One-step and two-step estimation of the effects of exogenous variables on technical efficiency levels

Wang, Hung-jen and Schmidt, Peter

October 2001

ONE-STEP AND TWO-STEP ESTIMATION OF THE EFFECTS OF EXOGENOUS VARIABLES ON TECHNICAL EFFICIENCY LEVELS

Hung-jen Wang
Academia Sinica

Peter Schmidt
Michigan State University

October, 2001

Revised March, 2002

REFERENCES

Battese, G.E. and T.J. Coelli (1988), "Prediction of Firm-level Technical Efficiencies with a Generalized Frontier Production Function and Panel Data," *Journal of Econometrics*, 38, 387-399

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.

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.

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

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.

Kumbhakar, S.C. and C.A.K. Lovell (2000), *Stochastic Frontier Analysis*, Cambridge University Press.

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.

Ritter, C. and L. Simar (1997), "Pitfalls of Normal-Gamma Stochastic Frontier Models," *Journal of Productivity Analysis*, 8, 167-182.

Simar, L., C.A.K. Lovell and P. Vanden Eeckaut (1994), "Stochastic Frontiers Incorporating Exogenous Influences on Efficiency," Discussion Papers No. 9403, Institut de Statistique, Université Catholique de Louvain.

1. INTRODUCTION

In this paper we are concerned with estimating the effects of exogenous variables on firms' levels of technical efficiency. To analyze this problem, we assume a standard stochastic frontier model in which the distribution of technical inefficiency may depend on exogenous variables. To be more specific, let y equal output (say, in logarithms); let x be a set of inputs; and let z be a set of exogenous variables that affect technical efficiency. The production frontier specifies maximal output as a function of x, plus a random (normal) error, and then actual output equals maximal output minus a one-sided error term whose distribution depends on z.

Many empirical analyses have proceeded in two steps. In the first step, one estimates the stochastic frontier model and the firms' efficiency levels, ignoring z. In the second step, one tries to see how efficiency levels vary with z, perhaps by regressing a measure of efficiency on z. It has long been recognized that such a two-step procedure will give biased results, because the model estimated at the first step is misspecified. The solution to this bias problem is a one-step procedure based on the correctly specified model for the distribution of y given x and z. In the one-step procedure the assumed relationship between z and technical efficiency is imposed in estimating the technology and the firms' efficiency levels, not just at the last stage of the exercise.

Although it is widely recognized that two-step procedures are biased, there seems to be little evidence on the severity of this bias. For example, Caudill and Ford (1993) provide evidence on the bias of the estimated technological parameters, but not on the efficiency levels themselves or their relationship to the explanatory variables z. The main contribution of this paper is to provide extensive Monte Carlo evidence of the bias of the two-step procedure. We find serious bias at all stages of this procedure. The size of the bias is very substantial and should argue strongly against two-step procedures.

We also provide some new theoretical insights into the bias problem. It is widely appreciated that the severity of the bias in estimation of the technological parameters (coefficients of x) depends on the magnitude of the correlation between x and z. However, we also explain why, if the dependence of inefficiency on z is ignored, the estimated firm-level efficiencies are spuriously underdispersed. As a result the second-step regression understates the effect of z on efficiency levels. Importantly, this is true whether or not x and z are correlated. Our simulations strongly confirm the relevance of this observation, since the two-step estimates of the effect of z on efficiency levels are seriously biased downward in all cases.

The paper also provides some arguments in favor of models that have the "scaling property" that, conditional on z, the one-sided (technical inefficiency) error term equals some function of z times a one-sided error distributed independently of z. Some but not all of the models in the literature have this property. We explain why this is a convenient and (to us) intuitively plausible property for a one-step model to have.

2. THEORETICAL DISCUSSION

2a. Basic Framework

As above, let y be log output. (We will not specify observational subscripts, for simplicity, but the discussion applies to either cross-sectional or panel data.) We let x be a vector of variables that affect the frontier (maximal) level of output, and z be a set of variables that affect the deviation of output from the frontier (technical inefficiency). We note that x and z may overlap. For example, the position of the frontier may depend on things other than those typically thought of as inputs, and the inputs may be among the factors that also affect technical efficiency. Our statistical model will specify a distribution for y conditional on x and z. Thus we treat x and z as "given" or "fixed," and as always this corresponds to an assertion of exogeneity (lack of feedback from y to x and z). One important implication of this viewpoint, which is sometimes missed, is that the variables z that determine inefficiency must not be functions of y. For example, if one variable in z is a measure of firm size, the size of the firm should be defined in terms of levels of inputs, not output.

Let $y^* \geq y$ be the unobserved "frontier". Then the linear stochastic frontier model asserts that, conditional on x and z, y^* is distributed as $N(x'\beta, \sigma^2)$. (The word "linear" refers to the fact that $E(y^*|x,z) = x'\beta$, which is linear in x.) This is consistent with the usual regression representation with an explicit error term v:

(1) $$y^* = x'\beta + v$$

where v is $N(0,\sigma_v^2)$ and is independent of x and z. The stochastic frontier model is completed by the assertion that, conditional on x, z and y^*, the actual output level y equals y^* *minus* a one-sided error whose distribution depends on z and perhaps some additional parameters δ. This is consistent with the composed-error representation:

(2) $$y = x'\beta + v - u(z,\delta) \quad , \quad u(z,\delta) \geq 0 \quad ,$$

where v is $N(0,\sigma_v^2)$ and is independent of x, z and u.

2b. Alternative Models and the Scaling Property

In the framework given above, different models correspond to different specifications for $u(z,\delta)$; that is, for the distribution of the technical inefficiency error term and the way that it depends on z. A common way to specify a model is to specify a distribution for u and then to allow the parameter(s) of that distribution to depend on z (and possibly other parameters δ). For example, suppose that u has a half-normal distribution, which we will denote by $N(0,\sigma_u^2)^+$, where here and elsewhere in this paper the superscript "+" refers to truncation on the left at zero. Then we can assume that the parameter σ_u (or σ_u^2) is a specified function of z, say $\sigma_u(z,\delta)$, so that u is distributed as

$N(0,\sigma_u(z,\delta)^2)^+$. This model has been considered by Reifschneider and Stevenson (1991), Caudill and Ford (1993) and Caudill, Ford and Gropper (1995). They consider different functional forms for $\sigma_u(z,\delta)$. For example, Claudill, Ford and Gropper specify that $\sigma_u(z,\delta) = \sigma \exp(z'\delta)$.[1]

Alternatively, we can model the dependence of $u(z,\delta)$ on z by writing it as

(3) $$u(z,\delta) = h(z,\delta)u^* \quad ,$$

where $h(z,\delta) \geq 0$ and where u^* has a distribution that does not depend on z. We will refer to the condition given in (3) as the *scaling property*. Then $h(z,\delta)$ will be called the *scaling function* and the distribution of u^* will be called the *basic distribution*. For example, this paper's simulations will be based on the model in which the scaling function is $h(z,\delta) = \exp(z'\delta)$ and the basic distribution is $N(\mu,\sigma^2)^+$.

