Using Least Squares and Tobit in Second Stage DEA Efficiency Analyses

John McDonald
Flinders University, Adelaide

Flinders Business School Research Paper Series
2008-3
ISSN 1441-3906

Address for correspondence:
Flinders Business School, Flinders University, GPO Box 2100, Adelaide 5001, Australia
E-mail: john.mcdonald@flinders.edu.au

Running title as for full title

Keywords: data envelopment analysis, measuring efficiency, tobit, least squares, fractional data.

Abstract
The paper examines second stage DEA efficiency analyses, within the context of a censoring data generating process (DGP) and a fractional data DGP, when efficiency scores are treated as descriptive measures of the relative performance of units in the sample. It is argued that the efficiency scores are not generated by a censoring process but are fractional data. Tobit estimation in this situation is inappropriate. In contrast, ordinary least squares is a consistent estimator, and, if White’s heteroskedastic-consistent standard errors are calculated, large sample tests can be performed which are robust to heteroskedasticity and the distribution of the disturbances. For a more refined analysis Papke and Wooldridge’s (1996) method has some advantages, but is more complex and requires special programming.
1. Introduction

My wise old professor used to say “If someone sells you a regression result much different from OLS, be suspicious - very suspicious.” And there is considerable merit in these words. In an interesting paper, Hoff (2007) advocates using tobit and ordinary least squares (OLS) in second stage data envelopment analysis (DEA) efficiency analyses stating “It is firstly concluded that the tobit approach will in most cases be sufficient in representing second stage DEA models. Secondly it is shown that OLS may actually in many cases replace tobit as a sufficient second stage DEA model.” In this paper, I come to a similar conclusion about OLS (although using a quite different argument), but advocate not using tobit. Let me set the scene.

It is common to analyse efficiency in two stages. Stage 1 is to use non-parametric DEA to calculate the efficiency with which output is produced from physical inputs.\(^1\) Stage 2 uses regression to relate efficiency scores to factors seen to influence efficiency. Some procedures have been developed that incorporate the influence of efficiency factors in the DEA analysis (see Cooper et al., 2000; Coelli et al., 1999; Fried et al., 1999; Grosskopf, 1996), but the two-stage procedure is very appealing both in terms of its simplicity and the way efficiency is described and interpreted. A Google search reveals hundreds of studies. Often at stage 2, the regression procedure used is two-limit tobit (2LT) with limits at zero and unity. Researchers who have used tobit at stage 2 include Bravo-Ureta et al. (2007), Latruffe et al. (2004), Fethi et al. (2002), Vestergaard et al. (2002), Ruggiero and Vitaliano (1999), Chilingerian (1995), Oum and Yu (1994) and Bjurek et al. (1992).

Hoff has compared the within-sample prediction performance (or fit) of 2LT, OLS, a quasi-maximum likelihood estimation (QMLE) method proposed by Papke and Wooldridge (PW, 1996) and the unit-inflated beta model of Cook et al. (2000) in a case study (the fishery of Danish liners and gillnetters over six months in 2002). In this particular example, Hoff found that OLS performed at least as well as the other methods. Tobit and the PW methods performed about as well, and the unit-inflated beta model, poorly. It is good statistical practice to carry out diagnostic or misspecification tests on estimated models to assess whether the models are well-specified, but this was not done.

In this paper, second stage strategies are reassessed. First (in sections 2-8), I review Hoff’s arguments within the context of a censoring DGP. I then argue (in
section 9) that DEA efficiency scores are not generated by a censoring DGP.\textsuperscript{2} They are a particular kind of fractional or proportional data. Tobit can be appropriate when the dependent variable data are generated by a censoring DGP, but is inappropriate when the data are fractional data. But, happily, OLS is an unbiased, consistent estimator, and, if heteroskedasticity is allowed for, (large sample) hypothesis tests can be validly undertaken. A careful OLS analysis will often be sufficient. For a more refined analysis, the gold standard is the QMLE procedure based on a Bernoulli log-likelihood function proposed by PW (1996). It is an asymptotically efficient method (within a broad class of estimators), but requires special computer programming and demands greater statistical expertise. For many applied researchers, familiar and easy to compute, OLS may be the way to go.

Throughout the paper when referring to DEA, I will deal with the single output, output-oriented case. In all sections except section 11, I will treat the DEA scores in the stage 2 analysis as descriptive measures of the relative efficiency of units in the sample (as, implicitly, Hoff and most applied researchers do). In section 11, I review recently published results on stage 2 analyses, when in stage 2, the scores are regarded as estimates of 'true' scores (relative to a 'true' frontier).

