Applied Economics, 1997, **29**, 459–469

Stochastic production function estimation: small sample properties of ML versus FGLS

ATANU SAHA, ARTHUR HAVENNER and HOVAV TALPAZ

Micronomics Inc., Los Angeles, Department of Agricultural Economics, University of California, Davis, USA and *Department of Statistics, ARO, the Volcani Center, Bet Dagan, Israel*

Just–Pope production functions have been traditionally estimated by feasible generalized least squares (FGLS). This paper investigates the small-sample properties of FGLS and maximum likelihood (ML) estimators in heteroscedastic error models. Monte Carlo experiment results show that in small samples, even when the error distribution departs significantly from normality, the ML estimator is more efficient and suffers from less bias than FGLS. Importantly, FGLS was found to seriously understate the risk effects of inputs and provide biased marginal product estimates. These results are explained by showing that the FGLS criteria being optimized at the multiple stages are not logically consistent.

I. INTRODUCTION

Appropriate production function formulation under risk has received considerable attention in production studies. Just and Pope (1978) have argued that traditional forms such as the Cobb–Douglas, Translog, etc., impose a risk-increasing role on all inputs. They proposed an alternative formulation that allows the effects of inputs on the deterministic component of production to be different than on the stochastic component. This specification admitted risk-increasing as well as risk-reducing inputs. Since then, the Just–Pope production function has been widely used in the applied production literature. Example of such studies include Griffiths and Anderson (1982), Hallam *et al.* (1989), Hassan and Hallam (1990), Kumbhakar (1993), Love and Buccola (1991), McCarl and Rettig (1983), Wan and Anderson (1985).

Just and Pope have offered two alternative methods of estimating their stochastic production function. In their 1978 article, they suggest a maximum likelihood (ML) procedure and discuss an iterative method for obtaining the ML estimates. However, in their 1979 paper, they lay out a three step estimation procedure that is essentially feasible generalized least squares (FGLS) under heteroscedastic disturbances. A principal objective of this paper is to explore the small-sample properties of these two estimation procedures.

Applied production economists have overwhelmingly chosen the second method. In fact, every Just–Pope estimation study cited above has used FGLS. To date, we have not found a single study in this area that reports ML estimates. Our paper's findings suggest that the ubiquity of FGLS may be unwarranted. Monte Carlo experiment results show, unless the error distribution departs significantly from normality, the ML estimator is substantially more efficient with a considerably smaller mean squared error than FGLS. Since FGLS appears to be the most widely used procedure in applied consumption analysis as well (see, for example, Fomby *et al.* (1988) and the references therein), the implications of our findings extend beyond production function estimation.

II. RELEVANT LITERATURE AND MOTIVATION

The Just–Pope production function can be written as:

$$y_i = f(\mathbf{x}_i, \beta) + g(z_i, \alpha)\varepsilon_i \tag{1}$$

where y denotes output, $\mathbf{x}$ and $\mathbf{z}$ are vectors of inputs which may contain common elements, β and α are the corresponding parameter vectors, ε is a random variable distributed with zero mean and unit variance, and $f(\cdot)$ and $g(\cdot)$ are possibly nonlinear functions. The subscript i denotes the ith observation, $i = 1, \ldots, n$. The underlying error in (1) has been assumed to have unit variance for notational simplicity; the points made in this and the next section hold for the more general case. Also, this assumption is relaxed in the paper's Monte Carlo experiments.

The production function in (1) can also be interpreted as an estimation equation with heteroscedastic errors:

$$y_i = f(\mathbf{x}_i, \beta) + u_i \quad (2)$$

$$u_i = g(\mathbf{z}_i, \alpha)\varepsilon_i \quad (3)$$

where now y denotes the dependent variable, $\mathbf{x}$ and $\mathbf{z}$ are regressors, and u is the disturbance term with zero mean and variance: $V(u_i) \equiv \sigma^2_{u_i} = g(\mathbf{z}_i, \alpha)^2$. Consequently, Just–Pope production function estimation issues are an integral part of the literature on estimation under heteroscedasticity.

Small sample properties of FGLS *vis á vis* ML have received scant attention in the econometrics literature on heteroscedasticity. Since data sets available for production function estimation are often limited, small sample estimator properties are of interest.

Taylor's 1977 paper was the first to provide a rigorous analysis of the small sample properties of FGLS. He derived conditions under which a FGLS estimator is preferred to ordinary least squares (OLS) for linear models in which the eigenvectors of the error covariance matrix are known. His result showed that for a small number of distinct error variances, FGLS dominates OLS for samples with a moderate degree of heteroscedasticity. Taylor's study, though insightful, does not provide answers to our inquiry because it does not derive conditions under which ML dominates FGLS in small samples.

Amemiya (1985) examined the asymptotic properties of FGLS estimators in models where the error variance is a linear function of regressors, that is, $\sigma_i^2 = \mathbf{z}_i\alpha$. This specification is widely used and fairly general because $\mathbf{z}_i$ elements may be nonlinear transformations of independent variables. Amemiya proposed an estimator for α that was asymptotically more efficient under normality than the estimators proposed by two earlier studies of Hildreth and Houck (1968), and Goldfield and Quandt (1972).

Magnus extended Amemiya's analysis by considering models where error variances depend upon an unknown parameter vector and a possibly nonlinear function, i.e. $\sigma_i^2 = g(\mathbf{z}_i, \alpha)$. He proposed simultaneous estimation of β and α through maximum likelihood and derived the properties of ML estimators. Based on Oberhofer and Kmenta's (1974) 'zig-zag' method, Magnus (1978) proposed an algorithm that leads to ML estimates.

Jobson and Fuller (1980) considered the case where $\sigma_i^2 = g(\mathbf{z}_i, \alpha, \beta^0)$, with β^0 being a subset (possibly whole) of the regression parameters β. Clearly, Jobson and Fuller's model is considerably more general than the ones in prior studies. They developed a FGLS procedure for estimating the parameters and presented their asymptotic properties, in particular, consistency and asymptotic normality. They also derived the asymptotic properties of ML estimators and showed that the FGLS estimate of β has the same asymptotic distribution as the ML estimate.

