CONSISTENT SETS OF ESTIMATES

FOR REGRESSIONS WITH ERRORS IN ALL VARIABLES

(Revised, March 1982 and December, 1982)

by

Steven Klepper and Edward E. Leamer*

Discussion Paper # 282

* Department of Social Sciences, Carnegie-Mellon University and Department of Economics, UCLA, respectively. Leamer's work has been supported by NSF grant SES78-09477. Comments from Dee Dechert, Steven Garber, Joseph Kadane, Rudolf Kalman, and David Levine and referees are gratefully acknowledged.

Consistent Sets of Estimates

For Regressions With Errors in All Variables

(Revised, March, 1982 and December, 1982)

by

Steven Klepper and Edward E. Leamer*

1. Introduction

In recent years there has been a resurgence of interest in the errors-in-variables model (EVM), or more generally unobservable variable models. In a sense, this is a reversion to the kind of models considered in the early history of econometrics. As Goldberger (1972, p. 993) notes in his retrospective essay on structural equation methods, "In the early days, economic equations were formulated as exact relationships among unobservable variables, and errors in the variables provided the only stochastic component in the observations. But, at least since the days of the Cowles Commission, the emphasis has shifted almost entirely to errors in the equations."

Goldberger goes on to speculate about the reasons for this shift in emphasis. He notes (p. 993) that the classical EVM, as exemplified by Friedman's permanent income model, is underidentified, and "underidentified models present no interesting problems of estimation and testing." The modern approaches to the EVM "solve" this problem by introducing sufficient prior information to identify the parameters (of interest) of the model. This prior information may take the form of a known value for the covariance matrix of the measurement errors, or, more recently, instrumental variable equations linking the true but unobservable variables to a set of observable proxies.

---

*Department of Social Sciences, Carnegie-Mellon University and Department of Economics, UCLA, respectively. Leamer's work has been supported by NSF grant SES78-09477. Comments from Dee Dechert, Steven Garber, Joseph Kadane, Rudolf Kalman, and David Levine and referees are gratefully acknowledged.

While the modern approaches to the EVM represent a considerable advance, they are not a panacea. Frequently, the additional information needed for identification is either not available or is not widely shared by researchers in the field. Moreover, it may be possible to extract some information from a data set even if the model is underidentified. For example, for the two-variable EVM, Frisch (1934, pp. 58-60) demonstrates that the interval computed from the two regressions corresponding to each variable as left-hand-side variable "consistently bounds" the true regression coefficient. The purpose of this article is to present a complete k-variable generalization of Frisch's result. Because of the possibility of perfect collinearity among the true variables, the k-variable analog of the interval defined by the two regressions can be an unbounded region. For this case especially, we show how inequality constraints on error variances can limit the set of estimates.

Our strategy involves exploiting the fact that the estimated covariance matrices of the true (unobservable) variables and the measurement errors must be positive semi-definite. Assuming that all variables are distributed normally, this enables us to identify a restricted set of estimates of the true regression coefficient vector which maximize the likelihood function of the measured (i.e., observed) data. More generally, for any distribution of the variables this restricted set of estimates consistently bounds the true regression coefficient vector.

We consider first the case in which no prior information is available to supplement the basic EVM and it is assumed that all variables are normally distributed. We establish the following two basic results.

1) Suppose there are k regressors, each measured with error.

By varying the left-hand-side variable it is possible to compute k + 1 different regression vectors, which can be expressed in a common, normalized form. We demonstrate that if the k + 1 normalized regressions are all in the same orthant then: (a) any estimate which is a maximum likelihood estimate of the true regression coefficient vector must be in the convex hull of the k + 1 regressions; (b) every point in the convex hull of the k + 1 regressions is a maximum likelihood estimate of the true regression coefficient vector.

2) If the k + 1 regressions are not all in the same orthant then the set of maximum likelihood estimates is unbounded. In particular, any estimate of any component is a maximum likelihood estimate provided the other components are suitably selected.

Elaborate and somewhat unclear proofs of proposition 1(a) may be found in Koopmans (1937), Reiersol (1945), and Dhondt (1960). Independently of our work, Patefield (1981) provides a clear and concise proof of proposition 1(a) and Kalman proves both propositions 1(a) and 1(b). Proposition 2 is conjectured, though not proved, in Leamer (1978). We provide simple proofs of both proposition 1(a) and propositions 1(b) and 2. We also provide a complete characterization of the case of two regressors measured with error, for which it is possible to define completely the set of maximum likelihood estimates when the three regressions are not in the same quadrant.

It is to be expected that most nonexperimental data sets will generate regressions lying in different orthants. Some further prior information will then be needed to bound the set of maximum likelihood estimates. The natural candidate is information concerning the "seriousness" of the measurement errors in the measured regressors. In particular, we assume that it is generally possible to answer the following questions:

(a) What is the maximum value of the squared multiple correlation (R^2) if the measurement error of the explanatory variables were removed?

(b) What is the minimum possible squared correlation (ρ^2) between a true explanatory variable and its measurement?

Provided the answers are not one and zero, these questions imply bounds on the variances of the measurement errors of the observed regressors which in turn can imply restrictions on the set of maximum likelihood points.[1]

The answers to questions (a) and (b) are denoted by R^{*2} and ρ_*^2, and the limited feasibility of mapping out the maximum likelihood region as a function of R^{*2} and ρ_*^2 is explored. Two mathematically tractable suggestions are offered. First, ρ_*^2 is set to zero and the maximum likelihood region is traced out as a function of R^{*2}. A useful statistic is the maximum value of R^{*2} for which the region is bounded and, incidentally, entirely contained in the orthant of the conventional least-squares coefficient estimator. Second, the region as a function of ρ_*^2 is enclosed in an ellipsoid which can be economically computed and reported. The critical value of ρ_*^2, below which this ellipsoid is unbounded, is $1 - \lambda_1$, where λ_1 is the smallest eigenvalue of the correlation matrix of the measured explanatory variables.

