Multiple Regression with Inequality Constraints: Pretesting Bias, Hypothesis Testing and Efficiency

MICHAEL C. LOVELL and EDWARD PRESCOTT*

This article analyzes, within the context of the standard multiple regression model, the problem of handling inequality constraints specifying the signs of certain regression coefficients. It is common econometric practice when regression coefficients are encountered with incorrect sign to delete the variables in question and reestimate the equation. This article shows that this procedure causes bias and can lead to inefficient parameter estimates. Furthermore, we show that grossly exaggerated statements concerning significance levels are likely to be made when other regression coefficients in the model are tested with the final regression obtained after deleting variables with incorrect sign.

1. INTRODUCTION

Tossing a fair coin is not necessarily a fair game—the artful player with a naive opponent can raise his probability of winning to 5/8 by suggesting “two out of three” whenever he fails to win on the first toss. The artful econometrician often fits several models experimentally to the same body of data in an attempt to find a model that looks right in terms of such criteria as the conformity of the signs of regression coefficients to a priori notions, their significance, the Durbin-Watson statistic and the magnitude of $R^2$. We shall consider the implications of a particular aspect of this art with regard to problems of parameter estimation and hypothesis testing.

Suppose that an econometrician is interested in the following model,

$$ y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + \cdots + \beta_k x_{kt} + \epsilon_t, \quad t = 1, \cdots, T $$

(1.1)

where $x_{it}$ denotes observation $t$ on the $i$th independent variable, and $\epsilon_t$ is an independently distributed random variable with finite variance $\sigma^2$ and zero expected value.\(^1\) Now suppose that economic theory suggests as part of the maintained hypothesis the condition $\beta_1 \geq 0$. We wish to evaluate the following two-step procedure for using the a priori information on the sign of $\beta_1$ in estimating the $\beta_i$:

For the first step, fit by least squares the regression

$$ y_t = b_1 x_{1t} + \cdots + b_k x_{kt} + \epsilon_t. $$

(1.2)

If $b_1 \geq 0$, use the $b_i$ as estimators of the $\beta_i$. Else, if $b_1 < 0$, use the second-step estimators $b_1^* = 0$ and, for $i > 1$, the coefficients $b_i^*$ obtained from the regression:

---

*Michael C. Lovell is professor of economics, Wesleyan University. Edward Prescott is on leave from the Department of Economics, University of Pennsylvania, as a Brookings Economic Policy Fellow, Department of Labor. The authors are indebted to Robert Summers, John W. Pratt and Arnold Zellner for helpful advice and comments on a preliminary draft. Computations were executed on the University of Pennsylvania computer. This article was supported by the Graduate School of Industrial Administration, Carnegie-Mellon University, Wesleyan University, the Wharton School of Finance and Commerce and the National Science Foundation.

\(^1\) The intercept may be introduced, in accordance with the standard convention, by making one of the explanatory variables identically equal to unity for all $t$. 
\[ y_t = b_2 x_{2t} + \cdots + b_k x_{kt} + \epsilon_t^*. \]  

(1.3)

Thus, the two-step estimator, call it \( \hat{b}_i \), of \( \beta_i \) is

\[ \hat{b}_i = \begin{cases} 
  b_i & \text{if } b \geq 0 \\
  b_i^* & \text{otherwise.}
\end{cases} \]  

(1.4)

This two-step procedure seems eminently reasonable.\textsuperscript{2} It constitutes a statistical application of quadratic programming.\textsuperscript{3} It is probably the approach most likely to be used in practice when regression coefficients with incorrect sign are encountered.\textsuperscript{4} Note that \( b_i \) will constitute the best linear unbiased estimator of \( \beta_i \) if the \( \epsilon_i \) are independently distributed with constant variance, but this does not imply that it is preferable to the non-linear estimator \( \hat{b}_i \). Indeed, \( \hat{b}_i \) is consistent.\textsuperscript{5} Further, it is the maximum likelihood estimator of \( \beta_i \) if the disturbance \( \epsilon_i \), in addition to the other restrictions, is normally distributed.

