European Journal of Operational Research 197 (2009) 792–798

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Interfaces with Other Disciplines

# Using least squares and tobit in second stage DEA efficiency analyses

John McDonald *

*Flinders Business School, Flinders University, GPO Box 2100, Adelaide 5001, Australia*

ARTICLE INFO

*Article history:*
Received 4 March 2008
Accepted 28 July 2008
Available online 7 August 2008

*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 [White, H., 1980. A heteroskedastic-consistent covariance matrix and a direct test for heteroskedasticity. Econometrica 48, 817–838] 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 [Papke, L.E., Wooldridge, J.M., 1996. Econometric methods for fractional response variables with an application to 401($k$) plan participation rates. Journal of Applied Econometrics 11 (6), 619–632] 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.[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 Papke and Wooldridge

* Tel.: +61 882967171.
*E-mail address:* john.mcdonald@flinders.edu.au

[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 Eq. (4), p. 428 of his paper, which is the censoring DGP (1) of this paper.

0377-2217/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.ejor.2008.07.039

(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, 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

$$y_i^* = x_i\beta + \varepsilon_\imath, \tag{1}$$

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

$$\begin{aligned}&\text{If } y_i^* \leqslant 0, \text{the efficiency score for the } i\text{th production unit, } y_i = 0,\\ &\text{if } y_i^* \geqslant 1,\ y_i = 1,\\ &\text{and if } 0 < y_i^* < 1,\ y_i = y_i^*\end{aligned}$$

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:

$$L = \prod_{y_i=0} \text{prob}(y_i = 0) \prod_{y_i=1} \text{prob}(y_i = 1) \prod_{0<y_i<1} f(y_i^*), \tag{2}$$

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, pp. 872–873), that the marginal effect with respect to the $m$th explanatory variable,

$$\frac{\delta E(y_i/x_i)}{\delta x_{im}} = \beta_m \cdot (\text{the probability that } y_i/x_i \text{ takes a non-limit value}), \tag{3}$$

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 $\varepsilon_\imath/x_i$ 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 ($x_i$, $i = 1, 2, \ldots, n$), 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 $y_i \geqslant 0$ when calculating the probability that $y_i$ takes a non-limit

[3] For 1LT with a limit at one, the DGP is: if $y_i^* \geqslant 1$, $y_i = 1$ and if $y_i^* < 1$, $y_i = y_i^*$. $L = \prod_{y_i=1} \text{prob}(y_i = 1) \prod_{y_i<1} f(y_i^*)$.

[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.