2. The two-limit tobit method

At stage 1, DEA is used to estimate frontier output given the physical input quantities and chosen production characteristics. In the example considered later, in the stage 1 DEA analysis, there is a single output, the analysis is output-orientated and constant returns to scale and strong free disposability of inputs were assumed. The production units were estates in a single county of England. Efficiency scores for production units were defined as the ratio of actual to the frontier value of (the net value of) output, and inputs consisted of three classes of capital, four categories of labour and three categories of land. The study is discussed in detail in McDonald (1997) and (1998). Efficiency scores must lie between 0 and 1 or equal 0 or 1. There are usually several values at 1, but often none at or close to 0.

If 2LT is used at stage 2, the unobservable latent or underlying regression is

\[ y_i^* = x_i \beta + \varepsilon_i, \]  

(1)

where the \( \varepsilon_i, x_i \) are normally, identically and independently distributed with mean, zero, and variance, \( \sigma^2 \), \( x_i \) is a 1 x k vector of observations on the constant and k-1 efficiency factor explanatory variables and \( \beta \) a k x 1 vector of unknown coefficients.

If \( y_i^* \leq 0 \), the efficiency score for the \( i \)th production unit, \( y_i = 0 \),
if \( y_i' \geq 1, y_i = I, \)

and if \( 0 < y_i' < 1, y_i = y_i' \)

The DGP postulates that the observed efficiency scores, \( y_i \), are the censored values of \( y_i' \), with censoring below zero and above one.

3. The structure of the likelihood

Given (1) is the DGP, the likelihood for a sample containing some \( y_i \)-observations = 0, some = 1, and some between 0 and 1 can be written:

\[
L = \prod_{y_i=0} \text{prob}(y_i = 0) \prod_{y_i=1} \text{prob}(y_i = 1) \prod_{0 < y_i < 1} f(y_i'),
\]

(2)

where \( f(y_i') \) is the density function of \( y_i' \), i.e., in this case, the normal density function.

If there are no \( y_i \)-observations = 0, then the first term will not appear in the likelihood function, and the likelihood functions for 2LT and one-limit tobit (1LT), with a limit at one, will be identical, and, consequently, the maximum likelihood estimates (MLE) of \( \beta \) and \( \sigma^2 \) for these methods, identical.\(^3\)

If there are no \( y_i \)-observations = 0 or 1, the first two terms will not appear and MLE are obtained by maximising the third term alone. This results in the OLS estimator, so for this case the 2LT and 1LT MLE and OLS estimates are identical.

But in the 2LT model, the MLE do not give the marginal (partial) effects of a change in the mean value of \( y_i/x_i \) with respect to a change in \( x_i \), the main focus of attention.

4. Two interpretations of marginal effects

In (2007, eq.10), Hoff reports the marginal effect in the 2LT model, commenting ‘That is the effect of the \( m \)th explanatory variable is a function of all explanatory variables as well as of all tobit regression parameters’.

While this is true, it may have been more pertinent to have indicated that it is a somewhat special combination of these quantities. The equation indicates the well-known result, see for example, Greene (2008, p.872-3), that the marginal effect with respect to the \( m \)th explanatory variable,
\[
\frac{\delta E(y_i|x_i)}{\delta x_{im}} = \beta_m, \text{(the probability that } y_i / x_i \text{ takes a non-limit value)},
\]

so the marginal effects are, in absolute value, less than or equal to the coefficient \( \beta_m \) values. Notice that all marginal effects are reduced in value by the same proportion. (3) is true whether or not the \( \epsilon_i / x_i \) in (1) are normally distributed.

If the probability that \( y_i \) takes a limit value is small, marginal effects will be similar to \( \beta \)-values. If there are no \( y_i = 0 \) observations, although 1LT and 2LT MLE are identical, the marginal effects are different. 2LT imposes the restriction that \( y_i \) cannot be less than zero, while 1LT does not, so the estimated probability that \( y_i \) takes a non-limit value is smaller for 2LT.

If there are no \( y_i = 0 \) or \( y_i = 1 \) observations, marginal effects will again be different. But the fewer limit values in the sample, the closer the 2LT and 1LT MLE will be to OLS and the closer we might expect the estimated probability that \( y_i \) takes a non-limit value would be to 1. Consequently, if there are not many \( y_i \)-limit values, we might expect 2LT and 1LT marginal effects to be similar to OLS marginal effects.

For the values of the explanatory variables in the sample \( (x_i, i = 1,2...n) \), an indication of the average probability that the efficiency scores, \( y_i \), equal the limit values is given by the relative frequency of observed score limit values and values in intervals close to the limit points.

