Cobb-Douglas production functions estimated by least-squares methods have been widely applied in agriculture. Estimates based on cross-sectional samples of farms, as most studies have been, almost typically result in some elasticities for land and labor which are negative. These negative coefficients, which are meaningless, can arise for several reasons. One is the prevalence of multicollinearity in random or unstratified farm samples where farmers with large labor inputs also are those with larger inputs of land and various forms of capital. (A farm sample stratified by the various input categories and magnitudes much in the manner of a factorial, central composite or other experimental designs to estimate the physical production surface, would not encounter this problem to the extent of the typical farm survey.) Another reason is reporting or measuring bias. For example, regardless of whether he actually does so, a farmer generally will report that he works twelve months per year; he is unlikely to concede that he is idle part of the time. Finally, other inadequacies of data also pose the potential of negative labor elasticities in Cobb-Douglas functions and are hard to explain under existing knowledge. These negative coefficients confuse the analysis and leave the researcher more or less empty-handed. Among others obtaining these results, Sahota simply considered his results absurd and Srivastava and Nagadevara could not attempt to explain theirs. The production function analysis conducted by Agrawal and Foreman, and Suryanarayana also resulted in negative elasticities for labor and capital services. Heady and Dillon and, recently, Doll have proposed reasons for the prevalence of negative elasticities. We attempt to resolve the problem of negative elasticities with a Bayesian model incorporating relevant prior restrictions.

Least-Squares Model

The model to be estimated is based on a production function from Srivastava and Nagadevara. The data pertain to the production of Desi wheat in Ferozpur district, Punjab (India) for the year 1968–69. The original sample included 150 farmers selected on a stratified random basis. From the 150 farmers, a cross-section sample of 90 farmers producing the Desi variety was selected. The production function estimated is of the form,

$$y_i = A x_{it}^{\beta_1} x_{it}^{\beta_2} x_{it}^{\beta_3} e_{it},$$

where $y_i$ = output of Desi wheat (rupees) of the $i$th farm, $x_{it} = $ area of the $i$th farm under wheat production (hectares), $x_{it} = $ human labor (adult man days) used by the $i$th farm, $x_{it} = $ value of seeds and fertilizers (owned and purchased) used by the $i$th farm (rupees), and $e_{it}$ are random disturbances.

The variables used here are slightly different from those of Srivastava and Nagadevara. They included irrigation charges in $x_3$ whereas we do not. For this reason, our estimates will differ from theirs when least-squares estimation is applied (table 1).

Elasticity of land, given by the estimate of $\beta_1$, is negative possibly due to the existence of multicollinearity. Existence of multicollinearity can be detected by examining the correlation matrix of $x_1, x_2$, and $x_3$ and its inverse which are given in equations (2) and (3), respectively:

$$\begin{pmatrix}
1.000 & 0.814 & 0.808 \\
0.814 & 1.000 & 0.861 \\
0.808 & 0.861 & 1.000
\end{pmatrix},
$$

$$\begin{pmatrix}
3.419 & -1.561 & -1.420 \\
-1.561 & 4.589 & -2.691 \\
-1.420 & -2.691 & 4.466
\end{pmatrix}.$$
Table 1. Least Squares Estimates of the Parameters

<table>
<thead>
<tr>
<th>Method</th>
<th>Log A</th>
<th>( \beta_1 )</th>
<th>( \beta_2 )</th>
<th>( \beta_3 )</th>
<th>( R^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Least squares</td>
<td>-0.246</td>
<td>-0.158</td>
<td>0.946</td>
<td>0.218</td>
<td>0.93</td>
</tr>
<tr>
<td>(0.264)</td>
<td>(-0.814)</td>
<td>(3.118)</td>
<td>(1.213)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*a The t-values are given in parentheses.

For this reason, the estimation method in our model differs from theirs.

In our Bayesian model the elasticities of land, labor, and fertilizer are taken to be nonnegative. Also, from analysis conducted elsewhere and from the existing knowledge in agriculture, we firmly believed that the elasticities do not exceed unity. No extra restriction is imposed on the returns to scale other than the one through the restrictions on the elasticities. Based on the above observations, the prior restrictions on the parameters in equation (1) are

(4) \( 0 \leq \beta_1 \leq 1 \),

(5) \( 0 \leq \beta_2 \leq 1 \),

and

(6) \( 0 \leq \beta_3 \leq 1 \).

