Journal of Econometrics 18 (1982) 275–279. North-Holland Publishing Company

A STATIONARY POINT FOR THE STOCHASTIC FRONTIER LIKELIHOOD

Donald M. WALDMAN*

University of North Carolina, Chapel Hill, NC 27514, USA

Received June 1980, final version received September 1981

The likelihood function for the stochastic frontier model is shown to possess an unusual stationary point which may or may not be a maximum. A condition is given to determine if the point is a maximum, and the result is interpreted in the context of specification and estimation.

1. Introduction

The specification and estimation of frontier or envelope functions has been a concern of econometricians for some time. The subject became of widespread interest with the introduction of a particularly attractive formulation of the model, the stochastic frontier of Aigner, Lovell and Schmidt (ALS) (1977) and Meeusen and Van den Broeck (1977). In that model, the random disturbance is composed of the sum of a symmetric and a one-sided random variable. This specification avoids the statistical problems of earlier formulations, while providing a convenient measure of the relative inefficiency. A recent survey may be found in Førsund et al. (1980).

The distributional assumptions most commonly made are the zero-mean normal distribution for the symmetric component and the same distribution truncated above zero for the one-sized component.[1] With this assumption, the likelihood equation is well defined, and ALS propose maximum likelihood estimation. This method has been widely applied [for examples, see (3), (4) and (5)]. However, recent experience with simulating the properties of competing estimators for this model has indicated that the likelihood function is not entirely well behaved.[2] We show that the likelihood function *always* has a stationary point at one particular set of parameter values.[3] A condition is given when this point is a local maximum and when it is a

*The author wishes to thank C. Bromley, D. Guilkey, R. Keener, C.A.K. Lovell, J. Olsen, P. Schmidt and V.K. Smith for helpful insights and discussions.

[1]This would be the case for a 'positive' frontier, i.e., a production function. For a cost function, the one-sided component would be truncated below zero.

[2]See Olsen, Schmidt and Waldman (1980). This point is elaborated in section 4.

[3]This was first noted, so far as we are aware, by Trevor Breusch.

saddle point. Existence of this point is troublesome since it involves a value of zero for the variance of the one-sided component of the disturbance, indicating no inefficiency relative to the frontier. Finally, we discuss the implications of the existence of this stationary point for the estimation of frontier models.

2. The model and the stationary point

The model ALS consider is the linear regression, for T observations

$$y_t=\boldsymbol{\beta}'\boldsymbol{x}_t+\varepsilon_t, \qquad \varepsilon_t=v_t-u_t,$$

where $(y_t, \boldsymbol{x}_t')$ is a $1\times(K+1)$ vector of output and inputs (the first element of $\boldsymbol{x}_t$ is a '1'), $\boldsymbol{\beta}$ is a $K\times 1$ vector of parameters, $v_t\sim N(0,\sigma_v^2)$ and u_t is the absolute value of a zero-mean normally distributed random variable with variance σ_u^2. All the v_t and the u_t are independent of one another and of the $\boldsymbol{x}_t$.

A convenient reparameterization of the two variances adopted by ALS is

$$\sigma^2=\sigma_u^2+\sigma_v^2, \qquad \lambda=\sigma_u/\sigma_v.$$

Thus the parameter vector is $\boldsymbol{\theta}'=(\boldsymbol{\beta}',\lambda,\sigma^2)'$. Note that λ represents the ratio of the standard deviation of the variance of (the distribution underlying) the one-sided component to that of the symmetric component, and hence is non-negative. The log-likelihood function (eq. 10 in ALS) is

$$l(\boldsymbol{\theta})=T\ln\sqrt{2/\pi}+T\ln\sigma^{-1}+\sum\ln(1-F_t)-\frac{1}{2\sigma^2}\sum\varepsilon_t^2, \tag{1}$$

where F_t is the standard normal cumulative distribution function evaluated at $\varepsilon_t\lambda\sigma^{-1}$. All summations above and in the sequel are over the set of T observations.