A second useful interpretation is the decomposition of tobit marginal effects of McDonald and Moffitt (1980). This shows that a change in \( x_{im} \) has two effects. It affects the conditional mean of \( y_i \) in the non-limit part of the distribution and also the probability that the observation will fall in the non-limit part of the distribution.\(^4\)

5. **Imposing a limit at zero**

Although there may be instances when a limit should be imposed at zero, in many applications, there are no zero efficiency scores and very few, if any, close to zero. In these cases will 2LT be a misspecification as Hoff claims? The above analysis indicates that the 2LT and 1LT MLE will be identical, but the marginal effects different. The 2LT marginal effects incorporate information that \( y_i \geq 0 \) when calculating the probability that \( y_i \) takes a non-limit value, whereas 1LT does not. There is no specification error and indeed 2LT uses more \textit{a priori} information than 1LT in calculating marginal effects and hence might be expected to be more asymptotically efficient.
Even so, when there are no zero scores or scores close to zero, ignoring the non-negative characteristic of the scores will have little effect, and the 1LT estimates should provide a good approximation.  

6. Tests for heteroskedasticity and non-normality

As Arabmazer and Schmidt (1984) point out, if in (1), the $\varepsilon_i / x_i$ are not normal or they are heteroskedastic, tobit MLE are not consistent estimators. It is important then, to carry out diagnostic or misspecification tests on estimated tobit models – something that is often not done in stage 2 efficiency analyses, and was not considered by Hoff.

Two useful misspecification tests are a test for normality (of the $y_i$) and a test for heteroskedasticity. Both Pagan and Vella’s (1989) conditional moment test and Andrews (1988) chi-square test (an extension of Pearson chi-square testing to parametric models with covariates) provide a useful framework for carrying out these tests.

7. Procedures that allow for heteroskedasticity and non-normality

If heteroskedasticity is detected, Greene (2008, pp. 875-6) shows how to allow for multiplicative heteroskedasticity (of the form $\sigma_i = \sigma \cdot \exp(h_i \cdot a)$, where $h_i$ is a 1 x p vector of variables and $a$ is a p x 1 vector of unknown constants) in the tobit model.

If both heteroskedasticity and non-normality are detected, then a transformation of the dependent variable may be appropriate. A simple approach is to take the logarithm of the dependent variable observations and relate them either to the explanatory variable observations or the logarithm of the observations. A more sophisticated approach is to transform the dependent variable by a Box-Cox transformation (Box and Cox, 1964). Poirier (1978) has analysed the use of these transformations when the transformed dependent variable is censored or truncated normal. Alternatively, the Box-Cox extended method (in which both the dependent and explanatory variables are subject to a power transformation) could be applied.

If non-normality is detected, survival models that assume the latent or underlying variable has a Weibull, logistic, or generalised log-logistic (Burr, type II) distribution can be estimated, see Greene (1997, pp. 612-7). Another approach is the lognormal regression model, see Amemiya (1973). A third approach is to estimate the model with a distribution free method such as Powell’s (1984) least absolute error procedure.
It is also possible to test and estimate models in which the probability of a limit (frontier) value is determined by a different relationship than the level of a non-limit (sub-frontier) value. Cooke et al.’s (2000) unit-inflated beta method allows for this, but it can also be carried out within the tobit context (see Cragg, 1971), when the $\epsilon_i / x_i$ in (1) are not normally distributed, when the Box-Cox or Box-Cox extended methods are used, and when the OLS or a QMLE method, such as that advocated by PW, is employed (see McDonald, 2008).

8. Some other considerations

Hoff compares the marginal effects of different procedures, but his argument for using tobit and OLS is that they provide a superior (or at least as good) fit. In evaluating these comparisons it is well to remember that, for limited dependent variable and discrete choice models, the best way of measuring fit is not obvious and naive methods can often be constructed that out-perform more appropriate procedures, particularly in unbalanced data situations (see for example Greene, 2008, p.792, for a discussion in the binary choice situation). The latter is partly because, in these models, MLEs and QMLEs do not optimise fit. OLS does maximise a fit criterion. It is the linear estimator that maximises $R^2$, so it is perhaps not surprising that Hoff finds it produces the best fit in four of the six months data he examines.

In many stage 2 efficiency analyses fit or prediction is not the key focus. The main interest is to discover which factors influence efficiency (significantly) and the size of factor marginal effects. There is some evidence that in limited dependent variable and choice situations, although the parameter estimates of alternative methods differ, the main inferences and marginal effects are often similar (see, for example, Greene, 2008 pp. 781-3 for binary choice models, pp. 873-4 for limited dependent models and p. 876 for heteroskedasticity in limited dependent models).

9. Efficiency scores are not censored or corner solution data

Sections 2-8, hopefully, clarify some issues raised by Hoff, but the key message is that efficiency scores are not censored values, the censoring model (1) does not describe how their values were generated, and, consequently, tobit is an inappropriate estimator.

Censored regression models, such as tobit, have been used to model situations when the dependent variable data have been censored or are the outcome of an optimization problem for which there is a corner solution. An example of censored data are wealth data for which wealth beyond some threshold value, say $200,000,
were recorded as $200,000. An example of corner solution data are the charitable contributions made by a family as a result of (conceptually) solving a utility optimization problem subject to a budget constraint. The size of the contribution depends on the parameter values of the utility function and budget constraint. For all utility values below a certain level the contribution will be zero. In this sense, zero is a corner solution.