The paper is organized as follows. In Section 2 we consider the general case of an arbitrary number of mismeasured regressors. In Section 3 we completely characterize the maximum likelihood region for the case of two regressors measured with error. In Section 4 we consider the use of prior information concerning the variances of the measurement errors in the measured regressors. An example is presented in Section 5 which illustrates the use of the principal results of the paper. Concluding remarks are offered in Section 6.

2. The Normal Errors-in-variables Structural Model

The normal errors-in-variables structural model is based on the following assumptions. An observation y_t is drawn from a normal distribution with mean $\beta_0 + \underset{\sim}{\beta}'\underset{\sim}{\chi}_t$ and variance σ^2, where $\underset{\sim}{\chi}_t$ is a (kxl) vector of unobservables and β_0, $\underset{\sim}{\beta}$ and σ^2 are the parameters of interest. The unobservables $\underset{\sim}{\chi}_t$ are measured by the vector $\underset{\sim}{x}_t$ which, conditional on $\underset{\sim}{\chi}_t$, is normally distributed with mean $\underset{\sim}{\chi}_t$ and covariance matrix $\underset{\sim}{D} = \text{diag}\,\{d_1, d_2, \ldots, d_k\}$. The unobservables $\underset{\sim}{\chi}_t$ come from a normal distribution with mean $\underset{\sim}{\bar{\chi}}$ and covariance matrix Σ.

These assumptions imply that the observable vector $(y_t, \underset{\sim}{x}_t')$ is normally distributed with moments

$$E(y_t, \underset{\sim}{x}_t') = (\beta_0 + \underset{\sim}{\beta}'\underset{\sim}{\bar{\chi}},\ \underset{\sim}{\bar{\chi}}')$$

$$V(y_t, \underset{\sim}{x}_t') = \begin{bmatrix} \sigma^2 + \underset{\sim}{\beta}'\underset{\sim}{\Sigma}\underset{\sim}{\beta} & \underset{\sim}{\beta}'\underset{\sim}{\Sigma} \\ \underset{\sim}{\Sigma}\underset{\sim}{\beta} & \underset{\sim}{\Sigma} + \underset{\sim}{D} \end{bmatrix}$$

Given a random sample of observations on $(y_t, \underset{\sim}{x}_t')$, the maximum likelihood estimates of β_0, $\underset{\sim}{\beta}$, and σ^2 can be found by setting these population moments equal to the corresponding sample moments. Given $\underset{\sim}{\beta}$, the k+1 sample means can be used to solve for the k+1 location parameters β_0, $\underset{\sim}{\bar{\chi}}$. This leaves the second order moments to estimate $\underset{\sim}{\beta}, \underset{\sim}{\Sigma}, \underset{\sim}{D}$, and σ^2:

$$s_y^2 = \sigma^2 + \underset{\sim}{\beta}'\underset{\sim}{\Sigma}\underset{\sim}{\beta} \quad (1)$$

$$\underset{\sim}{r}' = \underset{\sim}{\beta}'\underset{\sim}{\Sigma} \quad (2)$$

$$\underset{\sim}{N} = \underset{\sim}{\Sigma} + \underset{\sim}{D} \quad (3)$$

where s_y^2 is the sample variance of y, $\underset{\sim}{r}$ is the vector of sample covariances between y and $\underset{\sim}{x}$, and $\underset{\sim}{N}$ is the matrix of sample variances and covariances of $\underset{\sim}{x}$. For notational convenience, we have not distinguished between population parameters and estimates. Hereafter, $\underset{\sim}{\beta}$, $\underset{\sim}{\Sigma}$, $\underset{\sim}{D}$ and σ^2 refer to estimates.

An estimate $\underset{\sim}{\beta}$ is a maximum likelihood estimate if there exist positive semi-definite $\underset{\sim}{\Sigma}$, $\underset{\sim}{D}$ and σ^2 such that conditions (1), (2), and (3) are satisfied. Using (3) we can write (2) as $\underset{\sim}{r} = (\underset{\sim}{N}-\underset{\sim}{D})\underset{\sim}{\beta}$, from which we can derive $\underset{\sim}{D}$. If $\underset{\sim}{N}_i'$ is the ith row of $\underset{\sim}{N}$ then

$$d_i = \underset{\sim}{N}_i'(\underset{\sim}{\beta}-\underset{\sim}{b})/\beta_i \qquad (4)$$

where $\underset{\sim}{b}$ is the least-squares vector $\underset{\sim}{b} = \underset{\sim}{N}^{-1}\underset{\sim}{r}$. Also (2) can be inserted into (1) to obtain $\sigma^2 = s_y^2 - \underset{\sim}{r}'\underset{\sim}{\beta}$. Thus the positive semi-definite restrictions on σ^2, $\underset{\sim}{D}$ and $\underset{\sim}{\Sigma}$ can be written as

$$s_y^2 - \underset{\sim}{r}'\underset{\sim}{\beta} \geq 0 \qquad (5)$$

$$\underset{\sim}{N}_i'(\underset{\sim}{\beta}-\underset{\sim}{b})/\beta_i \geq 0 \qquad i = 1, \ldots, k \qquad (6)$$

$$\underset{\sim}{N} - \text{diag}\{\underset{\sim}{N}_i'(\underset{\sim}{\beta}-\underset{\sim}{b})/\beta_i\} \text{ pos. semi-definite} \qquad (7)$$