For finite samples, is the two-step procedure necessarily preferable to the use of the first-step estimators, the \( b_i \) that are derived without regard for the sign constraint? A simple contrived example which we present in Section 2 will confirm that our intuition is correct if it suggests that the two-step estimators, the \( \hat{b}_i \), may be biased. However, in evaluating any estimator, bias is not necessarily the most critical property to consider. We must also consider the efficiency of the two-step estimator; an appropriate measure is provided by the mean-square-error—\( E[(\hat{b}_i - \beta_i)^2] \). Our numerical example will reveal that contrary to intuition, estimators derived by ignoring inequality constraints may be more efficient (have a smaller expected square error) than the two-step estimators \( \hat{b}_i \). Since it is common econometric practice to report two-step estimators satisfying inequality constraints, it is of obvious interest to determine whether this approach, in practice, is likely to constitute an inefficient estimation procedure. In subsequent sections of this article, we shall explore this issue, and in addition, the question of bias. Our analysis will build in part on the work of Zellner [9], who was concerned with the distribution of \( b_1 \) in the case of simple regression.\textsuperscript{6}

It is common practice, for purposes of hypothesis testing, to report \( t \) statistics for whichever computer run yields the regression equation satisfying the sign

\textsuperscript{2} The problem is not totally unrelated to the commonly used procedure of deleting variables that have insignifi-

\textsuperscript{3} There exists a considerable literature, reviewed by [2], concerning the computational problem of minimizing

\textsuperscript{4} It is not the only procedure. Some practicing econometricians, considering a wrong sign indicative of a mis-

\textsuperscript{5} The asymptotic properties of a class of estimators of which \( \hat{b}_i \) is a member are developed in [4, Chapter 9].

\textsuperscript{6} Zellner's contribution is summarized in [4, p. 317]. Much of the argument of this article is readily generalized
to the case in which the switching rule hinges on whether a linear combination of the \( \beta_i \) is subject to an inequality

constraint.
constraints imposed by theory on the model. That is, for \( i > 1 \) the hypothesis testing analogue of the two-step estimation procedure at the \( \alpha \)-significance level is provided by the following decision rule:

\[
\text{Accept } H_0; \beta_i = 0, \quad i > 1 \text{ if } b_1 \geq 0 \text{ and } \left| t_i \right| \leq t_\alpha \\
\text{or} \\
b_1 < 0 \text{ and } \left| t_i^* \right| \leq t_\alpha^*,
\]

(1.5)

else reject \( H_0 \).

Here the statistic \( t_i \) is computed by standard formula along with regression Equation (1.2), which contains the variable \( x_{1t} \); \( t_i^* \) is computed from the second regression, which excludes the first explanatory variable; further, \( t_\alpha \) and \( t_\alpha^* \), the coefficients defining the widths of the acceptance regions, are determined for desired significance level \( \alpha \) from standard tables in the customary fashion. We demonstrate in Section 5 that the use of this two-step procedure in testing hypotheses about a model's parameters is quite likely to exaggerate significance levels; i.e., the probability of committing a Type I error, of rejecting the null hypothesis when it is true, may be much larger than the claimed significance level.

2. A CONTRIVED EXAMPLE

Here is a simple, contrived example designed to illustrate that estimates provided by the two-step procedure, the \( \hat{b}_i \), do not necessarily dominate the \( b_i \) estimates obtained by ignoring the inequality constraint. The example involves only two observations on each of two explanatory variables. We suppose that the econometrician correctly assumes that his data is generated by the following model:

\[
y_t = \beta_1 + \beta_2 x_{2t} + \epsilon_t \\
\beta_1 \geq 0 \\
E(\epsilon_t) = E(\epsilon_t x_{2t}) = 0.
\]

(2.1)

Table 1 reveals the possible sample outcomes that would be generated if the actual structure involves \( \beta_1 = .5, \beta_2 = 1, \) and \( P(\epsilon_t = 1) = P(\epsilon_t = -1) = 0.5 \), while the explanatory variable takes on the values \( x_{21} = -2 \) and \( x_{22} = 1 \). The first four columns of data on the table reveal, for each of the four possible outcomes of the experiment, the values of the dependent variables and the various estimators. Note that \( E(\hat{b}_1) = 5/8 \neq E(b_1) = \beta_1 = .5 \) and \( E(\hat{b}_2) = 41/40 \neq E(b_2) = \beta_2 = 1 \); thus, the two-step estimators are biased.

On the figure the four \( S_t \) lines indicate the regression outcome computed from each of the four equally likely samples. Because the disturbance is distributed independently of \( x_{it} \), their average slope is unity, although of course some lines are flatter and some are steeper. Only the first sample regression line, which has a slope of unity, has an inadmissible negative intercept. The dotted

---

1 We have a special case of the problem posed in Section 1, with \( x_{11} = x_{12} = 1 \), in order that \( \beta_1 \) will constitute the intercept.
S1* line indicates the second-step regression computed with the intercept excluded; line S1* is steeper than unity because point x21 is further than x22 from the origin. Consequently, the invocation of the sign constraint on the intercept serves to replace, with probability of 1/4, the line S1, which has a slope equal to β2, with the steeper line S1*. Since an error is introduced whenever Sample 1 is drawn, but none of the other sample outcomes are changed, it is clear that in this instance the consequence of the two-step procedure is inefficiency as well as bias.