Wooldridge (2002, p.517-521) shows that both of these situations can be modelled as censoring DGP similar to (1) of this paper and equation (4) of Hoff’s paper (but with censoring at different values). In the censored data example, $y_i^c$ corresponds to actual wealth and $y_i$ to the censored wealth data and, in the charitable contributions example, $y_i^c$ corresponds to the family utility value and $y_i$ to the charitable contribution.

The efficiency score data are not censored or corner solution data. The efficiency score generating process can better be described as a normalisation process. The DEA generates a production frontier using the production unit input-output data and the DEA assumptions (involving, e.g., returns to scale and disposability of inputs). In output-oriented analyses, a production unit’s efficiency score is determined as its actual output divided by the frontier output corresponding to the unit’s input values. This process normalises the maximum efficiency score to be one and all efficiency scores to lie on or within the unit interval. Although there may be multiple scores of one, there is no censoring. The process generates a particular kind of fractional or proportional data, or, if scores are multiplied by 100, percentage data.

The regression dependent variable, then, is not censored or corner solution data, but fractional data. A more suitable DGP for the scores would be the linear unit interval model,

\[ y_i = x_i \beta + u_i, \] (4)

where the $u_i/x_i$ are independently distributed with zero means, $0 \leq y_i \leq 1$, with the limit point $y_i = 1$ possessing positive probability.

(4) implies that, when $y_i = 1$, $u_i = 1 - x_i \beta$, with probability $= 1$ - probability that $y_i < 1$.

When $y_i < 1$, $-x_i \beta \leq u_i < 1 - x_i \beta$. 
The variance of $\frac{u_i}{x_i}$ involves a term, relating to the limit point $y_i = 1$, which depends on $x_i$. Consequently, the variance will usually depend on $i$ and the $\frac{u_i}{x_i}$ will usually be heteroskedastic.

Tobit is a fragile estimation procedure, inconsistent given DGP (1) if the $\frac{\varepsilon_i}{x_i}$ are not normal or heteroskedastic. Unless the efficiency scores are generated by a very special form of (4), 1LT and 2LT will provide inconsistent estimates.

The properties of OLS, given the data are generated by (4), parallel those of OLS in the linear probability binary discrete choice model (discussed by, for example, Greene, 2008, pp.770-793 and Judge et al., 1988, pp.753-768). OLS estimates of $\beta$ are consistent and asymptotically normal under general conditions, and hypothesis tests can be validly carried out if allowance is made for heteroskedasticity. One way of achieving the latter is to calculate standard errors using White’s (1980) method, an option available in many statistics packages. Large sample tests do not require that disturbances are normally distributed, but are valid given a range of distribution assumptions.

As with the linear probability model, there are some problems with DGP (4). First, unless the ranges of the $x_i$ are restricted, the expression for the probability that $y_i = 1$ (or $y_i$ equals an interval of values) may lie outside the unit interval. If all the explanatory variables are binary, this is not a problem. If some are continuous, then it is still possible to proceed with caution, bearing in mind that (4) will only be an approximation over a range of $x_i$ -values. Secondly, (4) forces individual explanatory variable marginal effects (the elements of $\beta$) to be constant over the range of their explanatory variable values, but, if this is considered an issue, the linear model can be augmented by non-linear functions of the explanatory variables (e.g., the squares of their values).

Some problems in using OLS to estimate $\beta$ in (4) are that it is not an asymptotically efficient estimator, and OLS estimation does not guarantee that an estimate of $x_i\beta$ lies in the unit interval, so estimated $y_i$ -values and predicted $y_i$ -values may not lie in that interval. The latter problem could be mitigated, to a degree, by applying inequality-restricted least squares, similar to that suggested in the binary choice model, but it is unclear this would be advantageous (see Judge et al., 1988, pp.759-761 and Greene, 2008, p.773, ft.2).

One way of overcoming some of the difficulties of DGP (4) is to transform $x_i\beta$ to produce models similar to logit and probit in the binary choice situation. The logit-like transformation of (4) would be $y_i = G(x_i\beta) + u_i$, where $G(.)$ is the logistic
cumulative distribution function. These models can then be estimated by non-linear least squares allowing for heteroskedasticity.\textsuperscript{9}

PW’s (1996) QMLE approach is an attractive alternative that provides robust estimates and testing procedures which are asymptotically efficient within a class of estimators (essentially all weighted non-linear estimators). It involves maximising a Bernoulli log-likelihood function and generating a robust variance matrix of coefficient estimates using equations (5), (7), (8) and (9) of their paper. PW focus on a logistic form of the estimator. On page 628, they describe how to calculate marginal (partial) effects but do not indicate how to calculate standard errors for these effects. Standard errors, evaluated at explanatory variable sample means, can be calculated from the marginal effects variance matrix, which (in their notation) is,

\[ AM = \hat{g}^2 (I_K + (1 - 2\hat{G})\hat{\beta}\bar{x})V (I_K + (1 - 2\hat{G})\bar{x}^T\hat{\beta}^T) , \]

where \( V \) is the variance matrix of the QMLE, \( \hat{\beta} \), (given in equation (9) of PW’s paper), \( \hat{g} \) and \( \hat{G} \) are evaluated at explanatory variable sample mean values and \( \bar{x} \) is the 1 x k vector of explanatory variable sample means.