The model in equation (1) can be rewritten as

(7) \( z_i = \beta_0 + \beta_1 y_{it} + \beta_2 y_{2i} + \beta_3 y_{3i} + u_i \),

where \( z_i = \log y_{it}, y_{1i} = \log x_{1i}, y_{2i} = \log x_{2i}, y_{3i} = \log x_{3i}, \) and \( \beta_0 = \log A \). The disturbances \( u_i \) are assumed to be independently and normally distributed with zero means and variances \( \sigma^2 \). In vector form, equation (7) can be written as:

(8) \[ \mathbf{z} = \mathbf{\beta}_0 + \mathbf{\beta}_1 \mathbf{y}_1 + \mathbf{\beta}_2 \mathbf{y}_2 + \mathbf{\beta}_3 \mathbf{y}_3 + \mathbf{u} \]

where \( \mathbf{z}, \mathbf{y}_1, \mathbf{y}_2, \) and \( \mathbf{u} \) are column vectors of observations and disturbances. The vector \( \mathbf{\beta} \) is a column vector whose elements are all equal to unity. The likelihood function based on equation (8) is

(9) \[ L = \frac{1}{\sigma^2(2\pi)^{n/2}} \exp \left\{ -\frac{1}{2\sigma^2} \left[ (n - 4)\mathbf{S}^2 + (\mathbf{\beta} - \hat{\mathbf{\beta}})' \mathbf{y}' \mathbf{y} (\mathbf{\beta} - \hat{\mathbf{\beta}}) \right] \right\} \]

where \( \mathbf{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3) \), \( \hat{\mathbf{\beta}} \) is the least-squares estimator of \( \mathbf{\beta} \), and \( \mathbf{y} \) is an \( n \times 4 \) matrix of observations of explanatory variables. The expression,

(10) \[ (n - 4)\mathbf{S}^2 = (\mathbf{z} - \mathbf{\hat{y}} - \mathbf{\hat{y}})' (\mathbf{z} - \mathbf{\hat{y}}), \]

is the usual residual sum of squares for least-squares estimation, and \( n \), which is the number of farms, is equal to 90 in our model. Without loss of generality, let \( (\mathbf{\beta} - \hat{\mathbf{\beta}})' \) and \( \mathbf{y}' \mathbf{y} \) be partitioned as

(11) \[ \mathbf{\delta}_1 = (\beta_1 - \hat{\beta}_1, \beta_2 - \hat{\beta}_2, \beta_3 - \hat{\beta}_3) \] and

\[ \mathbf{\delta}_2 = (\delta_{12}, \delta_{13}, \delta_{23}). \]

and

(12) \[ \mathbf{y}' \mathbf{y} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} \]

With this partitioning, the likelihood function in equation (9) can be written as

(13) \[ L = \frac{1}{\sigma^2(2\pi)^{n/2}} \exp \left\{ -\frac{1}{2\sigma^2} \left[ [(n - 4)\mathbf{S}^2 + \mathbf{\delta}_1' (\mathbf{h}_{11} - h_{12}h_{22}^{-1}h_{21}) \mathbf{\delta}_1 \right. \right. \]

\[ + (\delta_{21} + h_{22}^{-1}h_{21}h_{11} \delta_{12}) h_{22} \left( \delta_{22} + h_{22}^{-1}h_{21} \delta_{12} \right) \] \]

Prior Distribution

In the Bayesian analysis, the following prior distribution is used for the parameters \( \mathbf{\beta}' = (\beta_0, \beta_1, \beta_2, \beta_3) \) and \( \sigma \):

(14) \[ p(\mathbf{\beta}', \sigma) \propto p(\mathbf{\beta})p_2(\sigma), \]

where

(15) \[ p(\mathbf{\beta}) \propto \text{constant} \quad 0 \leq \beta_1 \leq 1, \quad 0 \leq \beta_2 \leq 1, \quad 0 \leq \beta_3 \leq 1, \]

and

(16) \[ p_2(\sigma) \propto \frac{1}{\sigma}. \]

In equations (14), (15), and (16), we follow Jeffreys, Tiao and Zellner, and Zellner. Ignorance about \( \sigma^2 \) is expressed in equation (16) in accordance with Jeffreys's invariance theory (p. 459).