From the point of view of efficiency, we note that for this example ̂b1 is preferable to b1 as an estimator of β1 because it has the smaller mean square error (MSE); that is, MSE(b1) = E[(b1 − β1)2] = 5/9 while with the two-step procedure we have MSE( ̂b1) = 53/144. The two-step procedure thus serves to cut by about one-third the expected loss in estimating β1, if the statistician is penalized in proportion to the square of his estimation error. For the other parameter, however, we have MSE(b2) = 2/9 while MSE( ̂b2) = 2/9 + 1/400. In this particular case, the unfortunate effect of utilizing the two-step procedure is to increase the mean-square-error.8

We explore in subsequent sections the pros and cons of the two-step estimation procedure. Section 3 derives a precise expression for the bias involved in the two-step procedure and shows how the econometrician can place bounds on the extent of the bias involved in practical applications. Then, Section 4 demonstrates that a quite natural restriction will insure that even for small samples the two-step estimation procedure yields more efficient estimators than those obtained by ignoring the inequality constraints. Section 5 is concerned with problems of hypothesis testing.

---

8. Rotemberg [7] independently constructed an example in which he also showed that MSE( ̂b2) can exceed MSE(b2). He conjectures that if the disturbance is normal and β is constrained to a convex set, the constrained least squares estimator cannot have larger expected quadratic loss than the unconstrained estimator. In Section 4, we prove the validity of a restrictive case for this conjecture.
The bias in the two-step estimation procedure is most easily analyzed with the aid of matrix notation. If we let \( y = \text{col}(y_i) \), \( x_i = \text{col}(x_{ii}) \), \( \epsilon = \text{col}(\epsilon_i) \), \( X_2 = [x_2, \cdots, x_k] \), and \( \beta_2 = \text{col}(\beta_2, \cdots, \beta_k) \) where \( \text{col} \) indicates a column vector, we can rewrite (1.1) as

\[
y = x_1\beta_1 + X_2\beta_2 + \epsilon. \tag{3.1}
\]

We now transform the problem into orthogonal form by defining the vector 
\[
x_1^* = x_1 - X_2(X_2'X_2)^{-1}X_2'x_1 = x_1 - X_2B_{1.2},
\]
where \( B_{1.2} \) is the vector of \( k-1 \) coeffi-
cient from the auxiliary regression of the first explanatory variable on the remaining. We then note that \( x_1 = x_1^* + X_1B_{1.2} \), and obtain on substitution into (3.1):

\[
y = x_1^*\beta_1 + X_2(\beta_2 + B_{1.2}\beta_1) + \epsilon \tag{3.2}
\]

Since \( X_2'x_1^* = 0 \), application of least squares to (3.2) yields the estimates

\[
b_1 = (x_1^*x_1^*)^{-1}x_1^*y \tag{3.3}
\]

\[
B_2 + B_{1.2}b_1 = (X_2'X_2)^{-1}X_2'y. \tag{3.4}
\]

Now the vector on the right of (3.4) is the set of regression coefficients of the second-step regression involving the deletion of the first variable, which we denote as \( B_2^* = \text{col}(b_2^*, \ldots, b_k^*) \). Thus we may rewrite (3.4) as:

\[
B_2^* = B_2 + B_{1.2}b_1. \tag{3.5}
\]

Further, our two-step estimator is

\[
b = \max(0, b_1) = b_1 - \min(0, b_1)
\]

and the \( k-1 \) component column vector

\[
\vec{b}_2 = \text{col}(b_2, \ldots, b_k) = B_2 + \min(0, b_1)B_{1.2}. \tag{3.6}
\]

To analyze the bias, note that \( E(b_1) = \beta_1 \) and \( E(B_2) = \beta_2 \); therefore:

\[
E\left[\begin{bmatrix} b_1 \\ \vec{b}_1 \end{bmatrix}\right] - \beta = E[\min(0, b_1)]\begin{bmatrix} -1 \\ B_{1.2} \end{bmatrix}. \tag{3.7}
\]