It is apparent from this brief review of the relevant studies that, while the asymptotic properties of FGLS versus ML estimators have been analysed and the asymptotic equivalence of the two methods demonstrated, their comparative properties in finite samples have remained unexplored. Consider, for example, the following citations from two leading econometric texts:

> Though the large sample properties of the feasible generalized least squares estimators in heteroscedasticity models are well defined, their small sample properties are not (Fomby *et al.*, 1988, p. 201).
>
> … most recommended estimators of β have identical asymptotic properties. Thus … there is no clear basis for choice — (it) seems reasonable to use any [FGLS] estimator for β … or to use maximum likelihood estimation. (Judge *et al.*, 1985, p. 455).

Thus, the existing evidence appears to provide 'no clear basis for choice' between ML and FGLS. In fact, we are unaware of any study to date that has examined the comparative performance of these two estimators in finite samples either analytically or through Monte Carlo experiments.

Our Monte Carlo findings on the superiority of ML estimates over FGLS have important implications for applied production studies. Factors influencing crop yield variability have received considerable attention in recent years (see, for example, Hazell, 1984; Singh and Byerlee, 1990). Stochastic production function estimation to determine the risk or yield-variability effects of inputs has become increasingly important. In Just–Pope production functions, the risk effects of inputs depend on the signs and magnitudes of the α parameters.[1] Our Monte Carlo results suggest that biased and inefficient FGLS estimates of α may seriously understate the risk effects of inputs. Thus the choice of estimation method may explain the observed insignificance of inputs' risk effects in several applied studies. For example, Wan and Anderson (1985) reporting FGLS estimation results of a Just–Pope production function, conclude: 'most of the estimates determining the marginal risk effects lack statistical significance. This suggests that the

[1] It is evident from (1) that the variance of output, $V(y)$, is $g(\mathbf{z},\alpha)^2$; consider the typical case where $g(\mathbf{z},\alpha) = G(\mathbf{z}^{\mathrm{T}}\alpha)$, with $G'(\cdot) > 0$; here the effect of the kth input on output variance is given by $\partial V(y)/\partial z_k = G'(\cdot)\alpha_k$, which has the same sign as α_k. Therefore, input z_k is risk-reducing if $\alpha_k < 0$, and risk-increasing if $\alpha_k > 0$.

measured controllable factors do not contribute very significantly to … production variability' (p. 85). Our analysis indicates that the insignificance of risk parameter estimates in the cited study may have resulted from inefficiency of FGLS and not from an absence of risk effects of inputs.

III. THE TWO ESTIMATION PROCEDURES

The three step FGLS estimation procedure is as follows. First, (2) is estimated through non-linear least squares. Second, the squared residuals from this estimation, $\hat{u}_i^2$, are used in the next stage:

$$\ln(\hat{u}_i^2) = \ln(g(z_i,\alpha)^2) + v_i \tag{4}$$

where $v_i = \ln \varepsilon_i^2$. Without loss of generality, the estimation equation in (4) can be rewritten as:

$$\ln(\hat{u}_i^2) = \alpha_0^* + \ln(g(z_i,\alpha)^2) + v_i^* \tag{4'}$$

where $v_i^* = v_i - E[v_i]$, and $\alpha_0^* = E[v_i] = E[\ln(\varepsilon_i^2)]$; consequently, $E[v_i^*] = 0$. Estimation of (4′) by least squares provides consistent estimates of α, denoted by $\hat{\alpha}$. The third step constitutes a weighted on-linear least squares regression of (2):

$$y_i^* = f^*(x_i,\beta) + u_i^* \tag{5}$$

where $y_i^* = y_i \cdot g(z_i,\hat{\alpha})^{-1}$, $f^*(x_i,\beta) = f(x_i,\beta)\cdot g(z_i,\hat{\alpha})^{-1}$ and $u_i^* = u_i \cdot g(z_i,\hat{\alpha})^{-1}$. Let the parameter estimates from the last stage be denoted by $\hat{\beta}$.

Amemiya (1985), and Jobson and Fuller (1980) have demonstrated that the second stage estimate $\hat{\alpha}$, though inefficient, is consistent. This is all that is required to ensure that the FGLS estimate $\hat{\beta}$ is consistent, distributed asymptotically normal and, in the presence of normally distributed errors, is asymptotically fully efficient, that is, its dispersion matrix, $V(\hat{\beta})$, attains the Cramér–Rao lower bound.

The situation is more complex in small samples. The v_is are serially correlated and heteroscedastic because each v_i is constructed from the same estimate of β in the first stage (see Amemiya, 1985 pp. 200–207). Amemiya demonstrates that serial correlation is absent in large samples. However, the problem of heteroscedasticity of v_i persists yielding inefficient estimates of the risk-effect parameters, α.

Within the framework of the FGLS procedure, modifications can be made to address the effects of heteroskedastic v_i on test of hypotheses (see Fomby *et al*., 1988 for a detailed discussion on this issue). White's proposed consistent estimator of the covariance matrix when the form of heteroscedasticity is unknown is suitable. Following White, the estimated dispersion matrix of α can be written as:

$$V(\hat{\alpha}) = (Z^{T}Z)^{-1} Z^{T}DZ(Z^{T}Z)^{-1} \tag{6}$$

where D is a diagonal matrix with ith element being $(\hat{v}_i^*)^2$, the squared residual from the estimation of (4′), and $Z = \partial \ln(g(z,\alpha)^2)/\partial\alpha$. This correction provides consistent standard errors for $\hat{\alpha}$, but the α estimates, and, therefore, the third stage estimation weights, $g(z_i,\hat{\alpha})^{-1}$, remain unchanged. Consequently, White's (1980) correction in (6) leaves the FGLS $\hat{\beta}$ estimates and their standard errors unaffected.

Maximum likelihood

In contrast to the three-step FGLS procedure outlined above, the method of maximum likelihood provides consistent and asymptotically fully efficient estimates of both sets of parameters, β and α, in a single stage. Since $y_i \sim N(f(x_i,\beta), g(z_i,\alpha)^2)$ under the assumption $\varepsilon_i \sim N(0,1)$, the log likelihood function is given by:

$$\ln L = -\frac{1}{2}\left[n\cdot\ln(2\pi) + \sum_{i=1}^{n}\ln(g(z_i,\alpha)^2) + \sum_{i=1}^{n}\frac{(y_i - f(x_i,\beta))^2}{g(z_i,\alpha)^2}\right] \tag{7}$$

Maximization of $\ln L$ with respect to α and β provides ML estimates. Most econometrics software packages will optimize (7) through an iterative method. The FGLS estimates of α and β can be used as starting values for the iterations.