The half-normal models described in the previous paragraphs have the scaling property. (It is equivalent to say that u is distributed as $N(0,\sigma_u(z,\delta)^2)^+$ or that u is distributed as $\sigma_u(z,\delta)$ times $N(0,1)^+$.) Models based on some other simple distributions, such as exponential, would also have the scaling property. (It is equivalent to say that u is distributed as exponential with parameter $\lambda(z,\delta)$, or that u is distributed as $\lambda(z,\delta)$ times an exponential variable with parameter equal to one.) However, not all commonly used models have this property. For example, Kumbhakar, Ghosh and McGuckin (1991), Huang and Liu (1994) and Battese and Coelli (1995) have considered the model in which u is distributed as $N(z'\delta,\sigma^2)^+$. (We will call this the KGMHLBC model.) This model does not have the scaling property because the variance of the pre-truncation normal is assumed to be constant (not dependent on z). Their model

[1] This functional form assumes that there is no intercept in $z'\delta$, so that overall scale is set by the constant σ. Equivalently, we could eliminate the overall constant σ if we add an intercept to $z'\delta$. Here, and elsewhere in this paper, we will omit intercept from $z'\delta$, and include a parameter that determines overall scale explicitly.

can be modified easily to have the scaling property, by letting the pre-truncation variance be proportional to the square of the pre-truncation mean. That is, the assumption that u is distributed as $h(z,\delta)$ times $N(\mu,\sigma^2)^+$ is equivalent to the assumption that u is distributed as $N[\mu h(z,\delta),\sigma^2 h(z,\delta)^2]^+$, so the latter model has the scaling property.

The use of the scaling property to generate models was suggested by Simar, Lovell and Vanden Eeckaut (1994). There is nothing sacred about this property, and it is ultimately an empirical matter whether models generated using it fit the data. However, it has some attractive features. The first of these is that it captures the idea, which we find intuitively reasonable and appealing, that the *shape* of the distribution of u is the same for all firms. The scaling factor $h(z,\delta)$ essentially just stretches or shrinks the horizontal axis, so that the scale of the distribution of u changes but its underlying shape does not. By way of contrast, consider the KGMHLBC model, which does not have the scaling property. Suppose for simplicity that $\sigma^2 = 1$. If the pre-truncation mean ($z'\delta$) equals three, say, the distribution of u is essentially normal, whereas if the pre-truncation mean equals minus three, the distribution of u is the extreme right tail of a normal, with a mode of zero and extremely fast decay of the density as u increases. On the other hand, in the truncated normal model with the scaling property (where u is distributed as $h(z,\delta)$ times $N(\mu,\sigma^2)^+$), the mean and standard deviation change with z, but the truncation point is always the same number of standard deviations from zero, so the shape does not change.

A second attractive feature of the scaling property is that it can generate very simple expressions for the effect of z on firm level efficiency or inefficiency, and these expressions do not require an assumption about the basic distribution (the distribution of u^*). For example, suppose that we pick as our scaling factor $h(z,\delta) = \exp(z'\delta)$.

Then $\delta = \partial \ln[u(z,\delta)]/\partial z$, and this is so regardless of the basic distribution. The simplicity of the interpretation of δ is of course nice, but the fact that this interpretation does not depend on the basic distribution is perhaps more fundamentally important. No similarly simple expression would exist for the KGMHLBC model, and the expression for the KGMHLBC model would rely on the truncated normal assumption. This feature of the scaling property is potentially important in empirical work, and it is also very important in our simulations, where we want to evaluate the bias in a two-step estimator. To do so we need to know what the second-step regression should be. With scaling factor $\exp(z'\delta)$, the second-step regression is a regression of ln(u) on z. This is so regardless of the basic distribution (the distribution of u*).

A third argument for the scaling property is that it makes possible estimation of β and δ, without having to specify the basic distribution. Let $u(z,\delta) = h(z,\delta)u^*$, as above, and let $\mu^* \equiv E(u^*)$, the mean of the basic distribution. Then we have

(4) $$E(y \,|\, x,z) = x'\beta - h(z,\delta)\mu^* \quad ,$$

and we can estimate β, δ and μ^* by nonlinear least squares. This possibility was noted by Simar, Lovell and Vanden Eeckaut (1994) and is discussed in Kumbhakar and Lovell (2000, section 7.3). This is potentially very useful because we can test important hypotheses, such as whether inefficiency depends on z, without having to make an assumption about the basic distribution.

2c. Why Is the Two-Step Estimator Biased?

In this section we will discuss the bias of the two-step estimator. Our discussion will be simpler if we assume that the scaling property holds, but the sense of the discussion does not depend on this.

It is widely agreed that the first step of the two-step procedure is biased if x and z are correlated. For example, see the discussion in Kumbhakar and Lovell (2000, p.

264). Some Monte Carlo evidence on the size of the bias, and also some explanation of its direction, are given by Caudill and Ford (1993). Basically, the first-step regression that ignores z suffers from an omitted variables problem, since $E(y|x,z)$ depends on z (see equation (4)) but the first-step regression does not allow for this. Standard econometric theory for least squares regression says that the estimate of β will be biased by the omission of z, if z affects y and if z and x are correlated. We are typically dealing, in the first step of the two-step procedure, with a maximum likelihood procedure, not with least squares, but this difference is unlikely to change the correctness of this conclusion, since empirically least squares and maximum likelihood are invariably very similar for coefficients other than the intercept. Also, the issue really is whether $h(z,\delta)$ is correlated with x, not whether z is correlated with x, but again as a practical matter this is not an important distinction. As pointed out by Caudill and Ford, the direction of the bias depends on the direction of the effect of z on u, and on the sign of the correlation between $h(z,\delta)$ and x. For example, if z is positively related to u (*in*efficiency), and if $h(z,\delta)$ is positively correlated with x, then neglecting z will cause the coefficient of x to be biased downward. Larger z will, other things equal, be associated with lower y and higher x, and thus the effect of x on y, not controlling for z, will appear smaller (less positive, or more negative) than it would if we controlled for z.