Now we note that \( E[\min(0, b_1)] < 0 \), unless \( b_1 < 0 \) is impossible (\( \vec{b}_1 = b_2 \)). Except in this uninteresting case, \( \vec{b}_1 \) has positive bias. Furthermore,

\[
E(\vec{b}_i) - \beta_i = E[\min(0, b_1)]b_{1.i}. \tag{3.7'}
\]

Here \( b_{1.i} \) denotes the relevant component of vector \( B_{1.2} \); i.e. \( B_{1.2} = \text{col}(b_{1.2}, \ldots, b_{1.k}) \). Thus, the remaining coefficients will be biased unless either \( \vec{b}_1 \) is unbiased or the \( i \)th explanatory variable enters with zero coefficient in the auxiliary regression of the first explanatory variable on the remaining \( k-1 \) explanatory variables.

If the disturbance term \( \epsilon \) in equation (1.1) is normally distributed, tables in [9, p. 10] which are reproduced in [4, p. 317] can be conveniently used to compute the bias as a function of the unknown parameter \( \beta_1 \), the variance \( \sigma_\epsilon^2 \) of the disturbances, and the explanatory variables.\(^9\) When \( \beta_1 \) is not known, a bound on the bias is implied by the knowledge that this parameter is non-negative, for we note that when \( \beta_1 = 0 \) the bias is maximum and then

\[
E(\vec{b}_i) - \beta_i = - E[\min(0, b_1)] = \sigma_\epsilon/\sqrt{2\pi} \approx .4\sigma_\epsilon,
\]

and for \( i > 1 \),

\[
E(\vec{b}_i) - \beta_i = b_{1.i}\sigma_\epsilon/\sqrt{2\pi} \approx .4\sigma_\epsilon b_{1.i}.
\]

\(^9\) The following formula obtained by integration, may be used to compute the bias:

\[
E(b_i) - \beta_i = \left[ \varphi(s) - f(s)\varphi(s) \right] b_{1.i},
\]

where \( s = \beta_1/\sigma_\epsilon \) and \( f \) and \( \varphi \) are the normal density and distribution functions respectively. This is the expression used by Zellner [9] to compute his table; Raiffa and Schlaifer [6, p. 356] tabulate the expression in brackets.
4. EFFICIENCY

Let us now compare estimator $b_i$ with the two-step estimator $\hat{b}_i$ in terms of the mean-square-error criterion. Our numerical example illustrated that it is possible for $\hat{b}_i$ to have a larger mean square error than $b_i$. But we shall demonstrate that $b_i$ will necessarily have the smaller mean square error if we require, as an additional restriction, that the disturbance $\epsilon$ be normally distributed.

In order to compute the difference in mean square error, we note

$$(b_i - \beta_i)^2 - (\hat{b}_i - \beta_i)^2 = (b_i - \hat{b}_i)(b_i + \hat{b}_i - 2\beta_i).$$  \hfill (4.1)

If $b_i \geq 0$, $b_i = \hat{b}_i$ and this expression is 0. If $b_i < 0$, $b_i = b_i^* = b_i + b_1, \beta_1$ so we obtain

$$\text{MSE}(b_i) - \text{MSE}(\hat{b}_i) = E[(b_i - \beta_i)^2 - (\hat{b}_i - \beta_i)^2]$$

$$= P[b_i < 0]E[-b_1, \beta_1(2b_i^* - 2E(b_i^*)) - b_1, \beta_1 + 2b_1, \beta_1 | b_i < 0].$$  \hfill (4.2)

Because $b_i$ and $b_i^*$ are independently distributed, taking the expectation yields

$$\text{MSE}(b_i) - \text{MSE}(\hat{b}_i) = P(b_i < 0)b_i^2, E[(b_i^2 - 2b_1\beta_1) | b_i < 0].$$  \hfill (4.3)

This expression is necessarily non-negative; for the term inside the expectations operator cannot be negative because of the condition on $b_i$. We have thus established that the two-step estimation procedure is more efficient than the unconstrained estimator, $b_i$. The relative gain in mean square error is

$$\frac{\text{MSE}(b_i) - \text{MSE}(\hat{b}_i)}{\text{MSE}(b_i)}$$

$$= P(b_i < 0)b_i^2, E[(b_i^2 - 2b_1\beta_1) | b_i < 0] / \text{MSE}_{b_i}$$

$$= \rho_{ii}^2 \int_{-\infty}^{0} t_i^2 \frac{2t_i}{\sqrt{2\pi}} \exp[-(t_i - \tau)^2/2]dt_i$$

$$= \rho_{ii}^2 \left[ \frac{\tau}{\sqrt{2\pi}} \exp(-\tau^2/2) + \int_{-\infty}^{-\tau} \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)dx \right]$$

$$= \rho_{ii}^2 \left[ \tau f(\tau) + (1 - \tau^2) F(-\tau) \right],$$  \hfill (4.4)

where $t_i = b_i/\sigma_{b_i}$, $\tau = \beta_1/\sigma_{b_1}$, $\text{MSE}(b_i) = \sigma_{\beta_i}^2$, $\rho_{ii} = -b_1, \sigma_{b_1}/\sigma_{b_i}$, and $f$ and $F$ are the normal density and distribution functions.