Any value of $\underset{\sim}{\beta}$ satisfying these restrictions is a maximum likelihood estimate.[2]

To gain insight into these conditions, consider (5) and (6). If the inequalities (5) and (6) are satisfied as equalities, the conditions can be written as

$$\underset{\sim}{y}'\underset{\sim}{e} = 0 \qquad (5')$$

$$\underset{\sim}{X}'\underset{\sim}{e} = \underset{\sim}{0} \qquad (6')$$

where $\underset{\sim}{y}$ and $\underset{\sim}{X}$ are respectively the vector and matrix of observations on y and $\underset{\sim}{x}$ with their means removed, and $\underset{\sim}{e} = \underset{\sim}{y} - \underset{\sim}{X}\underset{\sim}{\beta}$ is the vector of residuals. The k orthogonality conditions (6') define the usual least-squares vector $\underset{\sim}{b}$. The orthogonality condition (5') together with all but the pth orthogonality conditions from the set (6') define the pth reverse regression, which minimizes the residual sum-of-squares measured in the direction of variable x_p, that is $\underset{\sim}{e}'\underset{\sim}{e}/\beta_p^2$.

Each of these k+1 regressions satisfies the set of inequalities (5), (6) and (7). These estimates are formed by setting k of the k+1 variance parameters σ^2, d_1, d_2, ..., d_k equal to zero. In that event, the model can be written as $x_j = (\beta_o + \sum_{i \neq j} \beta_i x_i - y)/\beta_j + \varepsilon_j$, where ε_j has mean zero and variance d_j. In this form it is obvious that the coefficients are estimated by the j^{th} reverse regression.

A convenient way to compute these k + 1 regressions is to invert the moment matrix $\underset{\sim}{S}$:

$$\underset{\sim}{S}^{-1} = \begin{bmatrix} s_y^2 & \underset{\sim}{r}' \\ \underset{\sim}{r} & \underset{\sim}{N} \end{bmatrix}^{-1} = \begin{bmatrix} a_1 & a_2 & \cdots & a_{k+1} \\ \underset{\sim}{c}_1 & \underset{\sim}{c}_2 & \cdots & \underset{\sim}{c}_{k+1} \end{bmatrix}, \qquad (8)$$

where $\underset{\sim}{a}'$ is the first row of the inverse. Then the k + 1 regressions are $-\underset{\sim}{c}_j/a_j$, j = 1, 2, ..., k+1. This can be demonstrated by writing the system (5'), (6') as

$$\underset{\sim}{\delta} = \begin{bmatrix} \underset{\sim}{y}'\underset{\sim}{e} \\ \underset{\sim}{X}'\underset{\sim}{e} \end{bmatrix} = \underset{\sim}{S} \begin{bmatrix} 1 \\ -\underset{\sim}{\beta} \end{bmatrix}$$

where $\underset{\sim}{\delta}$ is a vector with element j equal to δ_j and all other elements equal to zero. Solving this system for $\underset{\sim}{\beta}$ yields $\underset{\sim}{\beta} = -\underset{\sim}{c}_j/a_j$.

We will now prove the following result.

<u>Theorem 1</u>: If the k + 1 regressions are all in the same orthant then the set of maximum likelihood estimates, that is, those values of $\underset{\sim}{\beta}$ satisfying (5), (6), and (7), is the convex hull of the k + 1 regressions.

Proof: The proof of the theorem is divided into two parts. First, we demonstrate that the set of maximum likelihood estimates is contained in the convex hull of the k + 1 regressions. Second, we demonstrate that every point in the convex hull satisfies (5), (6), and (7).

The first part of the proof follows from two propositions: (a) the region satisfying (5) and (6) is convex in each orthant; (b) at each of the k + 1 regressions, k of the k + 1 inequalities (5) and (6) are binding and the other inequality is satisfied but not binding. These propositions imply that if the k + 1 regressions are in the same orthant then the set of points in that orthant satisfying (5) and (6) is the convex hull of the k + 1 regressions. No points outside this orthant are feasible because the set of maximum likelihood points is connected. This follows from the continuity of the function $\underline{\beta}(\underline{D})=(\underline{N}-\underline{D})^{-1}\underline{r}$ for $\underline{N}-\underline{D}$ positive semi-definite.

The second part of the proof involves demonstrating that every point in the convex hull of the k + 1 regressions satisfies (7) (we have already demonstrated that (5) and (6) are satisfied). To establish this, it is convenient to suppress the normalization and to write the system of equations (1), (2), and (3) as

$$(9) \qquad (\underline{S}-\underline{\Delta})\underline{\gamma} = \underline{0}$$

$$\text{where } \underline{\Delta} = \begin{bmatrix} \sigma^2 & \underline{0}' \\ \underline{0} & \underline{D} \end{bmatrix} \quad \text{and } \underline{\gamma} = (1,-\underline{\beta}).$$

Without loss of generality, we assume that the variables are defined such that the k+1 regressions lie in the negative orthant, which implies, using (8), that $\underset{\sim}{S}^{-1}$ is strictly positive. Note that if $\underset{\sim}{\gamma}$ is strictly positive then $\underset{\sim}{\beta}$ must lie in the negative orthant. We establish the second part of the theorem by demonstrating that if $\underset{\sim}{\gamma}$ satisfies (9) and is strictly positive (and $\underset{\sim}{\Delta}$ satisfies (9) and is positive semi-definite) then $\underset{\sim}{S} - \underset{\sim}{\Delta}$ is positive semi-definite. This in turn implies that the submatrix $\underset{\sim}{N}-\underset{\sim}{D}$ of $\underset{\sim}{S}-\underset{\sim}{\Delta}$ is positive semi-definite, hence that (7) is satisfied.

