theta t + \beta_1 \ln area_i + \beta_2 \ln labor_i + \beta_3 \ln fert_i$$

$$+ \beta_{11}[½(\ln area_i)^2] + \beta_{12}(\ln area_i)(\ln labor_i) + \beta_{13}(\ln area_i)(\ln fert_i)$$

$$+ \beta_{22}[½(\ln labor_i)^2] + \beta_{23}(\ln labor_i)(\ln fert_i) + \beta_{33}[½(\ln fert_i)^2] + v_i - u_i \ .$$

In all of the models we estimate, we assume that v_i is $N(0, \sigma_v^2)$. Different models will make different assumptions about the distribution of u_i. Note that we use a single subscript "*i*" because we are ignoring the panel nature of the data. For example, since we have 43 farmers, *i* = 46 actually means the second year's observation on farmer number three.

The first model we estimate is the basic stochastic frontier model in which u_i is distributed as $N^+(0, \sigma_u^2)$. Our results agree with the results of previous analyses of these data, e.g. Coelli et al., p. 250, so we will not display them here. The likelihood value achieved was -74.410.

The second model we consider is the model of Stevenson (1980), in which u_i is distributed as $N^+(\mu, \sigma_u^2)$, so that the pre-truncation mean of u_i is not necessarily zero. Because μ and σ_u^2 are constant (do not vary over *i*) this model is a special case of both the pre-truncation parameterization of Wang (2002) and our post-truncation parameterization model. We estimated the model using our post-truncation parameterization software as a check on the program. Our results agreed quite closely but not exactly with the results in Coelli et al., p. 260. We achieve a slightly higher likelihood value of -71.316 compared to their -71.64. The main substantive difference is that our estimates imply a more highly truncated normal than theirs (*h* = $-\mu/\sigma$ of about 6, rather than their value of about 2). The value of *h* is numerically not very stable, in the sense that different starting values and different ways of bounding the parameter space in MATLAB led to sometimes very extreme values of *h*. In any case the one-sided error is a heavily truncated normal and the results may be sensitive to the accuracy of the calculation of the normal hazard function. Using either of the two likelihood values above, a likelihood ratio test would reject the basic stochastic frontier model in favor of Stevenson's model.

We now turn to models in which the distribution of u_i depends on the farm characteristics z_i listed above. The parameter estimates for these models are given in Table 2.

The first model we estimated was the RSCFG Model in which

(13) $$\mu_i = 0 \, , \sigma_i^2 = \exp(\gamma_0 + z_i{}'\gamma) \, .$$

These are the pre-truncation mean and variance. The implied post-truncation mean and variance are

(14A) $$\mu_{*,i} = \sqrt{\frac{2}{\pi}}\sigma_i = \exp[\gamma_0 + \frac{1}{2}\ln\left(\frac{2}{\pi}\right) + z_i'(\frac{1}{2}\gamma)]$$

(14B) $$\sigma_{*,i}^2 = \frac{\pi-2}{\pi}\sigma_i^2 = \exp[\gamma_0 + \ln\left(\frac{\pi-2}{\pi}\right) + z_i{}'\gamma] \, .$$

So a positive coefficient in γ indicates that an increase in the corresponding variable in z_i increases mean technical inefficiency, and it also increases the variance (and standard deviation) of technical inefficiency. (And, in fact, it increases the mean and the standard deviation by the same proportion.)

The regression coefficients for this model are unremarkable (similar to those for the basic stochastic frontier model or the Stevenson model). The coefficients ($\gamma's$) of *hhsize* and *nadults* in the σ_i^2 equation are insignificant at usual significance levels, while the coefficients of *age, edyrs* and *banrat* are significant. The results indicate that technical inefficiency is higher (and more variable) on average when the farmer is older and more educated, and when the fraction of bantog (upland) fields is lower. The effect of education is perhaps surprising.[2]

[2] The standard errors for this model, and for the next two models we will discuss, were calculated using the outer product of the gradient (OPG) version of the information matrix. Our attempts to calculate standard errors from the Hessian were not numerically stable, in the sense that small changes in starting values or details of the maximization led to small changes in the parameter estimates and in the likelihood values, but to substantial changes in the Hessian and the resulting standard errors. This did not occur with the OPG estimates.