Consideration of (4.4) reveals that the reduction in the mean square error obtained by using the two-step estimator under conditions of normality is proportional to $\rho_{ii}^2$, which can be calculated from the matrix of exogenous variables.\(^\text{11}\) Unless $x_{1i}$ and $x_{2i}$ display a high degree of colinearity, $\rho_{ii}^2$ will be small which implies the gain in efficiency will be small. Given $\rho_{ii}^2$, the gain in efficiency is maximized when $\beta_1 = 0$; indeed, for $\beta_1 = 0$ we find that the relative

---

\(^{\text{10}}\) Both $b_i^*$ and $b_i$ are linear combinations of the disturbance terms and therefore are jointly normal. Since uncorrelated jointly normal variates are independent, it suffices to show the covariance between $b_i^*$ and $b_i$ is 0. From (3.3) and (3.5) we derive

$$E \{ (b_i - \beta_i)[B_i - E(B_i)] \} = E \{ (\epsilon x_i^\ast x_{i1}^\ast - x_i^2s_{i1}^\ast - X_{i1}X_{i1} - X_{i1}^2s_{i1}^\ast) \} = 0$$

once we remember $x_{i1}^\ast x_{1i} = 0$.

\(^{\text{11}}\) Let $C = [X'X]^{-1}$; then $\rho_{11} = \epsilon_{11}/(\epsilon_{11}11).$
gain in mean square error is .5°C_{11}. As β increases the efficiency gain decreases; thus, for β/σ = 1, the difference is approximately C_{11}/10.

5. HYPOTHESIS TESTING

The true significance level of a test constructed at the .05 “significance level” by the two-step hypothesis testing procedure defined by equation (1.5) is revealed in Table 2. Clearly, the two-step hypothesis testing procedure may lead to exaggerated claims of significance. Inspection of the table reveals that the true probability of incorrectly rejecting the null hypothesis depends on both ρ_{11} and β/σ. The entries in Table 2 were calculated under the assumption that the disturbance ε in equation (1.1) is normally distributed. Although the probabilities reported are for the case in which the investigator appropriately incorporates knowledge of σ^2 in applying the two-step testing procedure, essentially the same results are obtained in the case of unknown variance with as few as six degrees of freedom.12

<table>
<thead>
<tr>
<th>Table 2. PROBABILITY OF INCORRECTLY REJECTING H_0: β = 0 WITH THE TWO-STEP TEST AT THE FIVE PERCENT SIGNIFICANCE LEVEL</th>
</tr>
</thead>
<tbody>
<tr>
<td>(</td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td>.50</td>
</tr>
<tr>
<td>.75</td>
</tr>
<tr>
<td>.90</td>
</tr>
<tr>
<td>.99</td>
</tr>
</tbody>
</table>

Table 2 reveals that the two-step hypothesis testing procedure exaggerates the significance level only when \( |ρ_{11}| \) is large and, at the same time, \( β_{1}/σ_b \) is of moderate magnitude. To see why this is so, recall first that under the null hypothesis

\[ P(|t_1| > t_{a}) = \alpha. \]

If, in addition, \( ρ_{11} = 0 \), then also \( t_1 = t_{1,*} = t_{1}, \)13 which explains why entries in the top row of Table 2 are all 0.05. Table 2 reveals that this effect persists even for \( ρ_{11} = 0.75 \).

Trouble can arise, however, with the larger values of \( |ρ_{11}| \) frequently encountered when dealing with highly collinear economic time series. To appreciate the circumstances in which the two-step procedure overstates the true significance level, we note that the probability of rejecting \( H_0: β = 0 \) with the two-step hypothesis testing procedure is

\[ P(|\hat{t}_1| > t_a) = P(|t_{1,*}| > t_{a,*} \cap b_1 < 0) + P(|t_{1}| > t_a \cap b_1 \geq 0). \] (5.1)

12 When the investigator knows σ, he may use the normal distribution to determine t_a, which defines the width of the acceptance region; further, σ_b = c_0σ, and \( t_1 = b_2/σ_b \). When σ is unknown the investigator must substitute its unbiased estimator and utilize the t distribution. The appendix analyzes the additional complications for the two-step hypothesis testing procedure in the case of unknown variance.