The proof of this result exploits the following property of symmetric nonnegative matrices.[3]

<u>Lemma 1</u>: If λ is the eigenvalue of a symmetric nonnegative matrix corresponding to a strictly positive eigenvector, then λ is the maximum eigenvalue of the matrix.

<u>Proof</u>: The lemma follows directly from: (1) the maximum eigenvalue of a nonnegative (symmetric) matrix is positive, with corresponding eigenvector strictly positive (this is the Perron-Frobenius theorem--cf. Gantmacher (1959, pp. 65-74)); (2) a symmetric matrix has at most one positive eigenvector. (This follows from the fact that the eigenvectors of a symmetric matrix are orthogonal.).

Lemma 1 is used to establish the following two results which complete the proof.

<u>Lemma 2</u>: $\underset{\sim}{S}-\underset{\sim}{\Delta}$ is positive semi-definite if and only if all the roots of $\underset{\sim}{S}^{-1}\underset{\sim}{\Delta}$ are less than or equal to one.

Proof: Since S is a real symmetric nonsingular matrix and Δ is real and symmetric, there exists a nonsingular matrix P such that $P'SP=I$ and $P'\Delta P=\Lambda=\text{diag}\{\lambda_1, \lambda_2,\ldots,\lambda_n\}$(cf. Gantmacher (1960, p. 314)). Therefore $S-\Delta$ can be expressed as $(P')^{-1}(I-\Lambda)P^{-1}$, which implies that $S-\Delta$ is positive semi-definite if and only if $\lambda_i \leq 1$, $i=1,2,\ldots, n$. But the λ_i are the roots of $S^{-1}\Delta$, which can be seen by rewriting the characteristic equation $|S^{-1}\Delta-\lambda I|=0$ as $|P'|\,|S|\,|S^{-1}\Delta-\lambda I|\,|P|=|P'\Delta P-\lambda P'SP|=|\Lambda-\lambda I|=0$.

Lemma 3: If γ and Δ satisfy (9) and γ is strictly positive and Δ is positive semi-definite then all of the roots of $S^{-1}\Delta$ are less than or equal to one.

Proof: Rewriting (9) as $(I-S^{-1}\Delta)\gamma=0$, it follows that γ is an eigenvector of $S^{-1}\Delta$ with corresponding eigenvalue equal to one. Given that γ is strictly positive, Lemma 1 then implies that the largest root of the nonnegative matrix $S^{-1}\Delta$ is one. □

The region defined by (5) and (6) is not convex across all orthants because the direction of inequalities (6) shift from orthant to orthant. For $k > 2$, the restrictions implied by (7) are complicated non-linear functions. It is therefore difficult to characterize completely the set of maximum likelihood estimates when the $k + 1$ regressions are not in the same orthant. But a complete characterization is made unnecessary by the fact that the set is unbounded. The set can be sandwiched between a pair of parallel hyperplanes and one linear combination of parameters is therefore estimable. But except under unlikely circumstances, all other linear combinations are unbounded. These properties of the set of maximum likelihood estimates are summarized in Theorems 2 and 3.

Theorem 2: Values of $\underset{\sim}{r}'\underset{\sim}{\beta}$ with $\underset{\sim}{\beta}$ satisfying (5), (6) and (7) are restricted to

$$s_y^2 \geq \underset{\sim}{r}'\underset{\sim}{\beta} \geq \underset{\sim}{r}'\underset{\sim}{b} \quad .$$

Proof: The first inequality is just (5). The second inequality follows by using $\underset{\sim}{r} = (\underset{\sim}{N}-\underset{\sim}{D})\underset{\sim}{\beta}$ to obtain $\underset{\sim}{r}'\underset{\sim}{\beta} - \underset{\sim}{r}'\underset{\sim}{b} = \underset{\sim}{\beta}'(\underset{\sim}{N}-\underset{\sim}{D})\underset{\sim}{\beta} - \underset{\sim}{\beta}'(\underset{\sim}{N}-\underset{\sim}{D})\underset{\sim}{N}^{-1}(\underset{\sim}{N}-\underset{\sim}{D})\underset{\sim}{\beta} = \underset{\sim}{\beta}'(\underset{\sim}{D}-\underset{\sim}{D}\underset{\sim}{N}^{-1}\underset{\sim}{D})\underset{\sim}{\beta} = \underset{\sim}{\beta}'\underset{\sim}{D}(\underset{\sim}{D}^{-1} - \underset{\sim}{N}^{-1})\underset{\sim}{D}\underset{\sim}{\beta}$, which is nonnegative because $(\underset{\sim}{N}-\underset{\sim}{D})$, hence $(\underset{\sim}{D}^{-1} - \underset{\sim}{N}^{-1})$, is positive semi-definite.

Theorem 3: If the k+1 regressions are not all in the same orthant then the set $\{\underset{\sim}{\psi}'\underset{\sim}{\beta}|\underset{\sim}{\beta}$ satisfying (5), (6), and (7)$\}$ is the set of real numbers for almost all vectors of constants $\underset{\sim}{\psi}$.

Proof: Theorem 3 is established by demonstrating that a subset of the set $\{\underset{\sim}{\psi}'\underset{\sim}{\beta}|\underset{\sim}{\beta}$ satisfying (5), (6), and (7)$\}$ is unbounded. Specifically, we consider the set $\{\underset{\sim}{\psi}'\underset{\sim}{\beta}|\underset{\sim}{\beta}$ satisfying (5), (6), and (7) and $d_i = 0$ for all but two suitably chosen values of i, $i = 1, 2, \ldots, k\}$. We prove that this set is unbounded by first demonstrating that requiring all but two of the d_i to equal zero effectively reduces the k + 1 variable problem to a three-variable problem. In the next section we establish that the three-variable problem is unbounded when the (three) regressions are not in the same orthant.