Table 3 gives the average partial effects (APE's) of the z's on μ_* and σ_*. For the RSCFG model there is not much additional information here since these APE's are proportional to γ. For example, from equation (14A), $\frac{d\mu_*}{dz} = \mu_* \cdot (\frac{1}{2}\gamma)$ and for the APE this would just be evaluated at the sample average value of μ_*. (And, for σ_*, similarly the APE would be $\frac{1}{2}\gamma$ multiplied by the average value of σ_*.) As a result a t-statistic for the significance of the APE of z_j on either μ_* or σ_* would be the same as the t-statistic for γ_j, as given in Table 2.[3]

The basic stochastic frontier model is a special case of this model (though the Stevenson model is not). We achieve a likelihood value of -65.89 for the RSCFG model (as opposed to -74.41 for the basic stochastic frontier model) and so the basic stochastic frontier model is rejected by a likelihood ratio test (chi-squared test with 5 degrees of freedom, statistic = 17.04).

The second model we consider is the model of Wang (2002) in which the pre-truncation mean (μ_i) and variance (σ_i^2) are parameterized as follows:

(15) $$\mu_i = \delta_0 + z_i{}'\delta \, , \sigma_i^2 = \exp(\gamma_0 + z_i{}'\gamma) \, .$$

Once again the parameter estimates are given in Table 2 and the APE's are given in Table 3.

We can see in Table 2 that almost none of the individual γ's or δ's is individually significant. The only individual coefficient that is significant at the 5% level is the coefficient of *edyrs* in the variance equation. However, the coefficients are jointly very significant. We achieve a likelihood value of -52.08, which is significantly larger than for the other models considered up to now. For example, this model reduces to the basic stochastic frontier model if we impose the 11 restrictions that $\delta_0 = \delta = \gamma = 0$, and this hypothesis is decisively rejected by the likelihood ratio

[3] A technical detail is that in practice we would evaluate the APE at the sample average value of $\hat{\mu}_*$ not μ_* (or $\hat{\sigma}_*$ not σ_*). However, due to the extra level of averaging, the order in probability of (average value of $\hat{\mu}_*$ minus average value of μ_*) is smaller than the order in probability of $(\hat{\gamma} - \gamma)$. This implies that we can treat the average value of $\hat{\mu}_*$ (or $\hat{\sigma}_*$) as a constant in calculating an asymptotic standard error for the APE.

test (statistic = 44.65). Similarly, the model becomes the Stevenson model if we impose the 10 restrictions that $\delta = \gamma = 0$, and this is also rejected by the likelihood ratio test (statistic = 38.46). Finally, it becomes the RSCFG model is we impose the 6 restrictions that $\delta_0 = \delta = 0$, and this hypothesis is rejected by the likelihood ratio test (statistic = 27.61).

The individual coefficients in the Wang model are hard to interpret because they indicate the effects of the *z*'s on the pre-truncation mean and variance, and a change in either the pre-truncation mean or the pre-truncation variance will affect both the post-truncation mean and the post-truncation variance. The average partial effects in Table 3 are therefore easier to interpret because they give the effects of the z's on the post-truncation mean and the standard deviation of *u*,[4] and these are the natural objects of interest.

The average partial effects of the *z*'s on the mean of *u* are of the same sign as in the RSCFG model except for *edyrs*. We no longer have the surprising result that education raises average inefficiency. The magnitudes of some of the partial effects are noticeably different, however. For the average partial effects of the *z*'s on the standard deviation of inefficiency, once again the sign is the same as in the RSCFG model except for *edyrs*, but the magnitudes of the partial effects are sometimes quite different.