Since

$$\mathrm{E}\left[\frac{\partial^2 \ln L}{\partial\alpha\partial\beta}\right]_{|\hat{\alpha}_{ML},\hat{\beta}_{ML}} = 0$$

the information matrix for (7) is block diagonal which implies that α and β can be estimated through Oberhofer and Kmenta's (1974) 'zig-zag' method without loss of efficiency. The iterative method essentially involves maximizing (7) with respect to β holding α at a given value and then maximizing with respect α holding β at the estimate from the preceding step. Iteration continues until convergence (see Greene, 1993, p. 367). However, with the advent of modern econometric software and nonlinear optimization techniques, it may be easier in most cases to optimize (7) directly with respect to α and β. The 'zig-zag' method may be necessary only in the most difficult cases where the $f(\cdot)$ and $g(\cdot)$ functions have a large number of parameters or are highly nonlinear in parameters.

Composition of FGLS and ML estimators

To compare the two estimators it is necessary to explicitly set out the criterion functions of the last two stages in FGLS. The second and third stage optimization problems in FGLS are:

$$\underset{\alpha}{\mathrm{Min}}\, S_2 = \sum_{i=1}^{n}\{\ln(\hat{u}_i^2) - \alpha_0^* - \ln(g(z_i,\alpha)^2)\}^2 \tag{8a}$$

$$= \sum_{i=1}^{n}\{\ln(y_i - f(x_i,\hat{\beta}))^2 - \alpha_0^* - \ln(g(z_i,\alpha)^2)\}^2$$

$$\underset{\beta}{\mathrm{Min}}\, S_3 = \sum_{i=1}^{n}\left(\frac{y_i - f(x_i,\beta)}{g(z_i,\hat{\alpha})}\right)^2 \tag{8b}$$

It is obvious from the comparison of (7) and (8b) that, for a given α, the FGLS and ML estimates of β would be identical. That is not true in the case of α. For a given value of $\hat{\beta}$, and attendant $\hat{u}_i$, the ML first order condition (foc) for α is:

$$\frac{\partial \ln L}{\partial \alpha} = \sum_{i=1}^{n} \left\{ \frac{\hat{u}_i^2}{g(z_i,\alpha)^2} - 1 \right\} \frac{g_\alpha(\cdot)}{g(z_i,\alpha)} = 0 \tag{9}$$

where g_α denotes $\partial g(\cdot)/\partial\alpha$. The corresponding FGLS foc, after some rearrangement, is:

$$\frac{\partial S_2}{\partial \alpha} = -4 \sum_{i=1}^{n} \{\ln(\hat{u}_i^2) - \alpha_i^* - \ln(g(z_i,\alpha)^2)\} \frac{g_\alpha(\cdot)}{g(z_i,\alpha)} = 0 \tag{10}$$

Clearly, the focs in (9) and (10) cannot be identical irrespective of the functional form for $g(\cdot)$. Also, iteration between the steps of FGLS will not lead to ML estimates.

The foregoing discussion of criterion functions at various stages of FGLS begs the question whether, like ML, one can estimate α and β in a single stage through FGLS. Given (2) and (3), the single-stage estimation problem is:

$$\underset{\beta,\alpha}{\text{Min}}\, S = \sum_{i}^{n} \varepsilon_i^2 = \sum_{i}^{n} \left(\frac{y_i - f(x_i, \beta)}{g(z_i,\alpha)} \right)^2 \tag{11}$$

Clearly for foc for β is the same as before. If an interior solution exists, the foc for α is:

$$\frac{\partial S}{\partial \alpha} = -\sum_{i}^{n} 2 \left(\frac{(y_i - f(x_i, \beta))^2}{g(z_i,\alpha)^3} \right) g_\alpha(\cdot) = 0 \tag{12}$$

A serious problem with the optimization in (11) is that the criterion function is not well formed with respect to α; that is, an interior solution for $\hat{\alpha}$ does not exist. Suppose, for the moment, α is a scalar and $g_\alpha(\cdot) > 0$ (< 0); then, since $g(z_i,\alpha)$ is in the denominator of the expression in (11), S is minimized as α goes to plus (minus) infinity. That is, α will either explode positively or negatively and the equality in (12) can never be satisfied. Similar comments apply to the elements of α when it is a vector. Thus, FGLS estimation is possible only because of the multi-stage nature of the procedure; direct optimization of a single well-formed GLS criterion is impossible. In contrast, because of the $\ln(g(z_i,\alpha)^2)$ term in addition to $[(y_i - f(x_i, \beta))/g(z_i,\alpha)]^2$ in the log-likelihood equation, the criterion function in ML is well defined, allowing a strictly interior solution for α.

It is $\hat{\alpha}$'s problem of unboundedness that necessitates a multi-stage estimation procedure in FGLS. This is markedly different from the usual nonlinear estimation problems in which a single criterion is well specified and multi-stage estimation is simply a computational procedure to find the optimum. In the FGLS case, the criterion (S in (11)) cannot be maximized directly; the multi-stage procedure becomes an integral part of the objective function by specifying two different criteria, S_3 of (8b) in combination with S_2 of (8a) (for which $\hat{\alpha}$ is defined).

The discussion of the two competing estimation methods can be summarized as follows: (a) both FGLS and ML provide consistent and asymptotically efficient estimates of β, but the FGLS estimate of α is asymptotically inefficient; (b) in small samples, the FGLS α estimate is likely to be biased and inefficient, and a biased α causes the estimate of β and its dispersion matrix to be biased as well; (c) the ML estimates of α and β are consistent and asymptotically fully efficient.

In general models of heteroscedasticity the inefficiency of α is of little concern because the FGLS estimate of β is efficient as long as $\hat{\alpha}$ is consistent. But in Just–Pope production function estimation, efficiency of $\hat{\alpha}$ matters since these parameters capture the risk effect of inputs.