13 When \( ρ_{11} = 0, b_1 = b_2 = b_3 \), but this implies \( t_1 = t_{1,*} = \hat{t}_1 \) only in the case of known σ. The result is only approximately true when σ is not known. See the appendix for further explanation.
From Equation (3.5) we observe that \( b_i^* \) will be subject to substantial bias when \( p_{it} \) and \( \beta_i \) both depart considerably from zero; in these circumstances, 
\[
P(|t_i^*| > t_i^*) \]
will exceed the claimed significance level by a substantial margin, and it is not surprising, in light of (5.1), that the two-step hypothesis testing procedure may incorrectly reject the null hypothesis with excessive frequency.

Even when the problem of collinearity is severe, the two-step test may incorrectly reject the null hypothesis with the claimed probability. In particular, when \( \beta_i = \beta_i = 0 \), from (3.5) the mean of \( b_i^* \) is zero. It is not surprising that under these conditions the two-step test has rejection probability \( \alpha \). This explains why all the entries in the first column of Table 2 equal 0.05 regardless of the degree of multicollinearity. At the other extreme, as \( \beta_i/\sigma_{bi} \) becomes large, 
\[
P(b_i < 0) \]
approaches zero and the two-step test becomes equivalent to the use of \( t_i \); thus, the actual probability of rejecting the null hypothesis approaches the claimed significance level as \( \beta_i/\sigma_{bi} \) becomes sufficiently larger.

It is necessary to concede that Table 2 cannot be used in practical applications to determine the true significance level generated by the two-step hypothesis testing procedure. An investigator can compute \( p_{it} \) from the matrix \( X \) of observations on the explanatory variables, but this reveals only the relevant row of Table 2. The true significance level of the two-step hypothesis testing procedure cannot be determined because only the sign rather than the magnitude of \( \beta_i \) is known. This table does reveal that the true significance level cannot depart too much from the stated level when the data are not very collinear. But the bottom rows of this table show that the range of indeterminacy may be quite large when dealing with highly correlated economic time series. Although an investigator willing to invoke a subjective probability distribution on the parameter \( \beta_i \) can evaluate the significance of the two-step testing procedure, he would want to formulate the problem from the beginning in Bayesian rather than classical terms.

Since it is impossible in practical applications to determine the true significance level of the two-step hypothesis testing procedure, it is essential to consider the possibility of always using the statistic \( t_i \) provided by regression equation (1.2), which includes \( x_{1t} \). An investigator who uses \( t_i \) for his test, regardless of the sign of \( b_i \), is in effect ignoring the a priori information that \( \beta_i \) is non-negative, but at least he knows the true significance level of his test. That the costs of ignoring a priori information may not be high in this particular application is suggested by the comparison of alternative test procedures presented in Table 3. For each test Table 3 reveals how the probability of rejecting the null hypothesis varies with \( \beta_i/\sigma_{bi} \) and \( \beta_i/\sigma_{bi} \) when \( p_{11} = 0.9 \). The upper part of Table 3 reveals the probability of rejecting \( H_0: \beta_i = 0 \) for the two-step testing procedure at a claimed significance level of \( \alpha = 0.05 \). The last two rows report the probabilities of rejecting the null hypothesis using the \( t_i \) statistic provided by regression (1.2) regardless of the sign of \( b_i \) at the 0.05 and the 0.11 significance levels; both these tests ignore the a priori information that \( \beta_i \geq 0 \). We note that the two-step hypothesis testing procedure has a higher probability of rejecting the null hypothesis when it is false than the simpler five percent significance level test that uses the results of regression (1.2) regardless of the sign of \( b_i \). This does not mean that the two-step procedure dominates; the true significance

\[ \text{A rigorous argument for this result can be found in the appendix.} \]
level of that test is not 0.05. For example, when $\beta_1/\sigma_{b_1} = 1.0$, the probability of incorrectly rejecting $H_0$ with the two-step test is eleven percent rather than five percent. A more meaningful comparison is facilitated with the test on the bottom row of Table 3 for $\alpha = .11$; it has approximately the same significance level as the two-step test. Observe that neither test dominates the other. It is unfortunate that in practical applications the very nature of the problem is such as to deny the investigator the information required to determine which procedure is more powerful. The test that ignores the sign condition does have the very important advantage that tables are available that can be used to determine the acceptance region yielding the claimed significance level.