A stationary point of (1) involves values for $\boldsymbol{\beta}$, λ, and σ^2 for which the first derivatives of (1) (see ALS, eqs. 11–13) are equated to zero. This must be true at the maximum likelihood solution, but it is also true for the point $\boldsymbol{\theta}^*=(\boldsymbol{b}',0,s^2)$, where

$$\boldsymbol{b}=\left(\sum \boldsymbol{x}_t\boldsymbol{x}_t'\right)^{-1}\sum \boldsymbol{x}_t y_t,$$

$$s^2=\frac{1}{T}\sum(y_t-\boldsymbol{b}'\boldsymbol{x}_t)^2=\frac{1}{T}\sum e_t^2,$$

the ordinary least squares estimates. This is because $\lambda=0$ implies the derivative of (1) with respect to λ vanishes and the derivatives with respect to

$\boldsymbol{\beta}$ and σ^2 simplify to the least squares normal equations. We examine the behavior of $l(\boldsymbol{\theta})$ at $\boldsymbol{\theta}^*$.

3. Analysis of $\boldsymbol{\theta}^*$

The matrix of second-order derivatives, as given by ALS (1977, app.), evaluated at $\boldsymbol{\theta}^*$ is

$$H(\boldsymbol{\theta}^*)=\begin{bmatrix} -s^{-2}\sum \boldsymbol{x}\boldsymbol{x}_t' & s^{-1}\sqrt{2/\pi}\sum \boldsymbol{x}_t & \boldsymbol{0} \\ s^{-1}\sqrt{2/\pi}\sum \boldsymbol{x}_t & -2T/\pi & 0 \\ \boldsymbol{0}' & 0 & -T/2s^4 \end{bmatrix}.$$

Unfortunately $H(\boldsymbol{\theta}^*)$ does not directly determine the behavior of $l(\boldsymbol{\theta})$ in the neighborhood of $\boldsymbol{\theta}^*$ since $|H(\boldsymbol{\theta}^*)|=0$. $H(\boldsymbol{\theta}^*)$ is non-positive definite, with $K+1$ negative characteristic roots and one zero root. The only interesting direction to examine $l(\boldsymbol{\theta})$, therefore, is in the direction determined by the characteristic vector associated with the zero root. This vector is

$$z'=(s\sqrt{2/\pi},\boldsymbol{0}',1,0).$$

where $\boldsymbol{0}$ is a $(K-1)\times 1$ vector of zeroes. We are interested then, in the sign of

$$\Delta l = l(\boldsymbol{\theta}^*+\mu z)-l(\boldsymbol{\theta}^*),$$

for arbitrarily small, positive μ. Since the parameter space is constrained by $\lambda \geqq 0$ and we have written the characteristic vector z to have the same sign as λ, we need search only in the direction $\mu>0$.

From (1) we have

$$l(\boldsymbol{\theta}^*)=T\ln\sqrt{2/\pi}-T\ln s - T\ln 2 - T/2,$$

and

$$l(\boldsymbol{\theta}^*+\mu z)=T\ln\sqrt{2/\pi}-T\ln s$$
$$+\sum \ln\left[1-F\{e_t\mu s^{-1}-\mu^2\sqrt{2/\pi}\}\right]-T/2-\mu^2 T/\pi,$$

where $F(\cdot)$ is the c.d.f. of the standard normal variate. Here we have made use of the fact that $\sum e_t=0$, $\sum e_t^2=Ts^2$, and $\sum e_t x_t=0$. Therefore,

$$\Delta l = -\mu^2 T/\pi + T\ln 2+\sum \log\left[1-F\{e_t\mu s^{-1}-\mu^2\sqrt{2/\pi}\}\right]$$
$$=-\mu^2 T/\pi+\sum \ln\left[2-2F\{e_t\mu s^{-1}-\mu^2\sqrt{2/\pi}\}\right]. \tag{2}$$