IV. MONTE CARLO EXPERIMENTS

The finite sample bias and inefficiency of FGLS *vis a vis* ML estimates cannot be determined *a priori*. For this we turn to the Monte Carlo experiments. These experiments were undertaken using the following Just–Pope production function:

$$y_i = f(x_i, \beta) + u_i \equiv \beta_0 \sum_{k=1}^{K} x_{ik}^{\beta_k} + u_i \tag{13a}$$

$$u_i = g(z_i,\alpha)\varepsilon_i \equiv \left[\exp\left\{ \sum_{j=1}^{J} \alpha_j \cdot z_{ij} \right\} \right]^{1/2} \varepsilon_i \tag{13b}$$

Equation 13b implies that $V(u_i) \equiv \exp\{\sum_{j=1}^{J} \alpha_j \cdot z_{ij}\} \cdot \sigma_\varepsilon^2$, where σ_ε^2 denotes the variance of the underlying error, ε. The exponential form for heteroskedasticity was chosen because it is widely used in the applied literature. Two design matrices were used in the experiments and they are available from the authors. The first design matrix, which we will call design matrix A, had 60 observations on two inputs, x_1 and x_2, plus a constant; the inputs entering the random part of the production function, z_1 and z_2, were identical to x_1 and x_2. The second design matrix, B, had 48 observations on six inputs, of which five, $\{x_1, \ldots, x_5\}$, plus the constant were included in the **x**-matrix, while the **z**-matrix had three inputs: $\{x_4, x_5, x_6\}$, the inputs x_4 and x_5 being common to matrices **x** and **z**. In both matrices the x_i vector included an intercept but z_i did not. Thus, in the second stage of FGLS an intercept coefficient was allowed – this corresponding to α_0^* in (4′) – but it was not used in forming the weights, $\exp\{\sum_{j=1}^{J} \hat{\alpha}_j \cdot z_{ij}\}^{1/2}$, for the third stage estimation. Also, in the second stage, $V(\alpha)$ was estimated using White's (1980) method given in (6); thus, in all cases, the reported FGLS standard errors for α were computed using White's procedure.

In the maximum likelihood (ML) procedure, the parameter vectors, α and β, and σ_ε^2 were jointly estimated by maximizing the log-likelihood function set out in (7) with appropriate modification to include σ_ε^2. ML estimation was undertaken using the 'zig-zag' method discussed earlier, that is, holding the elements of α and σ_ε^2 at arbitrary values, β was estimated, and then α and σ_ε^2 were jointly estimated holding β at its estimated value; iteration continued until convergence. This procedure is appropriate because the information-matrix is block diagonal, in $\{\alpha, \sigma_\varepsilon^2\}$ and β. In the interest of brevity, we report only α and β estimates in most cases; the σ_ε^2 estimates are provided only in results of experiments where different values of σ_ε^2 were used in generating the data. The Monte Carlo experiments undertaken in this study can be categorized into four groups. The details on each follow.

Group One: experiments using design matrix A

In all experiments in this group, 1000 sets of 60 observations on y were generated using design matrix A and the form in (13a, b), with the underlying error, ε_i, being a random draw from a standard normal distribution. The different sets of parameter values used in generating the data are reported in the column labelled 'actual values' in Table 1. It is clear from (13a, b) that larger values for the $g(\mathbf{z}_i, \alpha)$ part relative to the $f(\mathbf{x}_i, \beta)$ part of the production function increases the coefficient of variation (standard deviation/mean) of output, y. The sets of α and β parameter values were chosen to generate observations on y that differed considerably in coefficient of variation (CV). For each set of parameter values, the CV for each of the 1000 sets of 60 observations on y were computed; the figures in the Table are the average CVs.

Let $\hat{\theta}_k$ denote any element of the set $\{\hat{\alpha}_{1_k}, \hat{\alpha}_{2_k}, \hat{\beta}_{0_k}, \hat{\beta}_{1_k}, \hat{\beta}_{2_k}\}$, where the subscript k denotes the kth replication. The mean parameter estimate from 1000 replications, that is

$$\bar{\bar{\theta}} = \frac{1}{1000} \sum_{k=1}^{1000} \hat{\theta}_k$$

is reported as the coefficient estimate in Table 1. The t-ratios are also averages i.e.

$$\frac{1}{1000} \sum_{k=1}^{1000} \left(\frac{\hat{\theta}_k}{\text{S.E. } (\hat{\theta}_k)} \right)$$

The power of a test, defined as the probability of rejecting a false null hypothesis, in this case $H_0: \theta = 0$, was computed for the t-test as the proportion of t-ratios that exceeded the critical level at the 5% level of significance, that is: [number of $t(\hat{\theta}) > t^*$]/1000, where t^* is the critical table value of the t-statistics.

The mean-squared error is a measure of overall performance of an estimator and it reflects possible trade-off between unbiasedness and efficiency. Since true parameter values are never known in data analysis, the MSE can be meaningfully used only in Monte Carlo studies. The reported average MSE is:

$$\text{MSE}(\theta) = \frac{1}{1000} \sum_{k=1}^{1000} (\hat{\theta}_k - \bar{\bar{\theta}})^2 + (\bar{\bar{\theta}} - \theta)^2$$

where the first term in the expression reflects the variance of the estimated parameter across repetitions and the second term captures the degree of bias, θ being the actual parameter value. The last column in Table 1 presents the percentage difference in MSE between FGLS and ML estimators for each parameter.

The results in Table 1 show that for all parameters in the four sets the ML procedure yields a substantially lower MSE and a higher power of test than FGLS. Output CV, which reflects the 'noise to signal ratio' in the estimation equation, increases from 0.361 in set I to 3.341 in set IV. Not unexpectedly, with increasing 'noise', the MSE in both procedures, FGLS and MLE, increases. However, the difference in relative performance of the two procedures becomes markedly more pronounced at higher levels of CV. Also note, when CV is low (sets I and II), the difference between FGLS and ML MSE-s is relatively higher for α_1 and α_2 than for β_0, β_1 and β_2. But at higher levels of CV (sets II and III), this distinction vanishes – the percentage differences between FGLS and ML MSE-s are approximately the same for all parameters. The foregoing comments on the performance differences between the two procedures also apply in the case of power of the t-test, i.e. the ML power is generally greater than the FGLS power. Thus the probability of a Type II error (not rejecting a false null hypothesis) is higher under FGLS than ML. To investigate Type I errors (rejecting a true null hypothesis), we have calculated the average t-ratios: $t_\theta = (\hat{\theta} - \theta)/\text{SE}(\hat{\theta})$, testing the parameter estimates against the actual values. The ML values are all smaller than the FGLS values, indicating superior Type I error performance for ML, with the performance differential increasing as the coefficient of variation of y increases.