A second and less widely recognized problem is that the first-step technical efficiency measures are likely to be seriously underdispersed, so that the results of the second-step regression are likely to be biased downward. This is true regardless of whether x and z are correlated. To explore this point more precisely, suppose that x and z are independent, so that the first-step regression is unbiased. Thus, loosely speaking, the residual e is an unbiased estimate of the error ε = v-u. Also, suppose for simplicity that the scaling property holds. We now proceed to calculate the usual estimate of u, namely $u^* = E(u \mid \varepsilon = e)$, as in Jondrow et al. (1982) or Battese and Coelli (1988). This is a "shrinkage" estimator, where shrinkage is toward the mean, and this is intuitively reasonable because large positive ε will on average contain positive noise v, and should be shrunk downward toward the mean, while large (in absolute value) negative ε on average contain negative noise v, and should be shrunk upward toward the mean. The precise nature of the shrinkage depends on the distribution of u, and more importantly on the relative variances of v and u. For example, in the half normal case the value of u* (Jondrow et al., equation (2)) is a monotonic function of $\mu^* = -\varepsilon[\sigma_u^2/ (\sigma_u^2+\sigma_v^2)]$ and the way in which the shrinkage depends on the relative sizes of σ_u^2 and σ_v^2 is evident; but the same principle applies for other distributions. Larger variance of v (relative to u) means more noise in ε and calls for more shrinkage, and conversely.

Now, given the scaling property, it is evident that both the mean and the variance of u depend on z, and in the same direction. For example, if $h(z,\delta) = \exp(z'\delta)$ and $\delta > 0$, then large z will on average be associated with large u and also with large σ_u^2. So, compared to the case that σ_u^2 is constant, we should shrink (toward the mean) observations with large u less, and those with small u more. Saying the same thing, if we ignore the dependence of σ_u^2 on z, we will shrink the observations with large u too much, and the observations with small u too little, and our estimates of u will be underdispersed. That is, if the estimates of u are constructed ignoring the effect of z on σ_u^2, they will show less dependence on z than they should, and we should expect the second-step regressions to give downward biased estimates of the effect of z on u. From an econometric point of view, the problem is that u is measured with an error that is correlated with z, the regressor in the second-step regression.

Similar comments apply if we focus on the technical efficiencies $r \equiv e^{-u}$ rather than on u itself. Now the usual estimate is $r^* = E(r \mid \varepsilon = e)$, as given by Battese and Coelli (1988, equation (12), p. 391). Once again this is a shrinkage estimator, and ignoring the dependence of σ_u^2 on z leads to estimates that are underdispersed. So a second-step regression of some function of r* on z will suffer from the same downward bias as was discussed in the previous paragraph.

This bias in the second-step regression, due to underdispersion in the estimates of u that do not take into account the effect of z on u, does not seem to be systematically discussed in the literature. There is a brief discussion in Kumbhakar and Lovell (2000, p. 119), for a different measure of u (conditional mode rather than conditional mean), that clearly captures the essence of the above discussion. Our simulations will show that this bias is a serious (and perhaps surprisingly serious) problem.

3. SIMULATIONS

In this section we will conduct simulations to investigate the performance of the one-step and two-step estimators, in a model where inefficiency depends on some variables z. The one-step MLE will be based on the correctly specified model, and will therefore be consistent and asymptotically efficient. Thus the only interesting question for the one-step MLE is whether it performs well in finite samples of reasonable size. For the two-step estimator, we expect to find biased results, regardless of sample size, and the interesting questions are the severity of the bias and the way in which it depends on the various parameters of the model.

3a. Design of the Experiment

Our data follow the simple stochastic frontier model:

(5) $y_i = \beta x_i + v_i - u_i \quad , \quad i = 1,\ldots,N$.

All symbols are scalars. The v_i are i.i.d. $N(0,\sigma_v^2)$. The u_i are truncated normals scaled by an exponential function of a variable z_i; specifically,

(6) $u_i = \gamma \exp(\delta z_i) u_i^*$

where the u_i^* are i.i.d. $N(\mu,1)^+$. The vectors $(x_i,z_i)'$ are i.i.d. standard bivariate normal with correlation ρ. That is, there is only one input (x) and one variable that affects the distribution of inefficiency (z), and the parameter ρ controls their correlation. Finally, (x_i,z_i), v_i and u_i^* are mutually independent. Data were generated using the Stata random number generator. The number of replications of the experiment (for each case considered) was 2000.

From the point of view of experimental design, the parameters to be picked are β, δ, ρ, γ, σ_v, μ and N. Our strategy will be to pick a "Base Case" set of parameters, listed below. We will then vary each of the parameters, one at a time, holding the other parameters equal to their Base Case values.

Base Case Parameter Values: $\beta = 0$, $\delta = 1$, $\rho = 0.5$, $\gamma = 1$, $\sigma_v = 1$, $\mu = 0$, N = 200. For this set of parameters, average technical efficiency is $E(e^{-u}) = 0.5165$.

From the point of view of estimation, the parameters are β, δ, μ, γ and σ_v^2. In our Tables we report the mean, standard deviation and mean square error (MSE) for the estimates of these parameters. We also estimate each of the individual technical efficiencies, $r_i = \exp(-u_i)$, and we report the mean, standard deviation and MSE averaged over observations as well as replications. In addition, we report the correlation between the true and estimated r_i. Finally, for the two-step estimators of δ, we report (in addition to mean, standard deviation and MSE) the R^2 of the second-stage regression.

For the one-step estimates, we simply calculate the MLE based on the correctly specified model, and the estimates of the r_i that follow from this model. For the two-step estimates, we calculate the MLE with δ set equal to zero. That is, we estimate the truncated normal model in which the distribution of u is assumed not to depend on z. Then we calculate the estimates of the u_i and the r_i that follow from this model, and we calculate the second-step estimate of δ by regressing the logarithm of estimated u_i on z_i.

As a matter of curiosity, we also calculate a second-step estimate of δ based on the estimates of the u_i from the one-step model. That is, we regress the logarithm of the estimated u_i from the one-step model on z_i. In any actual application, this would be a silly thing to do because we would already have the one-step estimate of δ. In the present simulation, we do this because we want to see how much of the bias in the usual two-step estimator of δ is due to having estimated the u_i from an incorrectly specified model. Thus, in the tables, for the one-step model we have both the one-step estimate $\hat{\delta}$ and a second-step estimator $\hat{\delta}$-2S, whereas for the two-step model we have only $\hat{\delta}$-2S.

Our calculations were carried out in Stata and used the Stata numerical maximization routine to maximize the likelihood functions. As is often the case in simulations that involve numerical maximization, there were some problems with outliers, especially in estimation of γ and μ (the parameters of the truncated normal distribution). Our summary statistics are averages and are very sensitive to extreme outliers. In the end we simply truncated our results by discarding the replications with the 0.3% most extreme upper tail and lower tail estimates of μ and of γ. This would be a maximum of 24 replications (of 2000), but was usually only about half that amount, since replications with extreme estimates of μ also tended to have extreme estimates of γ, and vice-versa. This truncation of the results made very little difference for the parameters other than μ and γ.