Now we turn to the model of this paper, in which the post-truncation mean ($\mu_{*,i}$) and standard deviation ($\sigma_{*,i}$) are parameterized. In Section 2 above we considered two different specifications. In Specification 1, we have $\mu_{*,i}$ = exp($\delta_0 + z_i{}'\delta$) and $\sigma_{*,i}$ = exp($\gamma_0 + z_i{}'\gamma$). Unfortunately we were unable to estimate this model satisfactorily on this data set. We could estimate certain simplified versions of the model but not the full model. There was a lot of

[4] The partial effects of the z_j on μ_* and σ_*^2 are given by Wang (2002), pp. 244-245. To calculate the partial effect on the standard deviation σ_* we note that $\frac{d\sigma_*^2}{dz} = 2\sigma_* \frac{d\sigma_*}{dz}$ and therefore $\frac{d\sigma_*}{dz} = \frac{1}{2\sigma_*}\frac{d\sigma_*^2}{dz}$.

numerical instability and the algorithm would not converge. This is presumably because we need to impose the restriction that $\mu_{*,i} > \sigma_{*,i}$ for all i, which is equivalent to $\delta_0 + z_i'\delta > \gamma_0 + z_i'\gamma$ for all i. This is a very large set of restrictions, and the restricted parameter set is not compact. We therefore gave up on this parameterization.

We were able to estimate Specification 2 successfully. In this specification we have $\sigma_{*,i} = \exp(\gamma_0 + z_i'\gamma)$ and $\mu_{*,i} = \sigma_{*,i} + \exp(\delta_0 + z_i'\delta) = \exp(\gamma_0 + z_i'\gamma) + \exp(\delta_0 + z_i'\delta)$. Therefore the γ's determine the effect of the z's on the standard deviation of u, and also part of the effect of the z's on the mean of u. The δ's determine the effect of the z's on the difference between the mean and the standard deviation of u.

The parameter estimates are given in Table 2. The estimates of the β's and of σ_v are once again unremarkable. The estimates of the γ's and δ's are not directly comparable to the estimates from the Wang model (the two models do not contain the same parameters) but the level of individual significance of these parameters is comparable to what we had for the Wang model. To make substantive comparisons, it is probably best to compare the average partial effects of the environmental variables on the mean and standard deviation of inefficiency in the various models, since the interpretation of the partial effects does not depend on the model. For Specification 2 of our model, $\frac{d\sigma_*}{dz} = \sigma_* \cdot \gamma$ and $\frac{d\mu_*}{dz} = \mu_* \cdot \delta + \sigma_* \cdot (\gamma - \delta) = (\mu_* - \sigma_*) \cdot \delta + \sigma_* \cdot \gamma$. For the estimated APE's these would be evaluated at the sample average values of $\hat{\mu}_*$ and $\hat{\sigma}_*$, and at $\hat{\gamma}$ and $\hat{\delta}$. Since the estimated APE's are linear in $\hat{\gamma}$ *and* $\hat{\delta}$,[5] their asymptotic standard errors are easy to calculate.

The average partial effects from Specification 2 seem very reasonable, although admittedly

[5] For the same reason given in footnote 3, we can treat the average values of $\hat{\mu}_*$ and $\hat{\sigma}_*$ as constants in calculating the asymptotic standard errors.

many of them are not significantly different from zero. Most of the APE's are similar to those from the other two models. They are generally more similar to the results from the RSCFG model than to the results from the Wang model. The main differences across models are in the effects of education of the farmer (*edyrs*). In our model increasing the education of the farmer increases both the mean and standard deviation of inefficiency, which agrees with the results from the RSCFG model but not the results from Wang's model. However, the effect of education of the farmer on the mean of inefficiency is much smaller in our model than in the RSCFG model, and in fact is insignificantly different from zero. The main effect of education of the farmer is now on the variability (standard deviation) of inefficiency, and this effect is statistically significant. This set of results illustrates the potential of our model to distinguish effects of variables on the mean versus the variance of inefficiency.