10. An example

In the example referred to in section 2, at stage 1, (net) outputs of estates were related to 10 inputs. Efficiency scores, the ratios of actual to frontier output, were calculated by DEA. 96 or 17% of the 577 estates produced on the frontier (efficiency score = 1). The quartiles of the distribution were Q1, 0.435, median, 0.626 and Q3, 0.861. The lowest score was 0.159. These score distribution summary measures suggest that, in a stage 2 analysis, 2LT and 1LT marginal effects will be similar.

In the illustrative stage 2 analysis the efficiency scores were related to five factors seen as possibly affecting production efficiency. These were whether the estate was close to an urban centre (Urban), the size of the estate (Size), the grazing/arable mix of agriculture (Ag-mix), the tenure arrangement (Tenure), and the soil region location of the estate (whether or not it lay in the fertile loam soil area, Soil).

Table 1 lists 1LT, 2LT and OLS marginal effects and p-values that the factors significantly influenced efficiency.\textsuperscript{10} The 1LT and 2LT coefficient estimates were identical (because there were no zero score values). The key inferences (that the Size and Ag-mix variables were significant at the one percent level and the other variables not significant at the five percent level), are the same for 1LT, 2LT and OLS, and marginal effects for the significant variables similar. Note that the 2LT marginal effects are smaller in absolute value than the 1LT effects and the 1LT and 2LT
marginal effects smaller in absolute value than corresponding coefficient estimates. Equation (3) suggests that scaled least squares coefficient estimates (OLS coefficient estimates divided by the proportion of non-limit observations in the sample) might be good approximations to tobit coefficient estimates, and this seems to be the case\textsuperscript{11}.

Table 2 shows the 2LT and OLS marginal effects when the number of frontier observations is doubled (by duplicating the frontier observations, frontier observations now being 29\% of all observations) and trebled (38\%). The two methods again produce the same key inferences and, as expected, for the significant Size and Ag-mix variables, the 2LT and OLS marginal effects diverge as the proportion of frontier observations increases.

Table 1 also exhibits 2LT estimates when the errors are allowed to be multiplicatively heteroskedastic ($\sigma_i = \sigma \exp(h_i \alpha)$, with $h_i$ the 5 x 1 vector of observations on the five explanatory variables). The likelihood ratio test of $\alpha = 0$ (just) rejects homoskedasticity at the five percent significance level (the p-value was .049), but the key inferences are unchanged and the marginal effects for the Size and Ag-mix variables similar. More seriously, normality of the $y_i$ was rejected at the one percent significance level on Pagan and Vella’s conditional moment test, strongly suggesting that the 2LT (and 1LT) models are misspecified even when a censoring DGP is assumed.\textsuperscript{12}

Table 3 lists OLS estimates when p-values are calculated using heteroskedastic-consistent standard errors. Differences with the usual OLS p-values are minor. It also lists PW QMLE estimates. The main inferences are unchanged and the Ag-mix marginal effect is similar, but the Size marginal effect is somewhat larger.

11. Recent developments

Stage 1 DEA calculates efficiency scores for each unit in the sample. One interpretation (implicit in the analysis of the previous sections) is that the scores are simply descriptive measures of the relative performance of the units in the sample which can be treated in the same way as other regression variables in stage 2. Given this interpretation, the frontier can be viewed as an observed (within the sample) best practice construct. The fact that the score calculations are complex does not undermine this interpretation, indeed, economic data are often complex creations, good examples are macroeconomic aggregate observations used to estimate macro-models.
A second interpretation is that the scores are estimates of 'true' scores relative to a 'true' frontier and when the properties of stage 2 estimates are evaluated this should be taken into account. Recent, research embracing this interpretation includes the work of Banker and Natarajan (2008), Simar and Wilson (2007), Cazals et al. (2002) and Daraio and Simar (2005). The contribution of Banker and Natarajan is particularly impressive. They build on the path breaking paper of Banker (1993) in which Banker provided a formal statistical basis for DEA. He derived conditions under which DEA estimators are MLEs and consistent estimators. He also showed that the asymptotic distribution of the DEA estimators of efficiency deviations is identical to the true distribution of these deviations and developed large sample testing procedures of hypotheses (e.g., that one kind of production unit is more efficient than another). This paper represents a considerable advance in knowledge.\textsuperscript{13}