3b. Results for the Base Case Parameter Values

Case 1 in Table 1 corresponds to our base case parameter values: $\beta = 0$, $\delta = 1$, $\rho = 0.5$, $\gamma = 1$, $\sigma_v = 1$, $\mu = 0$, $N = 200$. We first note that the one-step estimates look pretty good. In particular, there is no evidence of significant bias. The parameters of the truncated normal base distribution (μ and γ) have rather large variances (their standard deviations are definitely not small relative to the parameter values themselves), reflecting the commonly cited view (e.g., Ritter and Simar (1997, p. 181)) that these are hard parameters to estimate. But this does not seem to cause any problems for the parameters of main interest (β, δ and the individual r_i).

Now consider the two-step estimates. The estimates of μ and γ are obviously very strange, with very large biases and variances. More importantly, we find exactly the types of bias that we expect in the parameters of main interest. First, the estimate of β is biased downward (mean = -0.31, compared to the true value = 0). This direction of bias is as expected given the positive correlation between x and z, and the positive relationship between z and the average level of u. Second, the estimates of the r_i (the technical efficiencies) are biased downward (mean = 0.45, compared to the mean of the actual r_i of 0.52). They are less strongly correlated with the true r_i and they are underdispersed, compared to the estimates from the one-step model. Third, the second-step estimator of δ is very significantly biased downward (mean = 0.35, compared to the true value = 1). All of these biases are in the expected direction, and the size of the biases is definitely not small. We would characterize the biases in the two-step estimates as serious. This is essentially the case against using two-step estimates.[2]

[2] An interesting curiosity is the bias in the two-step estimator of δ based on the estimates of u_i from the one-step model (mean estimate = 0.92, compared to the true value of $\delta = 1$). This is not just finite-sample bias (it persists with larger sample sizes). It reflects the fact that $\ln[E(u|z)] = z'\delta$, whereas the condition for unbiasedness of the second-step regression would be $E[\ln(u)|z] = z'\delta$, which does not hold.

We also consider the same parametric configuration as in the Base Case (Case 1), but where we set $\mu \equiv 0$ in estimation, instead of estimating μ. That is, we estimate the scaled half-normal model, which in this case is the correctly specified model. We call this Case 1A. For the one-step estimates, imposing $\mu = 0$ makes very little difference, except that the variance of the estimate of γ is substantially reduced. For the two-step estimates, surprisingly, imposing $\mu = 0$ makes things *worse* (even though $\mu = 0$ is a correct restriction). The bias of the estimate of β and of the two-step estimate of δ increases, and the estimates of the r_i are also more biased, and less correlated with the true r_i, compared to Case 1. We do not understand this result, but it does make clear that the problems with the two-step estimator in Case 1 do not primarily arise from the fact that we are poorly estimating the parameters of the truncated normal base distribution.

3c. Effects of Changing β

Changing β has no substantive effects on our results. The mean of the estimate (both one-step and two-step) of β changes by the same amount that β is changed, so that the bias, standard deviation and MSE of the estimates of β are unchanged. For all of the other parameters the estimates are identical before and after the change in β. Therefore, there is no need to tabulate these results.

3d. Effects of Changing δ

We now keep all other parameters at their Base Case values, but consider $\delta = 0.5$ (Case 2) and $\delta = 0$ (Case 3) in addition to $\delta = 1$ (Base Case). The results for these cases are given in Table 1.

For the one-step estimator, the true value of δ is not terribly important. Changing δ makes very little difference to the properties of the estimates of β, δ or σ_v. As $\delta \rightarrow 0$, our ability to estimate μ (and γ, to a lesser extent) deteriorates seriously.

The estimates of the individual r_i become slightly more biased, and noticeably less strongly correlated with the true r_i, when δ is small. Presumably this is so because when δ is large, z correlates with u and is useful to help estimate u. Still, except for μ and γ, the one-step estimators do fine even when $\delta = 0$.

For the two-step estimator, δ is a very important parameter because it dictates the degree of misspecification of the first step of the two-step procedure. As expected, the bias of the first-step estimate of β is effectively zero when $\delta = 0$, and it grows with δ. The individual r_i are seriously biased for $\delta = 1$ but not for the smaller values of δ. The two-step estimator of δ is biased for both $\delta = 1$ and $\delta = 0.5$, but the bias disappears as $\delta \rightarrow 0$. That is, the second-step estimator of δ may be able to tell us whether or not z affects u (whether or not $\delta = 0$)[3] but it cannot accurately estimate the effect of z on u when this effect exists. An interesting result is that the estimates of the r_i are better for the two-step procedure than for the one-step procedure, when $\delta = 0$. In this case the first-step estimator of β is unbiased, and we then estimate the r_i under the *correct* assumption that they do not depend on the z_i, so we ought to do well in this case. All in all, the two-step estimator performs well when the second step is not needed ($\delta = 0$) but is otherwise unreliable.

3e. Effects of Changing ρ

In Table 2 we report the results of our experiments in which we change the value of ρ, holding the other parameters constant at their Base Case values. Cases 4, 5, 6, 7 and 8 are defined by $\rho = -0.5, 0, 0.25, 0.75$ and 0.9. We also report again the results for the Base Case with $\rho = 0.5$.

For the one-step estimators, the value of ρ is not important. The results are

[3] We did not consider formal tests of the hypothesis that $\delta = 0$. Therefore we do not know whether a formal test based on the two-step estimator of δ would in fact be valid. However, at least the point estimate appears to be unbiased when $\delta = 0$.

essentially identical across all these cases.

For the two-step estimator, the value of ρ is important because it determines the sign and the size of the bias of the first-step estimate of β (which in turn influences the performance of the subsequent steps). First, we note that the results for ρ = -0.5 (Case 4) are essentially identical to those for ρ = 0.5 (Case 1). The bias of the estimate of β reverses sign and nothing else changes. Therefore we can effectively restrict our attention to positive values of ρ. Second, we note that the bias of the estimate of β grows as ρ grows, as expected. Third, it is interesting that the mean and the dispersion of the estimates of the individual r_i do not depend noticeably on ρ. They are biased and underdispersed even when ρ = 0, as was argued in Section 2. Finally, the second-step estimate of δ is seriously biased even when ρ = 0, though its bias does grow with ρ.

3.f Effects of Changing μ

In Table 3 we report the results of our experiments in which we change the value of μ, holding the other parameters constant at their Base Case values. Cases 9, 10, 11 and 12 are defined by μ = 1, 0.5, -0.5 and -1. We also report again the results for the Base Case with μ = 0. We note that changing μ changes the shape of the truncated normal distribution of u (inefficiency). We might anticipate, following the arguments of Ritter and Simar (1997), that estimation of this model will be harder when μ is positive and large. As $\mu \to \infty$, the distribution of u becomes normal and presumably becomes confounded with the normal distribution of v (statistical noise). However, as long as δ is non-zero, this argument may be less than compelling, because the distribution of u depends on z whereas the distribution of v does not, and this is another way to distinguish u from v. Furthermore, the degree of truncation is also relevant, and this decreases as μ increases. For example, when μ is large and positive, the degree of truncation is very small, and the shape of the distribution (if it can be separated from that of v) contains a lot of information about the parameters μ and σ^2. When μ is large (in absolute value) but negative, on the other hand, we observe only the extreme right tail of the distribution, whose shape may not be very informative about μ and σ^2, and the estimation problem may be harder just on that basis.