It is also evident from Table 1 (the column labelled t_0) that, while almost all ML estimates of non-zero α parameters are significant at the 1% level, several FGLS estimates of these parameter are insignificant even at the 5% level. Thus, if inferences were drawn from FLGS estimates, one would erroneously conclude that the two inputs do not have significant risk effects.

Group Two: the effect of different levels of σ_ε

The results of the second set of experiments are reported in Table 2. Unlike the experiments in the preceding group where the parameters were varied to change output CV, here parameters were held at a fixed value and different values of σ_ε were chosen in generating data on y using design matrix A. The actual σ_ε and the attendant output CV are reported in the first two columns of Table 2. In the interest of brevity, only the mean coefficient estimates and the MSE for various parameters, including σ_ε, are presented. Detailed results are available from the authors.

The range of output CV is smaller here than in the preceding set of experiments; as a consequence, the performance difference between FGLS and ML is less pronounced,

Table 1. *Monte Carlo results: estimation under various parameter values*

			Coefficient estimate		Mean squared error			Power of test		t_θ^* $H_0: \theta =$ actual value		t_0^{**} $H_0: \theta = 0$	
Set	Parameter	Actual value	FGLS	ML	FGLS	ML	Diff. (%)	FGLS	ML	FGLS	ML	FGLS	ML
I	α_1	− 0.100	− 0.0952	− 0.1010	0.0052	0.0025	51.92	0.299	0.570	0.92	0.86	1.55	2.23
	α_2	0.100	0.0952	0.0941	0.0038	0.0017	55.26	0.421	0.840	0.87	0.83	1.84	3.12
	β_0	1.200	1.2058	1.2072	0.0162	0.0158	6.79	1.000	0.984	0.92	0.87	10.82	10.74
	β_1	0.300	0.2990	0.2986	0.0016	0.0015	7.64	1.000	0.984	0.92	0.86	8.82	8.60
	β_2	0.600	0.6008	0.6006	0.0026	0.0025	6.49	1.000	0.984	0.92	0.86	13.64	13.21
	CV of y	0.361											
	Average				**0.0059**	**0.0048**	**25.62**	**0.744**	**0.872**				
II	α_1	0.000	− 0.0044	− 0.0084	0.0055	0.0028	48.56	—***	—***	0.92	0.85	0.92	0.85
	α_2	0.100	0.0947	0.0996	0.0037	0.0012	65.61	0.420	0.842	0.88	0.84	1.84	3.11
	β_0	1.500	1.5106	1.5094	0.0288	0.0266	7.64	1.000	0.999	0.91	0.87	10.09	10.07
	β_1	0.500	0.4984	0.4987	0.0023	0.0021	8.89	1.000	0.999	0.93	0.87	12.51	12.10
	β_2	0.500	0.5007	0.5006	0.0027	0.0025	7.38	1.000	0.999	0.92	0.86	11.18	10.95
	CV of y	0.430											
	Average				**0.0086**	**0.0070**	**27.62**	**0.855**	**0.960**				
III	α_1	0.250	0.1879	0.2437	0.0112	0.0027	75.83	0.739	0.993	1.27	0.86	3.03	5.18
	α_2	0.150	0.0726	0.1486	0.0138	0.0011	92.03	0.432	0.991	1.96	0.83	1.87	4.81
	β_0	2.000	1.7576	2.1391	1.5860	0.7333	53.76	0.661	0.993	2.01	0.87	2.11	2.71
	β_1	0.750	1.1377	0.7412	2.4002	0.0391	98.37	0.923	0.969	1.61	0.87	4.40	4.08
	β_2	0.200	0.2659	0.2066	0.8985	0.0205	97.71	0.280	0.300	1.67	0.82	1.67	1.53
	CV of y	1.679											
	Average				**0.9819**	**0.1593**	**83.54**	**0.607**	**0.849**				
IV	α_1	0.250	0.1951	0.2457	0.0108	0.0029	73.15	0.728	0.993	1.16	0.88	3.05	5.21
	α_2	0.150	0.0820	0.1473	0.0107	0.0012	88.79	0.402	0.996	1.65	0.87	1.80	4.76
	β_0	2.000	2.0459	2.4413	4.6846	3.9550	15.57	0.074	0.101	1.97	1.07	1.38	1.70
	β_1	0.750	1.2486	0.7605	3.3766	0.1001	97.04	0.670	0.713	1.42	0.89	2.93	2.67
	β_2	0.001	0.0250	− 0.0038	1.0822	0.0584	94.60	0.092	0.049	1.12	0.84	1.12	0.84
	CV of y	3.341											
	Average				**1.833**	**0.8235**	**73.83**	**0.393**	**0.570**				

* $t_\theta = (\hat{\theta} - \theta)/\mathrm{SE}(\hat{\theta})$, where θ denotes any element of the vector $\{\alpha_1, \alpha_2, \beta_0, \beta_1, \beta_2\}$. Its estimated value is $\hat{\theta}$ and SE denotes standard error.
** $t_0 = \hat{\theta}/\mathrm{SE}(\hat{\theta})$
*** The power of test is not relevant here since the null $H_0: \theta = 0$ is true.

Table 2. *Monte Carlo results: estimation under various levels of* σ_ε

			Coefficient estimates		Mean squared error		
Actual σ_ε	CV of y	Parameter*	FGLS	ML	FGLS	ML	Diff. (%)
		α_1	− 0.0953	− 0.0924	0.0052	0.0037	28.85
		α_2	0.0954	0.0929	0.0037	0.0016	56.76
0.5	0.3606	β_0	1.2016	1.2021	0.0040	0.0038	5.00
		β_1	0.2995	0.2995	0.0004	0.0004	2.56
		β_2	0.6003	0.6001	0.0007	0.0006	6.15
		Average**					19.86
		σ_ε	0.3607	0.5130	0.3075	0.0204	93.37
		α_1	− 0.0952	− 0.1010	0.0052	0.0025	51.92
		α_2	0.0952	0.0941	0.0038	0.0017	55.26
1.0	0.3772	β_0	1.2058	1.2072	0.0162	0.0151	6.79
		β_1	0.2990	0.2986	0.0016	0.0015	7.64
		β_2	0.6008	0.6006	0.0026	0.0025	6.49
		Average**					25.62
		σ_ε	1.4577	1.0389	4.9301	0.0788	98.40
		α_1	− 0.0951	− 0.1049	0.0052	0.0026	50.01
		α_2	0.0950	0.0965	0.0037	0.0017	53.91
1.5	0.4036	β_0	1.2122	1.2040	0.0370	0.0308	16.75
		β_1	0.2984	0.2987	0.0036	0.0031	12.29
		β_2	0.6016	0.6034	0.0060	0.0047	22.04
		Average**					31.00
		σ_ε	3.2651	1.5538	25.291	0.1787	99.29