Now using Bayes's theorem to combine the prior distribution in equations (14), (15), and (16) with the likelihood function in equation (13), the joint posterior distribution of the parameters is

(17) \[ p(\mathbf{\beta}_0, \mathbf{\beta}_1, \mathbf{\beta}_2, \mathbf{\beta}_3, \sigma) \propto \frac{1}{\sigma^{n+1}} \exp \left\{ -\frac{1}{2\sigma^2} \left[ [(n - 4)\mathbf{S}^2 + \mathbf{\delta}_1' \mathbf{h}_{11} \mathbf{\delta}_1 \right. \right. \]

\[ + (\delta_{21} + h_{22}^{-1}h_{21} \delta_{12}) h_{22} \left( \delta_{22} + h_{22}^{-1}h_{21} \delta_{12} \right) \] \]

where \( 0 \leq \beta_1 \leq 1, \quad 0 \leq \beta_2 \leq 1, \quad 0 \leq \beta_3 \leq 1. \)

Integrating equation (17) with respect to \( \beta_0, \beta_1, \beta_2, \beta_3, \) and \( \sigma \) successively gives the following marginal posterior distribution of \( \beta_1, \beta_2, \) and \( \beta_3 \):

(18) \[ p(\beta_1, \beta_2, \beta_3) \propto [(n - 4)\mathbf{S}^2 + \mathbf{\delta}_1' \mathbf{h}_{11} \mathbf{\delta}_1 \right. \]

\[ - h_{12} h_{22}^{-1}h_{21} \delta_{12} \] \]

where \( 0 \leq \beta_1 \leq 1, \quad 0 \leq \beta_2 \leq 1, \quad 0 \leq \beta_3 \leq 1. \)

The marginal posterior distribution \( p(\beta_1, \beta_2, \beta_3) \) is multivariate \( \mathbf{\tau} \) but truncated.

Bayesian Estimates of the Parameters

Two possible methods can be suggested to obtain the Bayesian point estimates of the parameters. The first method which maximizes the posterior
distribution will give modes as the estimates. This method resembles maximum likelihood estimation. The second method will give posterior means as the estimates which are optimal when the loss function is a quadratic one.

When the marginal posterior distribution, equation (18), is maximized with respect to $\beta_1$, $\beta_2$, and $\beta_3$ to obtain the modes, essentially the following quadratic programming problem is solved:

\[
\min (\beta_1 - \hat{\beta}_1, \beta_2 - \hat{\beta}_2, \beta_3 - \hat{\beta}_3) (h_{11} - h_{12} h_{22}^{-1} h_{21}) (\beta_1 - \hat{\beta}_1, \beta_2 - \hat{\beta}_2, \beta_3 - \hat{\beta}_3)',
\]
subject to $0 \leq \beta_1 \leq 1$, $0 \leq \beta_2 \leq 1$, and $0 \leq \beta_3 \leq 1$.

In our example, application of quadratic programming procedure, equation (19), gives zero as the mode of $\beta_1$. This is what it should be because truncation of $\beta_1$ occurs at zero. A zero estimate for $\beta_1$ is not very interesting, at least not in our case.

As suggested in the second method, posterior means and posterior variances of $\beta_1$, $\beta_2$, and $\beta_3$ are obtained from equation (18) by numerical integrations over the restricted space of $\beta_1$, $\beta_2$, and $\beta_3$. Posterior mean of $\beta_2$ is obtained from the following relation derived from equation (17) after simplification:

\[
(20) \quad \hat{\beta}_2 = \hat{\beta}_2 - h_{22}^{-1} h_{21} [\hat{\beta}_1 - \beta_1, (\beta_2 - \hat{\beta}_2), (\beta_3 - \hat{\beta}_3)',
\]
where $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$ are posterior means of $\beta_0$, $\beta_1$, $\beta_2$, and $\beta_3$. Table 2 summarizes the results of the Bayesian estimates.

Results obtained by the Bayesian approach are consistent with prior knowledge. Also, the posterior variances of $\beta_1$, $\beta_2$, and $\beta_3$ are lower than the variances given by least squares. Besides correcting $\beta_3$, the Bayesian approach also corrected other coefficients. We believe that more correct estimates of all the parameters are obtained by the Bayesian approach. Our belief is based on the premise that multicollinearity usually distorts the estimates of all the parameters and not just one. Restricted least squares will give zero estimate for $\beta_1$ in our example.

Conclusions

The method of estimation given in the preceding sections is quite general and is applicable to the class of problems in regression analysis when a subset or all of the parameters are known to lie within certain ranges based on a priori knowledge and when least squares produce results which are inconsistent with that knowledge. Prior distributions other than uniform can also be taken. Computational problems involving numerical integrations become increasingly difficult with the increase in the number of parameters, although they are not impossible to handle on large computers.

[Received February 1974; revision accepted January 1975.]

References