Our results indicate that, for the one-step estimator, the value of μ is not terribly important. It is true that, as μ moves from plus one to negative one, the estimates of μ and γ deteriorate considerably, which is consistent with the argument presented at the end of the preceding paragraph. But for the other parameters this does not make much difference.

For the two-step estimators, we see more differences as μ changes, but they are still not really important or striking. The estimates of μ and γ are very bad no matter what the true value of μ is. As μ moves from plus one to negative one, the estimates of β improve a little, and the estimates of the individual r_i perhaps improve a little. But these are not major changes.

3g. Effects of Changing γ

In Table 4 we report the results of our experiments in which we change the value of γ, holding the other parameters constant at their Base Case values. We consider $\gamma = 3$ (Case 13) and $\gamma = 5$ (Case 14) in addition to the Base Case value of $\gamma = 1$. The parameter γ represents pure scale in the distribution of u. Increasing γ while holding σ_v constant has the effect of increasing the size of inefficiency relative to noise, and should tend to make it easier to estimate technical inefficiency precisely.

For the one-step estimators, changing the value of γ makes relatively little difference. For larger γ, we estimate β a little worse, but we estimate δ and the individual r_i a little better. For the two-step estimators, γ makes more of a difference.

As γ increases, the bias of the estimate of β increases, but the bias of the second-step estimate of δ decreases quite noticeably.

3.h Effects of Changing σ_v

In Table 4 we also report the results of our experiments in which we change the value of σ_v, holding the other parameters constant at their Base Case values. We consider σ_v = 3 and 5 in addition to the Base Case value of σ_v = 1. Increasing σ_v increases the amount of statistical noise in the model, and should increase the bias and/ or variance for each of the estimated parameters. This turns out to be true. The differences are bigger for the two-step estimators than for the one-step estimators. The bias of the two-step estimator of β and especially of the two-step estimator of δ increases markedly with σ_v. For example, the mean of the two-step estimator of δ is 0.35 for σ_v = 1, 0.12 for σ_v = 3, and 0.05 for σ_v = 5. The true value of δ is one, so these are large biases indeed. No such bias problem exists for the one-step estimator, even for the largest value of σ_v.

3.i Effects of Changing N

In Table 5 we report the results of our experiments in which we change the sample size, N. We consider N = 500 and N = 1000 in addition to the Base Case value of N = 200. Naturally we can hope to estimate more precisely when the sample size is larger. However, we do not expect to see much else in these experiments, because the biases we have identified above are expected to persist asymptotically. The results are quite consistent with these expectations. The standard deviations of the estimators fall as N increases, but nothing else changes much.

4. CONCLUDING REMARKS

In this paper we have discussed models that allow one to estimate each firm's level of technical inefficiency and the way in which inefficiency depends on observable variables "z" (typically firm characteristics). Several such models have been previously suggested in the literature. They are typically estimated by maximum likelihood, in a single "step", and hence reference is made to "one-step" estimation or "one-step" models. This is in contrast to "two-step" methods, in which the first step is the estimation of a standard model that ignores the effect of z on inefficiency, and the second step is a regression of some measure of inefficiency on z.

The paper makes two contributions. First, we make some arguments in favor of the "scaling property" that the one-sided inefficiency error can be written as a function of z times a one-sided error independent of z. Second, we analyze the properties of the two-step estimator. We identify two sources of bias. The first step of the two-step procedure is biased for the regression parameters if z and the inputs "x" are correlated, as is well known. A less well known fact is that, even if z and x are independent, the estimated inefficiencies are underdispersed when we ignore the effect of z on inefficiency. This causes the second- step estimate of the effect of z on inefficiency to be biased downward (toward zero).

We perform Monte Carlo simulations to investigate the performance of the one-step and two-step estimators of a simple model that has the scaling property. The one-step estimators are based on a correctly specified model and are asymptotically optimal. We find that the one-step estimators also generally perform quite well in finite samples. The two-step estimators do not perform well. We find very significant bias in the first step, so long as x and z are correlated. We also find very significant bias in the second step, whether or not x and z are correlated, so long as inefficiency actually depends on z. These biases are substantial enough that we would recommend against using two-step procedures in any circumstances that we can envision.

Table 1: Base Case plus Changes in δ

Base Case: $\beta = 0$, $\rho = 0.5$, $\delta = 1$, $\mu = 0$, $\gamma = 1$, $\sigma_v = 1$; $E(e^{-u}) = 0.516$

CASE	change from base	param.	ONE-STEP mean	ONE-STEP s.d.	ONE-STEP MSE	ONE-STEP corr*	TWO-STEP mean	TWO-STEP s.d.	TWO-STEP MSE	TWO-STEP corr*
1	None	$\hat{\beta}$	-0.0010	0.0903	0.0082		-0.3064	0.1037	0.1046	
	(i.e.,	$\hat{\delta}$	1.0010	0.1189	0.0141		–	–	–	
	$\delta = 1$)	$\hat{\mu}$	0.0564	0.8150	0.6670		-37.2947	0.3833	1391.0389	
		$\hat{\gamma}$	1.0153	0.4089	0.1674		53.4673	7.0771	2802.8732	
		$\hat{\sigma_v}$	0.9910	0.0773	0.0061		1.0096	0.0922	0.0086	
		$E(\widehat{e^{-u}})$	0.5141	0.0340	0.0348	0.8113	0.4482	0.0234	0.0617	0.6564
		$\hat{\delta}$–2S	0.9180	0.1212	0.0214	0.7917	0.3539	0.0543	0.4204	0.2371
1A	μ set to 0	$\hat{\beta}$	-0.0033	0.0900	0.0081		-0.4595	0.1360	0.2296	
		$\hat{\delta}$	0.9963	0.1150	0.0132		–	–	–	
		$\hat{\mu}$	–	–	–		–	–	–	
		$\hat{\gamma}$	1.0007	0.1202	0.0144		2.3584	0.6666	2.2894	
		$\hat{\sigma_v}$	0.9864	0.0736	0.0056		1.3374	0.1629	0.1404	
		$E(\widehat{e^{-u}})$	0.5169	0.0302	0.0344	0.8128	0.3692	0.0303	0.0856	0.6218
		$\hat{\delta}$–2S	0.9133	0.1155	0.0209	0.7838	0.2380	0.0454	0.5827	0.2030
2	$\delta = 0.5$	$\hat{\beta}$	0.0002	0.0905	0.0082		-0.1748	0.0895	0.0386	
		$\hat{\delta}$	0.5069	0.1121	0.0126		–	–	–	
	$E(e^{-u})$	$\hat{\mu}$	0.4229	4.6666	21.9452		-17.4599	17.0405	595.0805	
	=0.520	$\hat{\gamma}$	1.0700	0.8234	0.6825		16.3313	15.3375	470.1703	
		$\hat{\sigma_v}$	0.9974	0.0882	0.0078		0.9823	0.0892	0.0083	
		$E(\widehat{e^{-u}})$	0.5162	0.0374	0.0445	0.6493	0.5160	0.0300	0.0523	0.5633
		$\hat{\delta}$–2S	0.4614	0.1117	0.0140	0.5759	0.1469	0.0463	0.1268	0.0733
3	$\delta = 0$	$\hat{\beta}$	-0.0013	0.0935	0.0087		-0.0015	0.0821	0.0067	
		$\hat{\delta}$	0.0014	0.1171	0.0137		–	–	–	
	$E(e^{-u})$	$\hat{\mu}$	5.1757	14.6694	241.8693		0.1210	4.1249	17.0209	
	=0.523	$\hat{\gamma}$	1.1652	2.0469	4.2150		1.3278	2.4939	6.3239	
		$\hat{\sigma_v}$	1.0159	0.1049	0.0113		1.0120	0.1022	0.0106	
		$E(\widehat{e^{-u}})$	0.5093	0.0468	0.0538	0.3976	0.5107	0.0451	0.0521	0.4891
		$\hat{\delta}$–2S	0.0012	0.1096	0.0120	0.2480	-0.0001	0.0224	0.0005	-0.0013