We achieve a likelihood value of -58.56. Our model nests the Stevenson model, which corresponds to the 10 restrictions that $\delta = \gamma = 0$, and the Stevenson model is rejected by the likelihood ratio test (statistic = 25.51). Our model also nests the basic stochastic frontier model, which corresponds to the 10 restrictions just given plus the one additional restriction that $\delta_0 = \gamma_0 + \ln\left(\sqrt{\frac{2}{\pi-2}} - 1\right)$. The basic stochastic frontier model is rejected by the likelihood ratio test (statistic = 31.70). Finally, our model nests the RSCFG model, which corresponds to the six restrictions that $\gamma = \delta$ and that $\delta_0 = \gamma_0 + \ln\left(\sqrt{\frac{2}{\pi-2}} - 1\right)$. The RSCFG model is rejected at the 5% level (statistic = 14.66, critical value = 12.59) but not as decisively as the other simpler models were rejected.

Our model does not nest Wang's model, or vice-versa. Our likelihood value of -58.56 is noticeably smaller than the likelihood value of -52.08 for the Wang model, so that our model clearly does not fit the data as well as the Wang model. However, because these are not nested models, we cannot say that this difference in likelihood values is statistically significant.

5. Simulations

In this section we report the results of a small simulation study. The point of the study is to see whether statistically reliable estimates can be achieved in either or both of our two specifications, in a very simple model.

The model we consider is of the form

(16) $$y_i = \beta x_i + v_i + u_i \ , \ i = 1,\ldots,n \ ,$$

where all symbols are scalars.[6] The x_i are iid standard normal and the v_i are iid $N(0, \sigma_v^2)$, where we consider σ_v = 0.3, 0.5 and 1.0. We consider three sample sizes, n = 200, 400 and 800. In all of our experiments we choose β = 0 (but it is estimated, so the true value of β is inconsequential).

We consider both specifications described in Section 2. The DGP is exactly the same in both cases but the model estimated is different. For Specification 1 we have $\mu_{*,i} = \exp(\delta_0 + \delta_1 z_i)$ and $\sigma_{*,i} = \exp(\gamma_0 + \gamma_1 z_i)$ where the z_i are iid standard normal. We generate the u_i as $N^+(0,1)$, which corresponds to $\delta_1 = \gamma_1 = 0$, $\delta_0 =$ -0.2258 and γ_0 = -0.5063. For Specification 2 we have $\sigma_{*,i} = \exp(\gamma_0 + \gamma_1 z_i)$ and $\mu_{*,i} = \sigma_{*,i} + \exp(\delta_0 + \delta_1 z_i)$. Again the z_i are iid standard normal and we generate the u_i as $N^+(0,1)$. In this specification this corresponds to $\delta_1 = \gamma_1 = 0$, $\delta_0 =$ -1.6340 and γ_0 = -0.5063.

Note that, in terms of the standard Aigner, Lovell and Schmidt (1977) notation, we have

[6] The fact that we have $v + u$ instead of $v - u$ is inconsequential. There are just a few sign changes in the likelihood.

$\sigma_u = 1$ and therefore $\lambda = \sigma_u/\sigma_v$ equals 3.33, 2 and 1 for σ_v = 0.3, 0.5 and 1.0, respectively.

We use 200 replications in our experiments, which is enough to answer the question posed above. We report the bias and RMSE of the estimates of each of the parameters.

Table 4 reports the results for Specification 1. The top panel shows the bias of the estimates. These numbers are not very encouraging. We generally estimate δ_1, γ_1 and σ_v reasonably well, but for the other parameters the bias is often large. Bias is often not too serious when σ_v is small and/or *n* is large, but for empirically relevant parameter values like σ_v = 1 (which corresponds to λ = 1) and *n* = 200 the bias is generally substantial.

The bottom panel in Table 4 reports the RMSE of the estimates. The conclusions are broadly similar to those for bias. The results are generally reasonably good when σ_v is small and/or *n* is large, but not otherwise.

Having reported these results, we must admit that we do not entirely trust them. There was a lot of numerical instability in the calculation of the estimates. Slightly different starting values or values of the tuning parameters in the maximization algorithm led to unreasonably large differences in the results. Also there were lots of outliers, and these drove many of the strange results reported in Table 4. The propensity of Specification 1 to generate occasional outliers in the results is empirically relevant, and so we chose not to trim such outliers. Our main conclusion from the simulations is that Specification 1, while attractive in principle, is unlikely to be empirically useful.