* The actual parameters were: $\alpha_1 = -0.1$; $\alpha_2 = 0.1$; $\beta_0 = 1.2$; $\beta_1 = 0.3$; $\beta_2 = 0.6$.
** The average is for all parameters except σ_ε.

although evident in all cases. But the difference is substantial in the case of σ_ε and the α parameters. In all three cases, the α_1 and α_2 estimates under FGLS are statistically insignificant; in contrast, the ML estimates of these parameters are significant at the 1% level (these results are not reported in Table 2).

The estimates of σ_ε in Table 2 show that FGLS MSE-s are between 15 to 141 times higher than ML MSEs for this parameter. The main reason for the poor performance of FGLS in estimating σ_ε lies in the high degree of bias in estimating $g(z_i,\alpha)$ in the second-stage of the three-stage FGLS estimation. Under FGLS

$$(\hat{\sigma}^2_\varepsilon)_{\text{FGLS}} = \sum_{i=1}^{60} \frac{(\hat{u}^*_i)^2}{57}$$

where $\hat{u}^*_i$ denotes the residuals from the third stage estimation, and $u^*_i = u_i \cdot g(z_i,\hat{\alpha})^{-1}$ (see Equation 5). The bias in the α estimates is translated to the estimate of $g(z_i,\hat{\alpha})^{-1}$, which is the weight in the third stage estimation. As a result the third stage regression did not fit well, yielding biased estimates for σ_ε. This problem does not arise in the ML procedure because α, β, and σ_ε are jointly estimated in a single stage by maximizing a well-formed criterion function.

Thus, the results from Group Two experiments suggest that, the FGLS procedure not only yields inefficient estimates of α and β, its estimate of σ^2_ε also suffers from an extremely high degree of bias. Recall, $g(z_i,\alpha)^2 \cdot \sigma^2_\varepsilon$ is the output variance in Just–Pope production function. Thus, a biased σ^2_ε estimate may lead to severely erroneous inferences if FGLS procedure is used, for example, in examining the variance effects of inputs or in analyzing the change in output variability over time or across regions.

Group Three: experiments using design matrix B

The Monte Carlo experiments in this group use design matrix B. Ten parameters were estimated, three αs, six βs, and σ^2_ε. The actual values of the parameters used in generating 1000 samples of 48 observations on y are given in Table 3. The underlying error in data generation was a standard normal variate. The method of computation of MSE, t-ratios, etc. are the same as before.

Table 3 results show that the average MSE difference between the two procedures is around 43%. The power of the t-test is considerably higher for the tests associated with α_2, α_3, β_2, and β_3, and essentially the same for the remaining parameters. Thus, ML is less likely than FGLS to make a Type II error. Both estimators make Type I errors on the

Table 3. *Monte Carlo results: estimation using design matrix B*

Parameter	Actual value	Coefficient estimate		Mean squared error			Power of test		t_θ^* $H_0: \theta =$ actual value		t_0^{**} $H_0: \theta = 0$	
		FGLS	ML	FGLS	ML	Diff. (%)	FGLS	ML	FGLS	ML	FGLS	ML
α_1	2.000	1.5276	2.2852	0.349	0.141	59.60	0.961	0.999	1.40	1.54	4.43	11.52
α_2	− 1.000	− 0.4374	− 1.4051	0.371	0.269	27.76	0.503	0.998	2.88	2.07	2.13	7.17
α_3	− 0.300	− 0.3043	− 0.4008	0.062	0.101	− 38.61	0.272	0.507	0.86	1.22	1.48	2.10
β_0	7.000	6.9799	7.0060	0.256	0.143	44.14	1.000	1.000	0.64	1.16	11.59	27.02
β_1	0.500	0.4981	0.4977	0.005	0.004	32.70	0.998	1.000	0.72	1.53	6.26	15.76
β_2	0.050	0.0551	0.0527	0.010	0.007	30.00	0.110	0.346	0.87	1.46	0.98	1.79
β_3	0.200	0.2030	0.2001	0.026	0.005	80.77	0.380	0.941	0.94	1.93	1.72	6.77
β_4	0.500	0.4970	0.5000	0.004	0.001	66.67	1.000	0.999	0.90	1.27	10.01	23.74
β_5	0.300	0.3012	0.3002	0.003	0.001	80.64	0.999	1.000	0.84	1.50	5.75	23.56
	Average			**0.121**	**0.075**	**42.96**	**0.691**	**0.866**				
Aver. sq. bias:												
for α_1–α_3		0.749×10^{-4}	0.082×10^{-4}									
for β_0–β_5		0.1800	0.0856									

* $t_\theta = (\hat{\theta} - \theta)/\text{SE}(\hat{\theta})$, where θ denotes any element of the vector $\{\alpha_1, \alpha_2, \beta_0, \beta_1, \beta_2\}$. Its estimated value is $\hat{\theta}$ and SE denotes standard error.
** $t_0 = \hat{\theta}/\text{SE}(\hat{\theta})$

α_2 test (the t_θ column in Table 3) with ML at less than the 5% level and FGLS even worse at less than the 1% level. However, the ML t_θs are worse (i.e. larger) for all other parameters. Thus, while the 5% decision criterion keeps ML from making any more false rejections of the true null than FGLS in our example, the increased power of the ML tests does come with the possibility of higher Type I error.

In consonance with the findings in the last two groups of experiments, Table 3 results show that, while two out of three FGLS estimates of αs are insignificant at the 1% level, all three ML estimates are highly significant. Thus, a common finding in all three groups of experiments is that FGLS substantially underestimates the significance of the risk effects of inputs in a Just–Pope production function.