Additional information on the relative merits of the two alternative testing procedures is presented in Table 4. Because only a moderate degree of multicollinearity is involved, the two-step test is significant at the claimed level of five percent. The two-step test is not uniformly more powerful. Indeed, neither test dominates the other. Clearly, it is not inappropriate to use the two-step hypothesis testing procedure when $|p_{t1}|$ is of small or moderate magnitude.

6. CONCLUSION

Our analysis suggests that for purposes of least square parameter estimation it is appropriate practice to delete a variable with incorrect sign. Although this customary procedure generates bias, it leads to more efficient parameter esti-

<table>
<thead>
<tr>
<th>Probability $\beta_1/\sigma_{b_1}$</th>
<th>$\beta_1/\sigma_{b_1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$-3.0$</td>
</tr>
<tr>
<td>0</td>
<td>.85</td>
</tr>
<tr>
<td>.25</td>
<td>.85</td>
</tr>
<tr>
<td>$P(</td>
<td>t_1</td>
</tr>
<tr>
<td>1.00</td>
<td>.85</td>
</tr>
<tr>
<td>2.00</td>
<td>.85</td>
</tr>
<tr>
<td>$P(</td>
<td>t_1</td>
</tr>
<tr>
<td>$P(</td>
<td>t_1</td>
</tr>
</tbody>
</table>

Table 4. POWER FUNCTIONS WHEN $\rho_{t1} = .50$
mates when the error term is normally distributed as well as independent of the explanatory variables. The customary procedure for hypothesis testing when sign constraints are present is inappropriate when dealing with highly collinear economic time series, for it can lead to substantially exaggerated claims of significance. A valid test, which may be more powerful than the customary procedure, is to always report the $t$ statistics calculated from the initial regression regardless of the signs of the parameter estimates.

APPENDIX: HYPOTHESIS TESTING

In order to simplify the analysis, we normalize by measuring $y_t$, $x_{it}$, and $x_{1t}$ in units such that $\sigma_e = \sigma_{b_1} = \sigma_{b_i} = 1$. Nonetheless, the results of this appendix are completely general for the arguments are presented in terms of parameters which are invariant to the required transformation. We assume $\epsilon$, the disturbance term, has unknown variance and, of course, that it is normally distributed.

Let $S$ be the sum of the squared residuals for regression (1.2), which includes $x_{1t}$. It has a chi-square distribution with $T - k$ degrees of freedom and is independent of both $b_1$ and $b_i$. Further, since $b_1^*$ is a linear combination of $b_1$ and $b_i$, $S$ is independent of $b_1^*$ as well. Using (3.5), we find

$$b_i/\sigma_b = (1 + \rho_{i2})^{1/2}b_i^* + \rho_{1i}b_1. \quad (A.1)$$

Consideration of (3.2) suggests the vector of residuals for regression (1.2) is

$$e = y - X_2B_2^* - x_1^*b_1, \quad (A.2)$$

and for regression (1.3), which excludes $x_{1t}$,

$$e^* = y - X_2B_2^*. \quad (A.3)$$

Since $S = e'e$, $S^*$ the sum of the squared residuals for the regression which excludes $x_{1t}$ is

$$S^* = e^*e^* = (e + x_1^*x_1^*b_1)'(e + x_1^*b_1)$$

$$= S + b_1x_1^*x_1^*b_1 + 2b_1x_1^*e. \quad (A.4)$$

But,

$$x_1^*e = x_1^*(y - X_2B_2^* - x_1^*b_1) = x_1^*y - x_1^*x_1^*b_1 = 0 \quad (A.5)$$

as $x_1^*X_2 = 0$, $x_1^*x_1^* = 1$, and $b_1 = (x_1^*x_1)^{-1}x_1^*y$. Thus, we obtain the result,

$$S^* = S + b_1^2. \quad \text{(A.5)}$$

Consideration of (3.2) suggests the vector of residuals for regression (1.2) is

$$e = y - X_2B_2^* - x_1^*b_1, \quad (A.2)$$

and for regression (1.3), which excludes $x_{1t}$,

$$e^* = y - X_2B_2^*. \quad (A.3)$$

Since $S = e'e$, $S^*$ the sum of the squared residuals for the regression which excludes $x_{1t}$ is