Table 5 reports the results for Specification 2. For this specification we did not encounter the numerical problems or large outliers that we encountered with Specification 1. The results are certainly better than they were for Specification 1, although comparisons are complicated by the fact that δ_0 and δ_1 are not the same parameters in the two specifications. In terms of bias (top panel), we now estimate four parameters (δ_1, γ_0, γ_1 and σ_v) reasonably well, but there are substantial biases for β and δ_0 except when σ_v is small and/or *n* is large. In terms of RMSE (bottom panel), there are fewer strange values than there were in Table 4. For this specification, σ_v is more important than *n* in determining the performance of the MLE, and we still have the problem that the results are not good when σ_v is large.

6. Concluding Remarks

In this paper we have considered stochastic frontier models in which the distribution of technical inefficiency is truncated normal. That is, in standard notation we have *u* distributed as $N^+(\mu, \sigma^2)$. We call μ and σ^2 the pre-truncation mean and variance. These can be distinguished from the actual ("post-truncation") mean and variance, $\mu_* = E(u)$ and $\sigma_*^2 = \text{var}(u)$. Previous models in the literature, notably Wang (2002), let μ and σ^2 depend on environmental variables (z) and parameters (θ). In this paper, we choose instead to parameterize the post-truncation mean and variance, μ_* and σ_*^2.

The main advantage of Wang's model is that it is easier to estimate. The likelihood is most naturally written in terms of the pre-truncation mean and variance. This simplifies programming and makes it more likely that the calculations will be numerically stable.

The advantage of our model is that the parameters are easier to interpret. The post-truncation mean and variance of *u* are the items of economic interest. In our model it is much easier than in Wang's model to separate the effect of an environmental variable on the mean of *u* from its effect on the variance of *u*, and either or both of these may be of interest. Also, because the interpretation of the parameters is clearer, it may be clearer what is and what is not a reasonable parameterization.

One of the main motivations of Wang's model is that it allowed a non-monotonic effect of environmental variables on the mean of *u*. Our model can also allow that, if we choose a non-monotonic function in our parameterization.