* *corr* is the correlation coefficient between the true and the estimated $E(e^{-u})$, and is the $\bar{R}^2$ of the 2nd-step regression of $\ln E(\hat{u})$ on z.

Table 2: Changes in ρ

Base Case: $\beta = 0$, $\rho = 0.5$, $\delta = 1$, $\mu = 0$, $\gamma = 1$, $\sigma_v = 1$; $E(e^{-u}) = 0.516$

CASE	change from base	param.	ONE-STEP mean	ONE-STEP s.d.	ONE-STEP MSE	ONE-STEP corr*	TWO-STEP mean	TWO-STEP s.d.	TWO-STEP MSE	TWO-STEP corr*
4	$\rho = -0.5$	$\hat{\beta}$	-0.0025	0.0892	0.0080		0.3042	0.1006	0.1027	
		$\hat{\delta}$	0.9996	0.1167	0.0136		–	–	–	
		$\hat{\mu}$	0.0593	0.8246	0.6831		-37.2760	0.8928	1390.2968	
		$\hat{\gamma}$	1.0161	0.4109	0.1690		53.4787	7.0192	2803.2616	
		$\hat{\sigma_v}$	0.9928	0.0773	0.0060		1.0102	0.0916	0.0085	
		$E(\hat{e^{-u}})$	0.5132	0.0350	0.0347	0.8120	0.4479	0.0233	0.0617	0.6565
		$\hat{\delta}$–2S	0.9168	0.1189	0.0211	0.7916	0.3531	0.0542	0.4214	0.2362
5	$\rho = 0$	$\hat{\beta}$	-0.0016	0.0861	0.0074		-0.0006	0.0992	0.0098	
		$\hat{\delta}$	1.0008	0.1167	0.0136		–	–	–	
		$\hat{\mu}$	0.0643	0.8327	0.6971		-37.3025	0.1570	1391.4985	
		$\hat{\gamma}$	1.0144	0.4238	0.1797		54.2774	7.2110	2890.4516	
		$\hat{\sigma_v}$	0.9923	0.0781	0.0061		1.0217	0.0918	0.0089	
		$E(\hat{e^{-u}})$	0.5135	0.0348	0.0348	0.8116	0.4484	0.0234	0.0559	0.7013
		$\hat{\delta}$–2S	0.9181	0.1195	0.0210	0.7927	0.4069	0.0548	0.3547	0.3079
6	$\rho = 0.25$	$\hat{\beta}$	-0.0016	0.0876	0.0077		-0.1494	0.1007	0.0324	
		$\hat{\delta}$	1.0008	0.1174	0.0138		–	–	–	
		$\hat{\mu}$	0.0616	0.8187	0.6737		-37.2979	0.2917	1391.2210	
		$\hat{\gamma}$	1.0134	0.4083	0.1668		54.0751	7.2026	2868.8151	
		$\hat{\sigma_v}$	0.9916	0.0778	0.0061		1.0187	0.0919	0.0088	
		$E(\hat{e^{-u}})$	0.5137	0.0343	0.0348	0.8116	0.4484	0.0234	0.0573	0.6907
		$\hat{\delta}$–2S	0.9181	0.1202	0.0211	0.7924	0.3943	0.0550	0.3699	0.2903
1	None	$\hat{\beta}$	-0.0010	0.0903	0.0082		-0.3064	0.1037	0.1046	
	(i.e.,	$\hat{\delta}$	1.0010	0.1189	0.0141		–	–	–	
	$\rho = 0.5$)	$\hat{\mu}$	0.0564	0.8150	0.6670		-37.2947	0.3833	1391.0389	
		$\hat{\gamma}$	1.0153	0.4089	0.1674		53.4673	7.0771	2802.8732	
		$\hat{\sigma_v}$	0.9910	0.0773	0.0061		1.0096	0.0922	0.0086	
		$E(\hat{e^{-u}})$	0.5141	0.0340	0.0348	0.8113	0.4482	0.0234	0.0617	0.6564
		$\hat{\delta}$–2S	0.9180	0.1212	0.0214	0.7917	0.3539	0.0543	0.4204	0.2371
7	$\rho = 0.75$	$\hat{\beta}$	-0.0012	0.0947	0.0090		-0.4828	0.1097	0.2451	
		$\hat{\delta}$	1.0010	0.1210	0.0146		–	–	–	
		$\hat{\mu}$	0.0459	0.8273	0.6862		-37.2003	1.6154	1386.4707	
		$\hat{\gamma}$	1.0203	0.4196	0.1764		52.3762	7.1109	2690.0566	
		$\hat{\sigma_v}$	0.9905	0.0767	0.0060		0.9934	0.0926	0.0086	
		$E(\hat{e^{-u}})$	0.5142	0.0341	0.0348	0.8111	0.4474	0.0233	0.0700	0.5922
		$\hat{\delta}$–2S	0.9176	0.1223	0.0217	0.7905	0.2785	0.0502	0.5231	0.1500
8	$\rho = 0.9$	$\hat{\beta}$	-0.0016	0.0998	0.0100		-0.6053	0.1159	0.3798	
		$\hat{\delta}$	1.0003	0.1254	0.0157		–	–	–	
		$\hat{\mu}$	0.0168	0.9107	0.8293		-36.8158	3.8123	1369.9273	
		$\hat{\gamma}$	1.0388	0.4837	0.2354		51.0769	8.3744	2577.7868	
		$\hat{\sigma_v}$	0.9898	0.0763	0.0059		0.9788	0.0926	0.0090	
		$E(\hat{e^{-u}})$	0.5144	0.0342	0.0349	0.8106	0.4467	0.0233	0.0777	0.5322
		$\hat{\delta}$–2S	0.9165	0.1264	0.0229	0.7884	0.2094	0.0449	0.6270	0.0858

* *corr* is the correlation coefficient between the true and the estimated $E(e^{-u})$, and is the $\bar{R}^2$ of the 2nd-step regression of $\ln E(\hat{u})$ on z.