Using Taylor's Theorem, we have the following two series:

$$F(x)=\tfrac{1}{2}+x/\sqrt{2\pi}-x^3/6\sqrt{2\pi}+\ldots,$$

$$\ln(1-x)=-x-x^2/2-x^3/3-\ldots \quad \text{for} \quad |x|<1.$$

Specializing these results to the problem and substituting into (2) yields, after some straightforward algebra,

$$\Delta l=(\mu^3/6s^3)\sqrt{2/\pi}\,[(\pi-4)/\pi]\sum e_t^3+0(\mu^4).^4 \tag{3}$$

Since $(\pi-4)/\pi<0$, Δl has locally the opposite sign as $\sum e_t^3$. If $\sum e_t^3<0$, the likelihood will be increased by moving away from $\boldsymbol{\theta}^*$ in the direction of z. However, $\boldsymbol{\theta}^*$ is a local maximum if $\sum e_t^3>0$.[5]

4. Implications

It is the case that

$$\operatorname{plim}((1/T)\sum e_t^3)=E[\varepsilon_t-E(\varepsilon_t)]^3=\sigma_u^3\sqrt{2/\pi}((\pi-4)/\pi),$$

which is negative for $\sigma_u>0$. Therefore as the sample size increases the probability that $\sum e_t^3>0$ and hence that $\boldsymbol{\theta}^*$ locates a local maximum goes to zero. In finite samples, the experience in simulations reported in Olsen, Schmidt and Waldman (1980) confirms that whenever $\sum e_t^3<0$, a local maximum of the likelihood is found at a point *other than* $\boldsymbol{\theta}^*$, and that whenever $\sum e_t^3<0$, no other local maximum exists.

For the purpose of estimation, this result suggests a strategy for the econometrician: having specified a stochastic frontier model, the third moment of the least squares residuals should be examined. If this quantity is positive, then it will always be the case, as proven above, that the least squares slope estimates and $\hat{\lambda}=0$ represent a local maximum of the likelihood. Further, the empirical evidence in Olsen et al. (1980) suggests that this point is a global maximum. If this quantity is negative the likelihood has a greater value at some other point where $\hat{\lambda}>0$.

[4]An alternative derivation of (3), suggested by T. Breusch, is to expand Δl directly by Taylor series. This process is simplified because the first term in the series drops out, since $\boldsymbol{\theta}^*$ is a stationary point, the second term drops out, since $H(\boldsymbol{\theta}^*)\ z=\boldsymbol{0}$, and only the third derivative terms corresponding to the non-zero elements of z need be evaluated.

[5]Note that this is true only under normality. ALS also propose the exponential distribution for u_t, although this is much less frequently adopted.

References

Aigner, D.J., C.A.K. Lovell and P. Schmidt, 1977, Formulation and estimation of stochastic frontier production function models, Journal of Econometrics 6, 21–37.

Førsund, F.R. and L. Hjalmarsson, 1979, Frontier production functions and technical progress: A study of general milk processing in Swedish dairy plants, Econometrica 47, 883–900.

Førsund, F.R., C.A.K. Lovell and Peter Schmidt, 1980, A survey of frontier production functions and of their relationship to efficiency measurement, Journal of Econometrics 13, 5–25.

Kopp, R.J. and V.K. Smith, 1978, The characteristics of frontier production function estimates for steam generating electric plants: An econometric analysis, Southern Economic Journal 47, 1049–1059.

Lee, L.F. and W.G. Tyler, 1978, The stochastic frontier production function and average efficiency: An empirical analysis, Journal of Econometrics 7, June, 385–390.

Meeusen, W. and J. van den Broeck, 1977, Efficiency estimation from Cobb–Douglas production functions with composed error, International Economic Review 18, 435–444.

Olson, J.A., P. Schmidt and Donald M. Waldman, 1980, A Monte Carlo study of estimators of stochastic frontier production functions, Journal of Econometrics 13, 67–82.