FIGURE 1

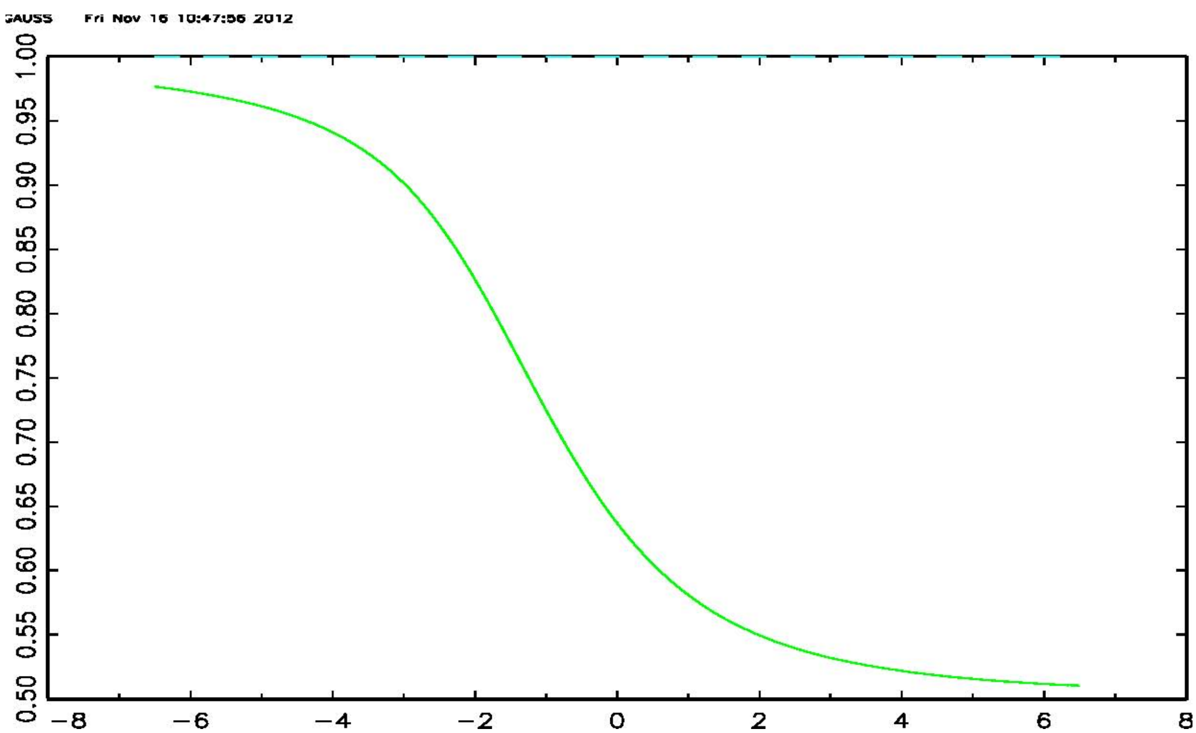

TABLE 1

Values of *h* and *H(h)*

h	*H(h)*		*h*	*H(h)*
-15	0.995575		1	0.580728
-14	0.994924		2	0.549317
-13	0.994118		3	0.531803
-12	0.993103		4	0.521655
-11	0.991803		5	0.515466
-10	0.990099		6	0.511497
-9	0.987805		7	0.508834
-8	0.984615		8	0.506975
-7	0.980000		9	0.505634
-6	0.972973		10	0.504639
-5	0.961539		11	0.503881
-4	0.941210		12	0.503293
-3	0.901465		13	0.502827
-2	0.826543		14	0.502452
-1	0.724730		15	0.502147
0	0.636620			

TABLE 2

Estimates of Models with Determinants of Inefficiency

		RSCFG MODEL			WANG MODEL			SPECIFICATION 2	
Variable		Coeff	|t-value|		Coeff	|t-value|		Coeff	|t-value|
β_0 (constant)		0.246	5.50		0.171	3.69		0.190	3.70
θ (time trend)		0.017	2.60		0.016	2.49		0.019	2.88
β_1 (area)		0.575	6.43		0.564	6.44		0.560	6.57
β_2 (labor)		0.215	2.46		0.232	2.78		0.240	2.85
β_3 (fert)		0.192	3.83		0.185	3.96		0.187	3.86
β_{11}		-0.352	1.39		-0.394	1.57		-0.305	1.25
β_{12}		0.577	2.87		0.528	2.49		0.551	2.91
β_{13}		0.011	0.07		0.060	0.43		-0.027	0.19
β_{22}		-0.601	2.21		-0.459	1.67		-0.550	2.12
β_{23}		-0.114	0.78		-0.133	0.99		-0.096	0.68
β_{33}		0.014	0.15		-0.013	0.16		0.035	0.38
σ_v		0.171	8.34		0.187	10.43		0.178	8.50
γ_0 (constant)		-3.053	3.27		2.022	1.34		-2.294	3.17
γ_1 (age)		0.028	2.06		0.008	0.41		0.013	1.11
γ_2 (edyrs)		0.112	3.36		-0.386	2.38		0.061	2.48
γ_3 (hhsize)		0.085	0.99		0.132	1.02		0.075	1.04
γ_4 (nadult)		-0.104	0.98		-0.099	0.61		-0.054	0.68
γ_5 (banrat)		-1.298	2.98		-0.687	1.13		-0.529	1.48
δ_0 (constant)					-7.940	0.95		-1.595	0.51
δ_1 (age)					0.034	0.59		0.044	1.06
δ_2 (edyrs)					0.685	1.02		-0.175	0.46
δ_3 (hhsize)					-0.008	0.04		-0.121	0.38
δ_4 (nadult)					-0.080	0.24		-0.084	0.39
δ_5 (banrat)					-0.844	0.60		-2.410	1.40
ln likelihood		-65.89			-52.08			-58.56	

TABLE 3

Average Partial Effects

Effects on mean of *u*

		RSCFG		WANG		SPEC. 2	
Variable		APE		APE		APE	Std. err.
age		0.0046		0.0053		0.0052	0.0028
edyrs		0.0186		-0.0054		0.0055	0.0206
hhsize		0.0140		0.0261		0.0117	0.0162
nadult		-0.0172		-0.0289		-0.0171	0.0188
banrat		-0.2150		-0.2316		-0.2499	0.1203