Table 3: Changes in μ

Base Case: $\beta = 0$, $\rho = 0.5$, $\delta = 1$, $\mu = 0$, $\gamma = 1$, $\sigma_v = 1$; $E(e^{-u}) = 0.516$

			ONE-STEP				TWO-STEP			
CASE	change from base	param.	mean	s.d.	MSE	corr*	mean	s.d.	MSE	corr*
9	$\mu = 1$	$\hat{\beta}$	-0.0013	0.0963	0.0093		-0.3914	0.1191	0.1674	
		$\hat{\delta}$	1.0002	0.0850	0.0072		–	–	–	
	$E(e^{-u})$	$\hat{\mu}$	1.0591	0.3951	0.1595		-37.2499	1.1759	1464.4405	
	=0.386	$\hat{\gamma}$	0.9902	0.1994	0.0399		82.0758	9.9469	6672.1806	
		$\hat{\sigma_v}$	0.9883	0.0864	0.0076		0.9683	0.1053	0.0121	
		$E(\hat{e^{-u}})$	0.3850	0.0294	0.0285	0.8387	0.3499	0.0197	0.0513	0.6891
		$\hat{\delta}$–2S	0.9317	0.0888	0.0125	0.8025	0.4986	0.0566	0.2546	0.3469
10	$\mu = 0.5$	$\hat{\beta}$	-0.0011	0.0932	0.0087		-0.3458	0.1103	0.1317	
		$\hat{\delta}$	1.0006	0.1013	0.0103		–	–	–	
	$E(e^{-u})$	$\hat{\mu}$	0.5680	0.5235	0.2785		-37.2813	0.7274	1427.9555	
	=0.454	$\hat{\gamma}$	0.9932	0.2639	0.0697		65.8462	8.2339	4272.7900	
		$\hat{\sigma_v}$	0.9897	0.0814	0.0067		0.9934	0.0980	0.0096	
		$E(\hat{e^{-u}})$	0.4525	0.0319	0.0326	0.8242	0.4006	0.0213	0.0574	0.6752
		$\hat{\delta}$–2S	0.9225	0.1044	0.0169	0.7905	0.4215	0.0557	0.3378	0.2868
1	None	$\hat{\beta}$	-0.0010	0.0903	0.0082		-0.3064	0.1037	0.1046	
	(i.e.,	$\hat{\delta}$	1.0010	0.1189	0.0141		–	–	–	
	$\mu = 0$)	$\hat{\mu}$	0.0564	0.8150	0.6670		-37.2947	0.3833	1391.0389	
		$\hat{\gamma}$	1.0153	0.4089	0.1674		53.4673	7.0771	2802.8732	
		$\hat{\sigma_v}$	0.9910	0.0773	0.0061		1.0096	0.0922	0.0086	
		$E(\hat{e^{-u}})$	0.5141	0.0340	0.0348	0.8113	0.4482	0.0234	0.0617	0.6564
		$\hat{\delta}$–2S	0.9180	0.1212	0.0214	0.7917	0.3539	0.0543	0.4204	0.2371
11	$\mu = -0.5$	$\hat{\beta}$	-0.0006	0.0876	0.0077		-0.2739	0.0986	0.0847	
		$\hat{\delta}$	1.0020	0.1371	0.0188		–	–	–	
	$E(e^{-u})$	$\hat{\mu}$	-0.7891	3.3635	11.3913		-37.2763	0.8090	1353.1526	
	=0.571	$\hat{\gamma}$	1.2321	1.9140	3.7154		44.1503	6.3150	1901.8065	
		$\hat{\sigma_v}$	0.9924	0.0740	0.0055		1.0190	0.0877	0.0080	
		$E(\hat{e^{-u}})$	0.5676	0.0358	0.0354	0.7998	0.4914	0.0258	0.0640	0.6347
		$\hat{\delta}$–2S	0.9188	0.1385	0.0258	0.8025	0.2975	0.0525	0.4963	0.1970
12	$\mu = -1$	$\hat{\beta}$	-0.0009	0.0863	0.0074		-0.2469	0.0945	0.0699	
		$\hat{\delta}$	1.0038	0.1564	0.0245		–	–	–	
	$E(e^{-u})$	$\hat{\mu}$	-2.3774	7.0546	51.6401		-37.2169	1.6121	1314.2578	
	=0.617	$\hat{\gamma}$	1.7497	3.4231	12.2739		37.0215	5.8916	1332.2415	
		$\hat{\sigma_v}$	0.9935	0.0713	0.0051		1.0252	0.0843	0.0077	
		$E(\hat{e^{-u}})$	0.6129	0.0373	0.0350	0.7893	0.5306	0.0287	0.0645	0.6111
		$\hat{\delta}$–2S	0.9235	0.1573	0.0306	0.8186	0.2505	0.0509	0.5644	0.1645

* *corr* is the correlation coefficient between the true and the estimated $E(e^{-u})$, and is the $\bar{R}^2$ of the 2nd-step regression of $\ln E(\hat{u})$ on z.

Table 4: Changes in γ and σ_v

Base Case: $\beta = 0$, $\rho = 0.5$, $\delta = 1$, $\mu = 0$, $\gamma = 1$, $\sigma_v = 1$; $E(e^{-u}) = 0.516$