Banker and Natarajan (2008) is also a very important paper. The authors provide a statistical foundation for two-stage analyses. They derive conditions under which two-stage procedures, consisting of DEA in stage 1 and OLS in stage 2 and DEA in stage 1 and MLE in stage 2, produce consistent estimators at stage 2. In their formulation, deviations from the production frontier may be due to movements in efficiency factors (which they refer to as contextual variables), the inefficiency of production units, or (two-sided) random noise. These three deviation components are assumed to be independent. Their results require that data are generated by a monotone increasing and concave production function which is separable from the stage 2 efficiency factor function. Separability requires that the efficiency factor variables are independent of the stage 1 input variables, although the input variables may be correlated with each other and the efficiency factor variables may be correlated with each other. Hypothesis testing of stage 2 parameters is not discussed.

Banker and Natarajan provide simulation evidence supporting two-stage procedures with DEA in stage 1. The evidence relates to parameter point estimates only. DEA based procedures with OLS, MLE or tobit in stage 2 performed as well in estimating efficiency factor effects and production unit efficiency as the best of the 12 one and two-stage parametric methods examined.

Hopefully, Banker and Natarajan's approach can be extended to deal with situations when estimation methods other than OLS and MLE are employed in stage 2 and efficiency factor variables are correlated with production function inputs. In addition, large sample testing procedures of stage 2 parameters need to be developed for all stage 2 estimation procedures and simulation evidence assembled.
Simar and Wilson's (2007) introduce a separability condition that implies that efficiency factors (called environmental variables) do not influence the frontier but can influence the efficiency scores of units. Their DGP is much more restrictive than that of Banker and Natarajan (2008) in that it does not allow for a two-sided noise term and the production unit efficiency terms are unit-specific truncated normal distributions. Simar and Wilson advocate a very complex seven-stage estimation procedure with double bootstrapping, which may be valid given their chosen DGP, but is not robust to plausible departures from it – in particular, that (true) efficiency scores in the second stage equation are unit-specific, truncated, normal random variables. Their procedure generates interval estimates that can be used to undertake large sample hypothesis tests of stage 2 parameters.\textsuperscript{14}

While progress has been made in this research program, problems remain in using the results in applied studies. Banker and Natarajan's results only relate to point estimation (not testing) and Simar and Wilson's DGP is a very special case, their results are not robust to departures from it and they do not provide misspecification tests.

Even if the theoretical problems surrounding inference can be solved, there are arguments for continuing to treat scores as descriptive measures at stage 2. Returning to the example of macroeconomic modelling, macroeconomic aggregate observations can be regarded as estimates of 'true' values of the variable at a point or period in time and it can be argued that macro-model inference should take this into account. Some reasons why this is usually not done include: it is unclear how to modify inference, it is thought it would lead to considerable complexity and perhaps only minor changes in inference, and it is thought that current techniques have been successful in providing insights and predictions. Similar arguments can be advanced for treating scores as descriptive measures at stage 2.\textsuperscript{15}

In epistemological terms the research program to treat scores as estimates of 'true' scores at stage 2 can be seen as a 'conventionalist' argument. 'Conventionalists' maintain that truth cannot be established by 'inductive' argument and regard truth as a matter of convention. They judge arguments, theories or procedures by criteria of convenience, such as simplicity, generality, degree of approximation or closeness of fit. The idea that efficiency scores should be treated as estimates of 'true' scores satisfies an abstract criterion of this kind. But theoretical niceties are of little concern to 'instrumentalists' who are ultimately only interested in the usefulness or predictive success of theories or procedures. There is considerable evidence testifying to the success of two-stage studies in which scores are treated as descriptive measures.
Hundreds of two-stage DEA studies have proven very useful in gaining insights into real world production processes and this knowledge has then been successfully conveyed to a wide audience. A strong 'instrumentalist' argument can be mounted to justify the descriptive interpretation of scores.16

12. Conclusion

There are good arguments for treating DEA efficiency scores as descriptive measures in second stage analyses.

The efficiency scores are not generated by a censoring process. They are fractional data. Stage 2 DEA efficiency analyses should be analysed using a DGP such as (4) that describes the dependent variable as fractional data. Given (4), tobit is an inappropriate estimation procedure. It is, in general, an inconsistent estimator and the best that can be said for it is that tobit estimates are often similar to OLS (and PW’s method) estimates. In contrast, OLS is a consistent estimator, and, if White’s heteroskedastic-consistent standard errors are calculated, tests can be performed which are valid for a range of disturbance distribution assumptions.

In this and PW’s study, OLS and PW’s QMLE produced similar inferences. PW’s procedure has the advantage over OLS that it is asymptotically more efficient, but it requires special programming and demands greater statistical expertise of researchers and those assessing the results of researchers. There is considerable merit in using familiar, easy to compute methods, such as OLS, which are understood by a broad community of people.