Without loss of generality, suppose that the two suitably chosen non-zero d_i are d_1 and d_2. Then $d_i = 0$ for $i = 3, 4, \ldots, k$ and equation (2) can be rewritten as

$$\begin{bmatrix} \underset{\sim}{\beta}_1 \\ \underset{\sim}{\beta}_2 \end{bmatrix} = \begin{bmatrix} \underset{\sim}{N}_{11} - \underset{\sim}{D}_{11} & \underset{\sim}{N}_{12} \\ \underset{\sim}{N}_{21} & \underset{\sim}{N}_{22} \end{bmatrix}^{-1} \begin{bmatrix} \underset{\sim}{r}_1 \\ \underset{\sim}{r}_2 \end{bmatrix} .$$

where $\underset{\sim}{D}_{11} = \text{diag}\{d_1, d_2\}$ and $\underset{\sim}{N}$, $\underset{\sim}{\beta}$ and $\underset{\sim}{r}$ (and subsequently $\underset{\sim}{\chi}$, $\underset{\sim}{x}$, and $\underset{\sim}{\psi}$) are partitioned conformably. Solving this system yields

$$\underset{\sim}{\beta}_2 = \underset{\sim}{N}_{22}^{-1}(\underset{\sim}{r}_2 - \underset{\sim}{N}_{21}\underset{\sim}{\beta}_1) \qquad (10)$$

and

$$r_* = \beta_1' \Sigma_*, \tag{2'}$$

where $r_* = r_1 - N_{12}N_{22}^{-1} r_2$ is the sample covariance of y and x_1 given x_2 and $\Sigma_* = N_{11} - D_{11} - N_{12} N_{22}^{-1} N_{21} = \Sigma_{11} - N_{12} N_{22}^{-1} N_{21}$ (using $\Sigma_{11} = N_{11} - D_{11}$ from (3)) is the sample covariance of χ_1 given x_2 (or equivalently χ_2). In addition, (3) implies $N_{11} = \Sigma_{11} + D_{11}$, which, after subtracting $N_{12} N_{22}^{-1} N_{21}$ from both sides, yields

$$N_* = \Sigma_* + D_*, \tag{3'}$$

where $N_* = N_{11} - N_{12} N_{22}^{-1} N_{21}$ is the covariance of x_1 given x_2 and $D_* = D_{11}$. Finally, using (10) to substitute for β_2, (1) can be rewritten as

$$s_{y*}^2 = \sigma^2 + \beta_1' \Sigma_* \beta_1, \tag{1'}$$

where $s_{y*}^2 = s_y^2 - r_2' N_{22}^{-1} r_2$ is the sample variance of y given x_2.

Equations (1'), (2'), and (3') have the same structure as (1), (2), and (3), except that equations (1'), (2'), and (3') pertain only to β_1 and all variances and covariances are conditioned on x_2. This demonstrates that the additional constraints $d_i = 0$ for $i = 3, 4, \ldots, k$ effectively reduce the problem to a three-variable model. If the set $\{\psi'\beta | \beta$ satisfying (5), (6), (7), and $d_i = 0$ for $i = 3, 4, \ldots, k\}$ is unbounded when the regressions are not in the same orthant (as is demonstrated in the next section for a suitable choice of x_1) then the corresponding set of values of $\psi'\beta = \psi_1'\beta_1 + \psi_2'\beta_2 = \psi_1'\beta_1 + \psi_2' N_{22}^{-1} (r_2 - N_{21}\beta_1)$ is unbounded provided $\psi_1' - \psi_2' N_{22}^{-1} N_{21} \neq 0$ and provided $\psi \neq r$ (Theorem 2).

All that is left to prove the result is a careful choice of $\underset{\sim}{x}_1$ so that the three regressions for the reduced problem are not in the same orthant. To do this, simply order and choose signs of the variables such that the inverse of the moment matrix (8) has the form

$$\begin{bmatrix} + & - & - & \cdot \\ - & + & - & \cdot \\ - & - & + & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{bmatrix}$$

The diagonal elements have to be positive, and the matrix has to be symmetric. The variables are signed so that the direct regression is in the positive orthant, that is, elements (1,2) and (1,3) are negative. The first reverse regression is chosen so that it is not in the positive orthant, and indices are chosen such that its first component is positive and its second is negative; that is element (3,2) is negative. $\square$

Theorems 1 and 3 can be easily amended if some of the variables are known to be measured correctly. If only $m < k$ of the explanatory variables are measured with error then only $m + 1$ regressions are computed and only the coefficients on the m mismeasured variables are checked for sign changes. This follows from the fact that equations (1), (2), and (3) can be written in terms of the $m \times 1$ subvector of β corresponding to the mis-measured variables, with the sample moments controlling for the error-free explanatory variables.

3. The Three-variable Case

In this section, we characterize completely the three-variable case. For convenience we can normalize such that

$$s_y^2 = 1$$

$$N = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

$$r' = [r_1 \quad r_2]$$

The least-squares estimates then are

$$[b_1, b_2] = [r_1 - \rho r_2,\ r_2 - \rho r_1]/(1-\rho^2).$$

The region of maximum likelihood estimates is described by the inequalities

$$1 - r_1\beta_1 - r_2\beta_2 \geq 0 \tag{11}$$

$$(\beta_1 + \rho\beta_2 - r_1)/\beta_1 \geq 0 \tag{12}$$

$$(\rho\beta_1 + \beta_2 - r_2)/\beta_2 \geq 0 \tag{13}$$

$$(r_1 - \rho\beta_2)/\beta_1 \geq 0 \tag{14}$$

$$(r_1 r_2 - \rho(r_1\beta_1 + r_2\beta_2))/\beta_1\beta_2 \geq 0 \tag{15}$$

Inequality (11) is implied by (5), inequalities (12) and (13) are implied by (6), and inequalities (14) and (15) are implied by (7).