CASE	change from base	param.	ONE-STEP mean	s.d.	MSE	corr*	TWO-STEP mean	s.d.	MSE	corr*
1	None	$\hat{\beta}$	-0.0010	0.0903	0.0082		-0.3064	0.1037	0.1046	
	(i.e.,	$\hat{\delta}$	1.0010	0.1189	0.0141		–	–	–	
	$\gamma = 1$,	$\hat{\mu}$	0.0564	0.8150	0.6670		-37.2947	0.3833	1391.0389	
	$\sigma_v = 1$)	$\hat{\gamma}$	1.0153	0.4089	0.1674		53.4673	7.0771	2802.8732	
		$\hat{\sigma_v}$	0.9910	0.0773	0.0061		1.0096	0.0922	0.0086	
		$\widehat{E(e^{-u})}$	0.5141	0.0340	0.0348	0.8113	0.4482	0.0234	0.0617	0.6564
		$\hat{\delta}$–2S	0.9180	0.1212	0.0214	0.7917	0.3539	0.0543	0.4204	0.2371
13	$\gamma = 3$	$\hat{\beta}$	-0.0006	0.1193	0.0142		-0.4270	0.1486	0.2044	
		$\hat{\delta}$	1.0008	0.0784	0.0061		–	–	–	
	$E(e^{-u})$	$\hat{\mu}$	0.0429	0.4471	0.2016		-37.2787	0.1688	1389.7299	
	=0.287	$\hat{\gamma}$	2.9739	0.6010	0.3617		150.8042	18.2079	22177.4423	
		$\hat{\sigma_v}$	0.9827	0.1088	0.0121		0.9857	0.1306	0.0173	
		$\widehat{E(e^{-u})}$	0.2859	0.0246	0.0294	0.8281	0.2505	0.0171	0.0420	0.7510
		$\hat{\delta}$–2S	0.8998	0.0795	0.0163	0.6227	0.6172	0.0626	0.1505	0.3698
14	$\gamma = 5$	$\hat{\beta}$	-0.0022	0.1452	0.0211		-0.4714	0.1850	0.2565	
		$\hat{\delta}$	1.0002	0.0693	0.0048		–	–	–	
	$E(e^{-u})$	$\hat{\mu}$	0.0359	0.3889	0.1525		-37.2733	0.1703	1389.3243	
	=0.201	$\hat{\gamma}$	4.9578	0.8545	0.7316		248.1483	29.9097	60015.2145	
		$\hat{\sigma_v}$	0.9740	0.1383	0.0198		0.9679	0.1607	0.0268	
		$\widehat{E(e^{-u})}$	0.2006	0.0202	0.0232	0.8329	0.1773	0.0150	0.0308	0.7750
		$\hat{\delta}$–2S	0.9098	0.0718	0.0133	0.5657	0.7200	0.0658	0.0827	0.4058
15	$\sigma_v = 3$	$\hat{\beta}$	-0.0060	0.2395	0.0574		-0.5306	0.2490	0.3436	
		$\hat{\delta}$	1.0126	0.2260	0.0512		–	–	–	
	$E(e^{-u})$	$\hat{\mu}$	3.9736	25.4427	662.7908		-33.7819	10.4015	1249.3507	
	=0.516	$\hat{\gamma}$	2.1501	5.4410	30.9122		53.3825	19.8900	3139.3434	
		$\hat{\sigma_v}$	2.9888	0.1892	0.0359		3.0798	0.2223	0.0557	
		$\widehat{E(e^{-u})}$	0.5057	0.0729	0.0541	0.7189	0.4047	0.0501	0.1026	0.3484
		$\hat{\delta}$–2S	0.9787	0.2293	0.0530	0.9306	0.1203	0.0464	0.7761	0.0852
16	$\sigma_v = 5$	$\hat{\beta}$	-0.0127	0.3923	0.1539		-0.6125	0.3905	0.5276	
		$\hat{\delta}$	1.0335	0.3555	0.1274		–	–	–	
	$E(e^{-u})$	$\hat{\mu}$	14.3686	38.2494	1668.7325		-24.2941	16.7575	870.8771	
	=0.516	$\hat{\gamma}$	3.1626	8.3682	74.6688		40.1590	30.4246	2458.6210	
		$\hat{\sigma_v}$	4.9773	0.2880	0.0834		5.1313	0.3230	0.1215	
		$\widehat{E(e^{-u})}$	0.4992	0.1132	0.0667	0.6915	0.3990	0.0954	0.1178	0.2299
		$\hat{\delta}$–2S	1.0160	0.3576	0.1281	0.9594	0.0485	0.0355	0.9066	0.0384

* *corr* is the correlation coefficient between the true and the estimated $E(e^{-u})$, and is the $\bar{R}^2$ of the 2nd-step regression of $\ln E(\hat{u})$ on z.

Table 5: Changes in N

Base Case: $\beta = 0$, $\rho = 0.5$, $\delta = 1$, $\mu = 0$, $\gamma = 1$, $\sigma_v = 1$; $E(e^{-u}) = 0.516$

CASE	change from base	param.	ONE-STEP mean	ONE-STEP s.d.	ONE-STEP MSE	ONE-STEP corr*	TWO-STEP mean	TWO-STEP s.d.	TWO-STEP MSE	TWO-STEP corr*
1	None	$\hat{\beta}$	-0.0010	0.0903	0.0082		-0.3064	0.1037	0.1046	
	(i.e.,	$\hat{\delta}$	1.0010	0.1189	0.0141		–	–	–	
	N=200)	$\hat{\mu}$	0.0564	0.8150	0.6670		-37.2947	0.3833	1391.0389	
		$\hat{\gamma}$	1.0153	0.4089	0.1674		53.4673	7.0771	2802.8732	
		$\hat{\sigma_v}$	0.9910	0.0773	0.0061		1.0096	0.0922	0.0086	
		$E(\hat{e^{-u}})$	0.5141	0.0340	0.0348	0.8113	0.4482	0.0234	0.0617	0.6564
		$\hat{\delta}$–2S	0.9180	0.1212	0.0214	0.7917	0.3539	0.0543	0.4204	0.2371
17	N=500	$\hat{\beta}$	-0.0005	0.0572	0.0033		-0.3062	0.0661	0.0981	
		$\hat{\delta}$	0.9995	0.0724	0.0052		–	–	–	
		$\hat{\mu}$	-0.0123	0.5056	0.2557		-37.1869	0.2026	1382.9029	
		$\hat{\gamma}$	1.0208	0.2393	0.0577		53.4005	4.3493	2764.7151	
		$\hat{\sigma_v}$	0.9973	0.0454	0.0021		1.0161	0.0553	0.0033	
		$E(\hat{e^{-u}})$	0.5160	0.0220	0.0338	0.8145	0.4470	0.0146	0.0615	0.6591
		$\hat{\delta}$–2S	0.9134	0.0736	0.0129	0.7911	0.3520	0.0336	0.4210	0.2375
18	N=1000	$\hat{\beta}$	-0.0004	0.0416	0.0017		-0.3070	0.0471	0.0965	
		$\hat{\delta}$	0.9977	0.0522	0.0027		–	–	–	
		$\hat{\mu}$	0.0041	0.3295	0.1085		-37.0701	0.2502	1374.2512	
		$\hat{\gamma}$	1.0056	0.1549	0.0240		53.3050	3.0675	2745.2215	
		$\hat{\sigma_v}$	0.9983	0.0315	0.0010		1.0168	0.0386	0.0018	
		$E(\hat{e^{-u}})$	0.5156	0.0154	0.0337	0.8149	0.4463	0.0104	0.0616	0.6581
		$\hat{\delta}$–2S	0.9114	0.0528	0.0106	0.7923	0.3513	0.0241	0.4214	0.2381

* *corr* is the correlation coefficient between the true and the estimated $E(e^{-u})$, and is the $\bar{R}^2$ of the 2nd-step regression of $\ln E(\hat{u})$ on z.