Acknowledgement

I acknowledge a debt to my DEA mentors, Shawna Grosskopf, Rolf Fare and Knox Lovell, William H. Greene for his excellent text and LIMDEP program and two reviewers who provided very helpful comments. Thanks, also, to my good friends Keith Hancock and Bob Wallace for their generous gifts which funded the research through a Flinders Business School grant and Missy Muggles and Coco Plus for their caring assistance, and, oh yes, especial thanks to my wise old professor.
### Table 1. One-limit tobit (1LT), two-limit tobit (2LT) and OLS estimates of a stage 2 analysis of Essex estate efficiency on factors affecting efficiency

<table>
<thead>
<tr>
<th>Factor</th>
<th>1LT marginal effect (p-value)</th>
<th>2LT marginal effect (p-value)</th>
<th>OLS marginal effect (p-value)</th>
<th>OLS coefficients $\beta_m$ (p-value)</th>
<th>1LT/2LT marginal effect</th>
<th>OLS scaled LS marginal effect (p-value)</th>
<th>2LT allowing for heteroskedasticity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Urban</td>
<td>-.0122 (.983)</td>
<td>-.0120 (.965)</td>
<td>-.0121 (.623)</td>
<td>-.0127 (.661)</td>
<td>-.0145 (.623)</td>
<td>-.0226 (.909)</td>
<td></td>
</tr>
<tr>
<td>Size</td>
<td>.000732 (.000)**</td>
<td>.000718 (.000)**</td>
<td>.000608 (.000)**</td>
<td>.000764 (.000)**</td>
<td>.000729 (.000)**</td>
<td>.000762 (.000)**</td>
<td></td>
</tr>
<tr>
<td>Ag-Mix</td>
<td>.000201 (.000)**</td>
<td>.000198 (.000)**</td>
<td>.000153 (.000)**</td>
<td>.000210 (.000)**</td>
<td>.000184 (.000)**</td>
<td>.000271 (.000)**</td>
<td></td>
</tr>
<tr>
<td>Tenure</td>
<td>.0214 (.962)</td>
<td>.0210 (.918)</td>
<td>.0209 (.309)</td>
<td>.0223 (.357)</td>
<td>.0251 (.976)</td>
<td>.0138 (.976)</td>
<td></td>
</tr>
<tr>
<td>Soil</td>
<td>.0203 (.722)</td>
<td>.0199 (.663)</td>
<td>.00587 (.852)</td>
<td>.0212 (.569)</td>
<td>.00707 (.569)</td>
<td>.0078 (.976)</td>
<td></td>
</tr>
</tbody>
</table>

**Notes:** * indicates significant at the five percent and ** significant at the one percent level. There were 577 production units, with 96 or 17% of the observations limit or frontier observations. Tobit marginal effects were calculated at explanatory variable sample means. OLS marginal effects are the OLS coefficient estimates. Scaled LS estimates are OLS marginal effects divided by the proportion of non-limit observations in the sample.
<table>
<thead>
<tr>
<th>Factor</th>
<th>Frontier observations</th>
<th>2LT</th>
<th>OLS</th>
<th>Frontier observations</th>
<th>2LT</th>
<th>OLS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>doubled</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urban</td>
<td>-.00756</td>
<td>-.00838</td>
<td>(.992)</td>
<td>-.00368</td>
<td>-.00532</td>
<td>(.817)</td>
</tr>
<tr>
<td>Size</td>
<td>.000720</td>
<td>.000543</td>
<td>(.000)**</td>
<td>.000703</td>
<td>.000487</td>
<td>(.000)**</td>
</tr>
<tr>
<td>Ag-Mix</td>
<td>.000227</td>
<td>.000155</td>
<td>(.000)**</td>
<td>.000234</td>
<td>.000146</td>
<td>(.000)**</td>
</tr>
<tr>
<td>Tenure</td>
<td>.0310</td>
<td>.0297</td>
<td>(.961)</td>
<td>.0370</td>
<td>.0342</td>
<td>(.073)</td>
</tr>
<tr>
<td>Soil</td>
<td>.0558</td>
<td>.0302</td>
<td>(.304)</td>
<td>.0751</td>
<td>.0417</td>
<td>(.131)</td>
</tr>
</tbody>
</table>

**Notes:** * indicates significant at the five percent and ** significant at the one percent level. When the frontier observations were doubled, the total number of observations was 673, of which 192 or 29% were frontier observations. When the frontier observations were trebled, the total was 769, of which 288 or 38% were frontier observations. Tobit marginal effects were calculated at explanatory variable sample means. OLS marginal effects are the OLS coefficient estimates.
Table 3. OLS and PW Quasi-maximum likelihood (PW) marginal effect estimates of a stage 2 analysis of Essex estate efficiency on factors affecting efficiency