The reverse regression estimates formed by letting (11) and (12) be equalities are

$$(\hat{\beta}_1,\hat{\beta}_2) = ((r_1r_2-\rho),\ (1-r_1^2))/b_2(1-\rho^2)\ .$$

The reverse regression estimates formed by letting (11) and (13) be equalities are

$$(\hat{\beta}_1,\hat{\beta}_2) = (1-r_2^2,(r_1r_2-\rho))/b_1(1-\rho^2).$$

The conditions for these vectors and (b_1,b_2) to be all in the same quadrant are $\mathrm{sgn}(b_1) = \mathrm{sgn}((r_1r_2-\rho)/b_2)$ and $\mathrm{sgn}(b_2) = \mathrm{sgn}((r_1r_2-\rho)/b_1)$. But these are equivalent, and can be written for emphasis as a theorem:

Theorem 4: The three regressions are in the same quadrant if and only if

$$\mathrm{sgn}(b_1)\mathrm{sgn}(b_2) = \mathrm{sgn}(r_1r_2-\rho). \tag{16}$$

A weaker sufficient condition is the following.

Theorem 5: Condition (16) holds if $\mathrm{sgn}(b_1)\mathrm{sgn}(b_2) = -\mathrm{sgn}\ (\rho)$.

Proof: It is enough to show this for the case $b_1 > 0$, $b_2 > 0$, $\rho < 0$, since the other cases are implied by a change in sign of one or both variables. The condition $b_2 > 0$ implies $r_2 - \rho r_1 > 0$. We can write $r_1r_2 - \rho(\underset{\sim}{r}'\underset{\sim}{b}) = (r_2-\rho r_1)b_1$ and from $b_1 > 0$ conclude $r_1r_2 - \rho(\underset{\sim}{r}'\underset{\sim}{b}) > 0$. The condition $R^2 < 1$ implies $\underset{\sim}{r}'\underset{\sim}{b} < 1$. Using these inequalities and $\rho < 0$, we derive $r_1r_2\rho^{-1} < \underset{\sim}{r}'\underset{\sim}{b} < 1$, and $r_1r_2 > \rho$. □

Using these two results and the inequalities (11) to (15) we may form three figures which represent the three distinct cases which can occur. Figures 1 and 2 illustrate Theorems 4 and 5 in which the three regressions are in the same quadrant. Figure 3 is like Figure 1 except that condition (11) has been adjusted so that the three regressions are not in the same quadrant. Note that no adjustment to (11) in Figure 2 can cause the three regressions to be in different quadrants. Also note that (14) is drawn only in Figure 3. This is generally implied by the other inequalities. A proper proof of the unboundedness of the region depicted in Figure 3 requires the tedious checking of the inequalities (11) to (15) along the line $\underset{\sim}{r}'\underset{\sim}{\beta} = c$ for $\underset{\sim}{r}'\underset{\sim}{b} < c < 1$. This is left to the reader.

4. Prior Restrictions

In most applications in economics, the possibility of measurement error in the regressors is ignored. This is not because it is generally believed that economic data are measured without error. Rather, it is a consequence of the fact that the procedures available to deal with measurement error in the regressors often require the use of prior information which in some sense is more heroic than the assumption of correctly measured regressors. Left to choose between the lesser of evils, the typical researcher generally opts for the simpler, more traditional approach.

If interest focuses only on the signs of the true regression coefficients, the traditional approach may not be that bad. In Section 2 we saw that if the $k + 1$ regressions are in the same orthant, the maximum likelihood region is entirely contained in the same orthant as $\underset{\sim}{b}$. But in many instances it is to be expected that the $k + 1$ regressions will not all be in the same orthant, in which case Theorem 3 implies that the traditional data summary conveys no information about the signs of the true coefficients. In other instances, the researcher may be interested in the magnitudes as well as the signs of the true coefficients.

The situation is not as bleak as it appears when it is recognized that in the average application the researcher may possess considerable prior information concerning the seriousness of the measurement errors in the observed regressors. In particular, we assume that the researcher is able to answer the following two questions:

(a) If there were no measurement errors in the observed x's, how large could be the squared multiple correlation of y with the x's?

(b) Among the k regressors, how small could the squared correlation be between an observed regressor and its true counterpart?

The material in Section 2 assumes that the answer to (a) is one and that the answer to (b) is zero. These are unreasonable answers. No one would really expect to see the R^2 go to one as the measurements become more accurate, and no one thinks that the variance in the measured x's is entirely due to measurement error. But inequality (5) allows the R^2 to be one, and the inequalities (7) allow the measurement error variances to be as large as the sample variances. If more restrictive but still sensible assumptions are used, the set of maximum likelihood estimates may shrink substantially.