Effects on standard deviation of *u*

		RSCFG		WANG		SPEC. 2	
Variable		APE		APE		APE	Std. err.
age		0.0035		0.0012		0.0030	0.0027
edyrs		0.0140		-0.0069		0.0145	0.0058
hhsize		0.0106		0.0073		0.0179	0.0172
nadult		-0.0130		-0.0074		-0.0128	0.0189
banrat		-0.1625		-0.0575		-0.1261	0.0850

TABLE 4

Bias of MLE of Post-Truncation Parameterization Model: Specification 1

n	σ_v		$\hat{\beta}$	$\hat{\delta}_0$	$\hat{\delta}_1$	$\hat{\gamma}_0$	$\hat{\gamma}_1$	$\hat{\sigma}_v$
200	0.3		-0.0955	0.0954	0.0041	-0.2344	-0.0026	-0.0534
400	0.3		-0.0435	0.0465	0.0022	-0.1703	0.0046	-0.0230
800	0.3		-0.0304	0.0320	-0.0028	-0.7049	-0.0036	-0.0130
200	0.5		-0.2522	0.2369	0.0062	-0.8523	0.2206	-0.1249
400	0.5		-0.1585	0.1585	0.0029	-0.0641	0.0076	-0.0672
800	0.5		-0.1215	0.1106	-0.0039	-0.0712	0.0040	-0.0503
200	1.0		-0.7494	0.3866	0.0198	-2.5659	-0.1612	-0.2022
400	1.0		-0.7710	0.5198	-0.0313	-0.3749	0.0324	-0.2131
800	1.0		-0.5986	0.4697	0.0021	0.0807	-0.0047	-0.1700

RMSE of MLE of Post-Truncation Parameterization Model: Specification 1

n	σ_v		$\hat{\beta}$	$\hat{\delta}_0$	$\hat{\delta}_1$	$\hat{\gamma}_0$	$\hat{\gamma}_1$	$\hat{\sigma}_v$
200	0.3		0.1643	0.1865	0.0531	3.3636	0.1126	0.0987
400	0.3		0.1064	0.1341	0.0379	1.7701	0.0581	0.0743
800	0.3		0.0931	0.1197	0.0300	5.8868	0.0611	0.0714
200	0.5		0.3646	0.3863	0.0861	9.3163	2.9910	0.1819
400	0.5		0.2644	0.2719	0.0412	1.7215	0.0487	0.1210
800	0.5		0.2301	0.2926	0.0356	1.3463	0.0760	0.0992
200	1.0		1.1682	1.0775	0.3724	13.382	3.1643	0.3384
400	1.0		1.0434	0.9461	0.2452	4.0405	0.3904	0.3315
800	1.0		0.8375	0.6480	0.0707	1.1404	0.0988	0.2735

TABLE 5

Bias of MLE of Post-Truncation Parameterization Model: Specification 2

n	σ_v		$\hat{\beta}$	$\hat{\delta}_0$	$\hat{\delta}_1$	$\hat{\gamma}_0$	$\hat{\gamma}_1$	$\hat{\sigma}_v$
200	0.3		-0.0782	0.1869	0.0006	0.0047	0.0016	-0.0527
400	0.3		-0.0391	0.0793	0.0084	0.0086	0.0000	-0.0253
800	0.3		-0.0261	0.0307	-0.0218	0.0082	0.0004	-0.0169
200	0.5		-0.1906	0.3331	-0.0025	0.0278	0.0118	-0.0999
400	0.5		-0.1174	0.2147	0.0014	0.0328	0.0012	-0.0577
800	0.5		-0.1068	0.2032	-0.0062	0.0377	0.0020	-0.0461
200	1.0		-0.6399	0.8744	0.0272	-0.0281	-0.0244	-0.2195
400	1.0		-0.6218	0.8780	0.0040	0.0861	0.0165	-0.1906
800	1.0		-0.5040	0.6872	-0.0293	0.0811	0.0445	-0.1627