Table 3 also shows that, in contrast to the findings in the preceding two groups of experiments, the MSE difference between FGLS and ML is more pronounced in the case of β than the α parameters. While the average squared bias difference between FGLS and ML for β parameters is about 0.0944, the corresponding figure for the three α parameters is only 0.667×10^{-4}. To understand the inferential consequences of biased β estimates, we computed the marginal product (MP) of inputs using FGLS and ML parameter estimates. Since the mean of the stochastic part of the Just–Pope production function, $g(z,\alpha)\varepsilon$, is zero, the expressions for the MPs, evaluated at input means, involve only $\hat{\beta}$. The MPs were computed using each of the 1000 sets of $\hat{\beta}$ values, with the inputs held at the sample means. The average of these 1000 MP estimates for each input under both procedures are reported in Table 4 in addition to the true MP values. The squared bias and the MSE were also calculated for each of the 1000 repetitions and then averaged. Table 4 reveals that the average squared bias under FGLS is almost 32% higher and average MSE 58% larger than under ML. Note, the bias in MP estimates is not caused by α estimates, it stems solely from the bias in estimating β. This result, in conjunction with the findings discussed earlier, suggests that not only is FGLS likely to underestimate the risk effects of inputs but it may also provide considerably biased estimates of the mean effects.

Group Four: *experiments with non-normal error distribution*

It may be argued that the foregoing experiments are biased in favour of ML because the errors in the estimation equations are normally distributed. To explore the relative performance of the two procedures under non-normality we undertook additional experiments in which errors distributed as a χ^2 variate with various degrees of freedom were used in generating the observations on y. In the interest of brevity we used only the five-parameter model in these experiments. The data-generating parameters values were the same as in Group Two experiments.

The χ^2 pdfs skewness, and the attendant departure from normality, is inversely proportional to its degree of freedom,

Table 4. *Marginal product estimates under FGLS and ML using matrix B*

		Estimated value		Mean squared error		
	Actual value	FGLS	ML	FGLS	ML	Diff. (%)
MP of x_1	4.997	5.031	5.033	0.478	0.325	32.01
MP of x_2	0.636	0.716	0.686	1.742	1.118	35.82
MP of x_3	2.315	2.427	2.355	3.784	0.732	80.65
MP of x_4	7.056	7.082	7.141	0.446	0.141	68.39
MP of x_5	4.537	4.604	4.600	0.662	0.178	73.11
Average				**1.422**	**0.499**	**58.00**
Av. squared bias		0.00502	0.00342			

denoted by v (see Fig. 1). The Monte Carlo experiments were repeated by generating χ^2-errors, with v set at 5, 10, 15, and 20. In each case we computed 1000 sets of parameter estimates. Detailed experiment results and individual parameter estimates are available from the authors. In Table 5 we report only the average MSE and average power of the t-test for the five parameters.

The second column in Table 5 provides test results for departure from normality (see Kiefer and Salmon, 1983 for a description of the test). The null hypothesis is that the distribution is normal and the test statistic and its P-value are reported. As is also evident from Fig. 1, the skewness and kurtosis of the χ^2- distribution are quite different from those of a normal distribution for all the values of v considered.

Table 5 results show that at all levels of v, the ML outperforms FGLS in terms of MSE and the power of the t-test. However, the performance difference between the two procedures decreases as the χ^2 distribution's departure from normality becomes pronounced, that is, as the parameter v decreases. Thus, in empirical applications ML is likely to outperform FGLS in terms of unbiasedness and efficiency even when the error distribution departure from normality is markedly pronounced, a somewhat unexpected result.

Before concluding this section, it may be worth pointing out that we repeated all experiments in the four groups using several other parameter values and functional forms for the $f(\cdot)$ and $g(\cdot)$ functions (in (13)). The performance difference between FGLS and ML prevailed in all these experiments. We have not presented these results in the interest of brevity, but they are available on request. Based on these and the reported results, we feel confident to assert that this study's principal conclusions regarding the small sample superiority of ML over FGLS are robust to functional form and parameter choice.

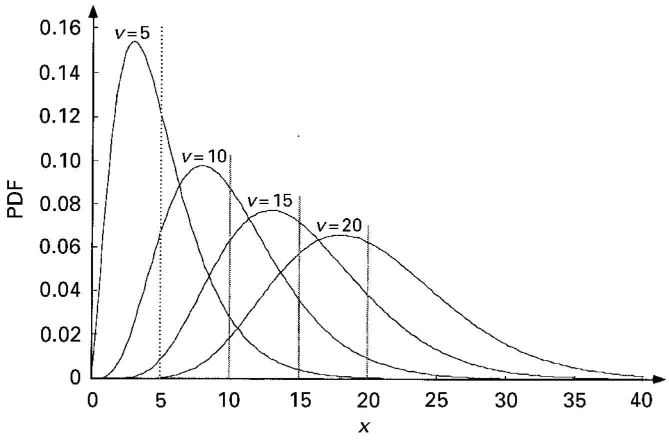

Fig. 1. *Chi-squared probability density functions with four different degrees of freedom, v. The vertical lines locate the respective means*

V. CONCLUDING COMMENTS

In the applied production literature, stochastic production functions have been traditionally estimated using a multi-stage feasible generalized least squares procedure. The objective of this study was to investigate whether, in small samples, greater efficiency and lower bias of the estimators can be attained by the alternative of a single-stage maximum likelihood method. We explored the issue through Monte Carlo experiments. Our findings can be summarized as follows: (i) FGLS substantially underestimates the significance of the parameters in the random component of the production function. This result holds despite White's correction in estimating the variance-covariance matrix of these parameters. (ii) Relative to FGLS, ML yields considerable improvement in the probability of rejecting a false null hypothesis (i.e. power of a test) in all cases. (iii) ML estimates have considerably smaller MSE relative to FGLS for parameters that enter the random as well as the nonrandom part of the estimation equation. Bias is also markedly pronounced in the FGLS estimates of the variance of the underlying error. (iv) The estimated marginal product of inputs were found to be substantially biased when they were computed using the FGLS parameter estimates. (v) The MSE difference in the parameter estimates under the two procedures holds even when the error distribution departs from normality. (v) Although the performance difference between the two procedures diminishes as the underlying error distribution's departure from normality increases, the