The true R^2, $\underline{\beta}'\underline{\Sigma}\underline{\beta}/(\sigma^2 + \underline{\beta}'\underline{\Sigma}\underline{\beta})$, is estimated by $\underline{r}'\underline{\beta}/s_y^2$, and the restriction that this estimate is less than R^{*2} implies the inequality

$$R^{*2}s_y^2 - \underline{r}'\underline{\beta} > 0 \tag{17}$$

This is the same as inequality (5) if $R^{*2} = 1$. The squared correlation between χ_i and its measurement x_i, $\Sigma_{ii}/(\Sigma_{ii}+d_i)$, is estimated by $(N_{ii}-d_i)/N_{ii} = (N_{ii}\beta_i - \underline{N}_i'(\underline{\beta}-\underline{b}))/N_{ii}\beta_i$ and the restriction that this estimate of the squared correlation is between ρ_*^2 and ρ^{*2} implies the inequalities

$$\rho^{*2} \geq (N_{ii}\beta_i - \underline{N}_i'(\beta - b))/N_{ii}\beta_i \geq \rho_*^2 \quad i = 1, \ldots, k \tag{18}$$

Inequalities (6) and (7) imply (18) with $\rho^{*2} = 1$ and $\rho_*^2 = 0$.

Inequality (17) is easy to work with since it involves only a trivial adjustment to (5). Similarly, if ρ^{*2} is not one, the inequalities (6) are trivially altered. But if ρ_*^2 is not zero, k additional inequalities are added to the problem and the region of maximum likelihood estimates can become considerably more complex. A complete way to report results would be to characterize fully the maximum likelihood region for a set of values for ρ_*^2, ρ^{*2}, and R^{*2}. This is likely to consume more journal space than is justified and it is computationally expensive. We propose instead two alternatives.

The first sets $\rho_*^2 = 0$ and $\rho^{*2} = 1$ and explores the nature of the maximum likelihood region as a function of R^{*2}. We demonstrate that there exists a maximum value of R^{*2}, denoted by R_m^{*2}, for which the maximum likelihood region remains bounded and, incidentally, entirely contained in the orthant containing $\underset{\sim}{b}$. The vertices of the maximum likelihood region are described as a function of R^{*2} for values of $R^{*2} \leq R_m^{*2}$. The second alternative sets $R^{*2} = 1$ and encloses the maximum likelihood region within an ellipsoid. A condition which ensures that the ellipsoid is bounded is identified.

If $\rho_*^2 = 0$ and $\rho^{*2} = 1$, the problem is straightforwardly altered by replacing s_y^2 by $R^{*2}s_y^2$, as in (17). Consider then the k + 1 adjusted regressions which can be computed by inverting the adjusted moment matrix

$$\begin{bmatrix} R^{*2}s_y^2 & \underset{\sim}{r}' \\ \underset{\sim}{r} & \underset{\sim}{N} \end{bmatrix}^{-1} = \begin{bmatrix} \cdot & -(R^{*2}s_y^2)^{-1}\,\underset{\sim}{r}'\underset{\sim}{C}^{-1} \\ \cdot & \underset{\sim}{C}^{-1} \end{bmatrix}, \tag{19}$$

where $\underset{\sim}{C} = \underset{\sim}{N} - \underset{\sim}{r}(R^{*2}s_y^2)^{-1}\underset{\sim}{r}'$. The adjusted k + 1 regressions are formed by dividing this matrix by the negative of the first row, as described above the proof of Theorem 1. Following the proofs of Theorems 1 and 3, it is then possible to establish the following theorem.

Theorem 6: If the k + 1 adjusted regressions are in the same orthant then the set of maximum likelihood estimates, that is, those values of $\underset{\sim}{\beta}$ satisfying (17), (6) and (7), is the convex hull of the k + 1 adjusted regressions. If the k + 1 adjusted regressions are not in the same orthant and if $\underset{\sim}{\psi}$ is a vector of constants, then the set of maximum likelihood estimates for $\underset{\sim}{\psi}'\underset{\sim}{\beta}$ is the set of real numbers for almost all vectors of constants $\underset{\sim}{\psi}$.

The k + 1 adjusted regressions can be computed directly from (19). The reader may verify that the first regression is just the direct regression $b = N^{-1}r$, independent of R^{*2}. The other k adjusted regressions are found by inverting C:

$$C^{-1} = N^{-1} - N^{-1}r(r'N^{-1}r-R^{*2}s_y^2)^{-1}r'N^{-1}$$

$$= ((r'N^{-1}r-R^{*2}s_y^2)N^{-1}-bb')/(r'N^{-1}r-R^{*2}s_y^2)$$

and computing

$$r'C^{-1} = b' - r'N^{-1}r(r'N^{-1}r-R^{*2}s_y^2)^{-1}b'$$

$$= - R^{*2}s_y^2b/(r'N^{-1}r-R^{*2}s_y^2)$$

Thus the ith adjusted regression, formed by dividing a column of the inverse moment matrix by the negative of its first element, is

$$\hat{\beta}_{(i)}(R^{*2}) = - ((r'N^{-1}r-R^{*2}s_y^2)S_{(i)}-bb_i)/b_i \tag{20}$$

where $S_{(i)}$ is the ith column of N^{-1}. If $R^{*2} = 1$, this is just the ith reverse regression

$$\hat{\beta}_{(i)} = b + (s_y^2-r'N^{-1}r)S_{(i)}/b_i \quad .$$

We can, therefore, write the ith adjusted regression as a weighted average of the ith reverse regression and the direct regression:

$$\hat{\beta}_{(i)}(R^{*2}) = ((1-R^{*2})\, s_y^2b + (R^{*2}s_y^2-r'N^{-1}r)\hat{\beta}_{(i)})/(s_y^2-r'N^{-1}r) \tag{21}$$

It may be noted that the lowest value of R^{*2} consistent with (19) being positive semi-definite is the squared multiple correlation between y and the measured x's: $r'N^{-1}r/s_y^2$, in which case $\hat{\beta}_{(i)}(R^{*2})$ is just b. Using either

(20) or (21) it is easy to trace out the linear path taken by the ith adjusted regression as R^{*2} varies from $\underset{\sim}{r}'\underset{\sim}{N}^{-1}\underset{\sim}{r}/s_y^2$ to one.