<table>
<thead>
<tr>
<th>Factor</th>
<th>OLS marginal effect (p-value)</th>
<th>PW marginal effect (p-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Urban</td>
<td>-.0121 (.635)</td>
<td>-.0116 (.656)</td>
</tr>
<tr>
<td>Size</td>
<td>.000608 (.000)**</td>
<td>.000928 (.000)**</td>
</tr>
<tr>
<td>Ag-Mix</td>
<td>.000153 (.000)**</td>
<td>.000159 (.000)**</td>
</tr>
<tr>
<td>Tenure</td>
<td>.0209 (.317)</td>
<td>.0114 (.599)</td>
</tr>
<tr>
<td>Soil</td>
<td>.00587 (.869)</td>
<td>.00279 (.897)</td>
</tr>
</tbody>
</table>

**Notes:** p-values are based on heteroskedastic-consistent standard errors. * indicates significant at the five percent and ** significant at the one percent level. There were 577 production units, with 96 or 17% of the observations limit or frontier observations. PW marginal effects were calculated at explanatory variable sample means. OLS marginal effects are the OLS coefficient estimates.
References


Friedman, M., 1953, The methodology of positive economics, in his Essays in positive economics, University of Chicago Press, 3-43.


McDonald, J., 2008. Allowing for different frontier and sub-frontier relationships in second stage DEA efficiency analyses. Flinders Business School, 19, Flinders University.


1 Hoff (2007) describes DEA.

2 Hoff's rationale for using tobit at stage 2 is that DEA scores “resemble corner solution variables”. When he calculates the tobit likelihood function he uses equation (4), p. 428 of his paper, which is the censoring DGP (1) of this paper.

3 For 1LT with a limit at one, the DGP is: If \( y_i^* \geq 1 \), \( y_i = 1 \) and if \( y_i^* < 1 \), \( y_i = y_i^* \).

\[
L = \prod_{y_i \leq 1} \text{prob}(y_i = 1) \prod_{y_i < 1} f(y_i^*).
\]

4 More specifically, the tobit marginal effect is equal to (the change in the conditional mean of \( y_i \), given that \( y_i \) takes a non-limit value) times (the probability that \( y_i \) takes a non-limit value) plus (the change in the probability that \( y_i \) takes a non-limit value) times (the conditional mean of \( y_i \), given that \( y_i \) takes a non-limit value), see McDonald and Moffitt (1980, eq. 5) for details that relate to the one-limit at zero tobit case.

5 In statistical analyses it is not unusual to ignore this kind of information. For example, the heights of male or female adult populations, the IQs of adults and exam marks of students are often regarded as normally distributed. The approximation is usually good, even though these variables, unlike the normal random variable, must be non-negative and bounded from above in the positive direction.

6 This limits the use of White’s asymptotic covariance estimator. As White indicated (1982, p.4) “it is the consistency of the QMLE for the parameters of interest in a wide range of situations which ensures its usefulness as a basis for robust estimation techniques.”

7 This contrasts with the MLE in the normal linear regression model. In that model, the MLE is the OLS estimator, so it maximises \( R^2 \).

8 The main conditions are that the \( u_i / X_i \) have finite variances and weak heterogeneity conditions are satisfied, so a central limit theorem applies, see White (1980).

9 The calculations can be made using probit or logit software that accepts y-values that are fractional values (as well as 0 or 1) and then calculating standard errors using White’s method.

10 For binary explanatory variables, marginal effects are the differential effect when the variables take their binary values (and other explanatory variables are set equal to their mean values). Table 1 gives estimated effects based on partial derivatives. These are often a good approximation. The significant variables, Size and Ag-mix, are not binary variables.

11 The approximation argument is as follows: (3) suggests setting the tobit marginal effect estimate = the tobit coefficient estimate.(an estimate of the non-limit probability). The scaled least squares estimate = the OLS coefficient estimate/(an estimate of the non-limit probability). But the estimated tobit and OLS marginal effects are similar and the estimated OLS marginal effect is the the OLS coefficient estimate, so, the scaled least squares estimate is approximately = the tobit marginal effect estimate/(an estimate of the non-limit probability), which is approximately = the tobit coefficient estimate.

12 The test for normality is strictly speaking a test for symmetry and mesokurtosis.

13 Korostelev et al. (1995) established similar results and investigate the rate of convergence of DEA estimators.

14 Given their unusual DGP, it is unclear what interpretation can be placed on Simar and Wilson's simulation evidence. Cazals et al. (2002) and Daraio and Simar (2005) provide an alternative approach. In these papers efficiency factors are introduced into the production process by defining a conditional efficiency measure and non-parametric stage 2 methods are advocated.

15 In many countries, macro aggregate variable observations are revised as more information becomes available. Some early studies that examine the effect of revisions on the properties of macro-variable
time series and regressions are McDonald (1974, 1975 and 1976).

Boland (1973) and Frazer and Boland (1983) review 'instrumentalism' and 'conventionalism' and the influence of 'instrumentalism' in economics via the thinking of Friedman (1953).