Table 5. *Monte Carlo results: estimation under chi-squared error distribution*

χ^2 skewness parameter** v	Normality test*** statistics (P-value)	Average mean squared error*			Average power of test*		
		FGLS	ML	Diff. (%)	FGLS	ML	Diff. (%)
5	30.40 (0.000)	0.0404	0.0377	6.68	0.7868	0.8374	6.04
10	11.60 (0.003)	0.1169	0.0753	35.58	0.6976	0.7820	10.79
15	6.93 (0.031)	0.2313	0.1173	49.29	0.6302	0.7104	11.29
20	4.90 (0.086)	0.5213	0.1746	66.51	0.5220	0.6182	15.56

* The MSE and power of test figures are averages for all five parameters.
The data on y were generated using: $\alpha_1 = -0.1$; $\alpha_2 = 0.1$; $\beta_0 = 1.2$; $\beta_1 = 0.3$; $\beta_2 = 0.6$.
** The χ^2 distribution is more skewed, and its departure from normality more pronounced, at lower values of v.
*** The test for normality is based on the third and fourth moments of the error distribution. The null hypothesis is that the distribution is normal.

ML estimates continue to outperform the FGLS estimates in both power and MSE.

The observed inefficiency of FGLS relative to ML is not in disagreement with Just and Pope's analysis. In their 1978 article, they show that the FGLS estimate of α is asymptotically inefficient and they argue cogently in favour of using ML. In the context of the loglinear functional form considered in their study, they conclude: 'maximum likelihood estimation of [α] is more than twice as efficient as step estimation asymptotically' (p. 80). Yet, instead of using ML, most applied production studies have opted for the three-stage FGLS estimation technique proposed in Just and Pope's (1979) paper.

The implications of the paper's findings are not limited to the literature on Just–Pope production function but extend to estimation models with heteroscedastic errors. In the general econometrics literature it has been demonstrated that a consistent estimate of the variance equation parameters (i.e. α) is sufficient to ensure that the FGLS estimate of β is consistent and asymptotically fully efficient. Consequently, FGLS's inefficiency in estimating α is of little concern in securing the desired asymptotic properties of $\hat{\beta}$. But in small samples the analytical results are not definitive and several theorists have argued that one may use either FGLS or ML. Our Monte Carlo results show that one should not be ambivalent in the choice of estimation procedures: FGLS yields not only inefficient estimates of α but, more importantly, its performance is considerably inferior to ML in estimating β.

REFERENCES

Amemiya, T. (1985) *Advanced Econometrics* (Harvard University Press, Cambridge).

Buccola, S. T. and McCarl, B. A. (1986) Small-sample evaluation of mean-variance production function estimators, *American Journal of Agricultural Economics*, **68**, 732–8.

Fomby, T. B., Hill, R. C. and Johnson, S. R. (1988) *Advanced Econometric Methods* (Springer-Verlag).

Goldfield, S. M. and Quandt, R. E. (1972) *Nonlinear Methods in Econometrics* (Amsterdam, North Holland).

Greene, W. H. (1993) *Econometric Analysis*, 2nd edn (Macmillan).

Griffiths, W. E. and Anderson, J. R. (1982) Using time-series and cross-section data to estimate a production function with positive and negative marginal risks, *Journal of the American Statistical Association*, **77**, 529–36.

Hallam, A., Hassan, R. M. and D'Silva, B. (1989) Measuring stochastic technology for the multi-product film: the irrigated farms of Sudan, *Canadian Journal of Agricultural Economics*, **37**, 495–512.

Harvey, A. (1976) Estimating regression models with multiplicative heteroskedasticity, *Econometrica*, **44**, 461–5.

Hassan, R. M. and Hallam, A. (1990) Stochastic technology in a programming framework: a generalized mean-variance farm model, *Journal of Agricultural Economics*, **41**, 196–206.

Hazell, P. (1984) Sources of increased instability in Indian and U.S. cereal production, *American Journal of Agricultural Economics*, **66**, 302–11.

Hildreth, C. and Houck, J. P. (1968) Some estimators for a linear model with random coefficients, *Journal of the American Statistical Association*, **63**, 584–95.

Jobson, J. D. and Fuller, W. A. (1980) Least squares estimation when the covariance matrix and parameter vector are functionally related, *Journal of the American Statistical Association*, **75**, 176–81.

Judge, G. G., Griffiths, W. E., Hill, R. C., Lütkepohl, H. and Lee, T. (1985) *The theory and Practice of Econometrics*, 2nd edn (John Wiley and Sons).

Just, R. E. and Pope, R. D. (1978) Stochastic specification of production functions and economic implications, *Journal of Econometrics*, **7**, 67–86.

Just, R. E. and Pope, R. D. (1979) Production function estimation and related risk considerations, *American Journal of Agricultural Economics*, **61**, 276–84.

Kiefer, N. and Salmon, M. (1983) Testing normality in econometric models, *Economics Letters*, **11** 123–8.

Kumbhakar, Sabul, C. (1993) Production risk, technical efficiency, and panel data, *Economics Letters*, **41**, 11–16.

Love, A. and Buccola, S. T. (1991) Joint risk preference-technology estimation with a primal system, *American Journal of Agricultural Economics*, **73**, 765–74.

Magnus, J. R. (1978) Maximum likelihood estimation of the GLS model with unknown parameters in the disturbance covariance matrix, *Journal of Econometrics*, **7**, 281–312.

McCarl, B. A. and Rettig, R. B. (1983) Influence of hatchery smolt releases on adult salmon production and its variability, *Canadian Journal of Fisheries and Aquatic Sciences*, **40**, 1880–6.

Oberhofer, W. and Kmenta, J. (1974) A general procedure for obtaining maximum likelihood estimates in generalized regression models, *Econometrica*, **42**, 579–90.

Singh, A and Byerlee, D. (1990) Relative variability in wheat yields across countries and over time, *Journal of Agricultural Economics*, **41**, 21–33.

Taylor, W. E. (1977) Small sample properties of a class of two stage Aitken estimators, *Econometrica*, **45**, 497–508.

Wan, G. H. and Anderson, J. R. (1985) Estimating risk effects in Chinese foodgrain production, *Journal of Agricultural Economics*, **41**, 85–93.

White, H. (1980) A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity, *Econometrica*, **48**, 817–38.