$$S^* = e^*e^* = (e + x_1^*x_1^*b_1)'(e + x_1^*b_1)$$

$$= S + b_1x_1^*x_1^*b_1 + 2b_1x_1^*e. \quad (A.4)$$

But,

$$x_1^*e = x_1^*(y - X_2B_2^* - x_1^*b_1) = x_1^*y - x_1^*x_1^*b_1 = 0 \quad (A.5)$$

as $x_1^*X_2 = 0$, $x_1^*x_1^* = 1$, and $b_1 = (x_1^*x_1)^{-1}x_1^*y$. Thus, we obtain the result,

$$S^* = S + b_1^2. \quad \text{(A.5)}$$

Given the above discussion, the relevant $t$ statistics for the two-step test can be represented as

$$t_i = \frac{b_i/\sigma_{b_i}}{[S/(T - k)]^{1/2}} = \frac{(1 - \rho_{i2})^{1/2}b_i^* + \rho_{1i}b_1}{[S/(T - k)]^{1/2}} \quad (A.6)$$

and

$$t_i^* = \frac{b_i^*}{[S^*/(T - k + 1)]^{1/2}} = \frac{b_i^*}{[(S + b_1^2)/(T - k + 1)]^{1/2}}. \quad (A.7)$$

The random variables $S$, $b_1$, and $b_i^*$ are independently distributed. $S$ has a chi-square distribution with $T - k$ degrees of freedom, while $b_1$ and $b_i^*$ are normally
distributed with means $\beta_1/\sigma_b$ and $(\beta_i/\sigma_b - \rho_{ij} \beta_j/\sigma_b)/(1 - \rho_{ij}^2)^{1/2}$ respectively and variances 1.

When $\beta_1 = \beta_i = 0$, the likelihood of $(t_1, b_1)$ equals the likelihood of $(-t_i, -b_i)$. This implies $P(\{ t_i > t_a \cap b_1 \geq 0 \}) = P(\{ t_i > t_a \})/2 = \alpha/2$. For $t_i^*$ and $b_1$, the almost identical argument implies $P(\{ t_i^* > t_a^* \cap b_1 < 0 \}) = \alpha/2$ as well. Substituting into (5.1), we find the probability of incorrectly rejecting $H_0$ when $\beta_1 = 0$ is $\alpha$, the claimed significance level for the two-step test.

That the two-step testing procedure is approximately valid when $\rho_{ij} = 0$ is easily demonstrated once we note that this condition implies that $b_1$ and $b_i$ are independently distributed; since $S$ is also independently distributed, we have under $H_0$

$$P(\{ | t_i | > t_a \cap b_1 \geq 0 \}) = P(b_1 \geq 0)P(\{ | t_i | > t_a \}) = P(b_1 \geq 0)\alpha.$$ 

Furthermore,

$$P(\{ | t_i^* | > t_a^* \cap b_1 < 0 \}) = P(b_1 < 0)P(\{ | t_i^* | > t_a^* \}) = P(b_1 < 0)\alpha,$$

provided there are more than a few degrees of freedom. Hence, we have upon substitution into (5.1) that $P(\{ | t_i | > t_a \}) = \alpha$ under the null hypothesis when $\rho_{ij} = 0$.

To see how the entries in Tables 2, 3, and 4 were calculated, we first define

$$p(b_1, S) = \begin{cases} P(\{ | t_i | > t_a \mid b_1, S \}) & \text{for } b_1 \geq 0 \\ P(\{ | t_i^* | > t_a^* \mid b_1, S \}) & \text{for } b_1 < 0. \end{cases}$$ (A.8)

Inspection of (A.6) and (A.7) reveals that this function can be evaluated using the normal distribution function. Next we observe

$$P(\{ | t_i | > t_a \}) = E[p(b_1, S)].$$ (A.9)

This expression was evaluated using numerical integration for several of the parameters $\rho_{ij}, \beta_1/\sigma_b$, and $\beta_i/\sigma_b$ when $T-k=6$ and 18 as well as when $\sigma_i$ is known. In this latter case $S$ will not be an argument of $p(\cdot, \cdot)$ as $t_i = b_i/\sigma_{b_i}$ and $t_i^* = b_i^*$. Note that all these parameters are invariant to the units in which the variables are measured. The difference in power between the two-step test and the simple test which uses only $t_i$ and ignores the sign constraint is approximately the same when $T-k=6$ and 18 and when the variance of the disturbance term is known. We were not surprised to find the power is greatest for known variance and least for $T-k=6$ and unknown variance. When $T-k=6$ the significance level of the test is not exaggerated by as great a margin as when there are more degrees of freedom.

REFERENCES