RMSE of MLE of Post-Truncation Parameterization Model: Specification 2

n	σ_v		$\hat{\beta}$	$\hat{\delta}_0$	$\hat{\delta}_1$	$\hat{\gamma}_0$	$\hat{\gamma}_1$	$\hat{\sigma}_v$
200	0.3		0.1643	0.5386	0.1858	0.0916	0.0701	0.0979
400	0.3		0.1045	0.4083	0.1475	0.0616	0.0472	0.0630
800	0.3		0.0858	0.3471	0.1275	0.0466	0.0346	0.0468
200	0.5		0.3277	0.7630	0.2569	0.5170	0.1179	0.1665
400	0.5		0.2450	0.6297	0.1924	0.1681	0.0660	0.1139
800	0.5		0.2165	0.5797	0.1455	0.0990	0.0399	0.0920
200	1.0		0.9339	1.2477	0.2466	1.3860	0.6133	0.3342
400	1.0		0.9059	1.2656	0.2137	0.6769	0.2572	0.3022
800	1.0		0.7895	1.1142	0.1797	1.0746	0.3800	0.2717

REFERENCES

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

Alvarez, A., C. Amsler, L. Orea and P. Schmidt (2006), “Interpreting and Testing the Scaling Property in Models Where Inefficiency Depends on Firm Characteristics,” *Journal of Productivity Analysis*, 25, 201-212.

Barrow, D.F. and A.C. Cohen Jr. (1954), “On Some Functions Involving the Mills Ratio,” *Annals of Mathematical Statistics*, 25, 405-408.

Battese, G.E. and T.J. Coelli (1995), "A Model for Technical Inefficiency Effects in a Stochastic Frontier Production Function for Panel Data," *Empirical Economics,* 20, 325-332.

Bera, A.K. and S.C. Sharma (1999), “Estimating Production Uncertainty in Stochastic Frontier Production Function Models,” *Journal of Productivity Analysis*, 12, 187-210.

Caudill, S.B. and J.M. Ford (1993), "Biases in Frontier Estimation Due to Heteroskedasticity," *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 Heteroskedasticity," *Journal of Business and Economic Statistics*, 13, 105-111.

Coelli, T.J., D.P. Rao, C.J. O’Donnell and G.E. Battese (2005), *An Introduction to Efficiency and Productivity Analysis,* 2nd edition, Springer.

Greene, W.H. (2012), *Econometric Analysis*, 7th edition, Pearson.

Horrace, W.C. (2012), “Moments of the Singly Truncated Normal Distribution,” unpublished manuscript, Syracuse University.

Huang, C.J. and J.-T. Liu (1994), "Estimation of a Non-Neutral Stochastic Frontier Production Function," *Journal of Productivity Analysis*, 5, 171-180.

Just, R.E. (1975), “Risk Aversion under Profit Maximization,” *American Journal of Agricultural Economics*, 57, 347-352.

Just, R.E. and R.D. Pope (1979), “Production Function Estimation and Related Risk Considerations,” *American Journal of Agricultural Economics*, 61, 276-284.

Just, R.E. and R.D. Pope (2003), “Agricultural Risk Analysis: Adequacy of Models, Data, and Issues,” *American Journal of Agricultural Economics*, 85, 1249-1256.

Kumbhakar, S.C., S. Ghosh and J.T. McGuckin (1991), "A Generalized Production Frontier Approach for Estimating Determinants of Inefficiency in U.S. Dairy Farms," *Journal of Business and Economic Statistics*, 9, 279-286.

Reifschneider, D. and R. Stevenson (1991), "Systematic Departures from the Frontier: A Framework for the Analysis of Firm Inefficiency," *International Economic Review*, 32, 715-723.

Stevenson, R.E. (1980), “Likelihood Functions for Generalized Stochastic Frontier Estimation,” *Journal of Econometrics*, 13, 57-66.

Wang, H.-J. (2002), “Heteroscedasticity and Non-Monotonic Efficiency Effects of a Stochastic Frontier Model,” *Journal of Productivity Analysis*, 18, 241-253.