Equation (20) or (21) can also be used to solve for R^{*2}_m, the maximum value of R^{*2} consistent with all adjusted regressions in the same orthant. Referring to equation (21) and assuming that b_j and $\hat{\beta}_{ij}$ are opposite in sign, we can solve for the values of R^{*2} at which $\hat{\beta}_{ij}(R^{*2})$ and b_j assume the same sign:

$$\begin{aligned} R^{*2} &\leq (s_y^2 b_j - \underset{\sim}{r}'\underset{\sim}{N}^{-1}\underset{\sim}{r}\hat{\beta}_{ij})/s_y^2(b_j-\hat{\beta}_{ij}) \\ &= (\underset{\sim}{r}'\underset{\sim}{N}^{-1}\underset{\sim}{r}(b_j-\hat{\beta}_{ij})+b_j(s_y^2-\underset{\sim}{r}'\underset{\sim}{N}^{-1}\underset{\sim}{r}))/s_y^2(b_j-\hat{\beta}_{ij}) \\ &= \hat{R}^2 + (1-\hat{R}^2)/(1-(\hat{\beta}_{ij}/b_j)) \quad , \end{aligned}$$

where $\hat{R}^2$ is the squared sample correlation coefficient $\underset{\sim}{r}'\underset{\sim}{N}^{-1}\underset{\sim}{r}/s_y^2$. This value is minimized across all i and j when the ratio $\hat{\beta}_{ij}/b_j$ is as large a negative number as possible. Therefore

$$R^{*2}_m = \hat{R}^2 + (1-\hat{R}^2) \min_{i,j} \left((1-(\hat{\beta}_{ij}/b_j))^{-1}\right), \qquad (22)$$

where the indices i,j select only those estimates opposite in sign, $\hat{\beta}_{ij}/b_j < 0$.

Another way to report this number is to compute the proportion of the gap between $R^2 = \hat{R}^2$ and $R^2 = 1$ that can be attributed to measurement error without causing the k + 1 regressions to be in more than one orthant. Denoting this by g,

$$g = \begin{cases} 1 & \text{if the } k+1 \text{ regressions are in the same orthant} \\ (R^{*2}_m-\hat{R}^2)/(1-\hat{R}^2) & \text{otherwise} \end{cases} \quad .$$

The statistic g is a convenient summary statistic to indicate the sensitivity of conventional inferences to measurement error in the regressors. A value of g equal to one indicates that the k + 1 unadjusted regressions are in the same orthant. In this instance, no prior restrictions at all are needed to ensure that the maximum likelihood region is bounded and entirely contained in the orthant containing $\underset{\sim}{b}$. Alternatively, a value of g equal to zero indicates that allowing even the smallest amount of measurement error in any of the regressors will cause the maximum likelihood region to be unbounded. In this latter instance, conventional data summaries which report only $\underset{\sim}{b}$ may be a particularly poor summary of the data unless the possibility of even small measurement errors in the regressors can be ruled out a priori.

The above discussion is based on the case where $\rho_*^2 = 0$ and $\rho^{*2} = 1$. Provided $\rho_*^2 = 0$, the problem is straightforwardly altered to accommodate a value of $\rho^{*2} \neq 1$. As above, all that is required is to alter the moments N_{ii} to $N_{ii}\rho^{*2}$. But it is more likely that a researcher would want $\rho_*^2 \neq 0$ than $\rho^{*2} \neq 1$. We propose to deal with this more complicated case by enclosing the maximum likelihood region by an ellipsoid. Using (2) and (3) we can write the estimate of $\underset{\sim}{\beta}$ as a function of $\underset{\sim}{D}$ as

$$\underset{\sim}{\beta} = (\underset{\sim}{N}-\underset{\sim}{D})^{-1}\underset{\sim}{r} \tag{23}$$

The inequalities (18) constrain the diagonal elements of $\underset{\sim}{D}$ to lie between $N_{ii}(1-\rho^{*2})$ and $N_{ii}(1-\rho_*^2)$. A weaker restriction allows $\underset{\sim}{D}$ to be non-diagonal but constrains $\underset{\sim}{D}$ to $(1-\rho^{*2})\operatorname{diag}\{N_{11}, N_{22}, \ldots, N_{kk}\} \equiv \underset{\sim}{D}_* \leq \underset{\sim}{D} \leq \underset{\sim}{D}^* \equiv (1-\rho_*^2)\operatorname{diag}\{N_{11}, N_{22}, \ldots, N_{kk}\}$, where $\underset{\sim}{A} \leq \underset{\sim}{B}$ means $\underset{\sim}{B} - \underset{\sim}{A}$ is positive semi-definite. The estimator $\underset{\sim}{\beta}$ can be written as $((\underset{\sim}{N}-\underset{\sim}{D}^*)+(\underset{\sim}{D}^*-\underset{\sim}{D}))^{-1}\underset{\sim}{r}$ with $\underset{\sim}{0} \leq \underset{\sim}{D}^* - \underset{\sim}{D} \leq \underset{\sim}{D}^* - \underset{\sim}{D}_*$. Written this way the estimator appears to be a posterior mean with a prior located at the origin with a prior variance matrix $\underset{\sim}{V}$ bounded from below, $(\underset{\sim}{D}^*-\underset{\sim}{D}_*)^{-1} \leq \underset{\sim}{V}$. Theorem 2 in Leamer (1981) then implies the following.

Theorem 7: Any β satisfying (23) with $D_* \leq D \leq D^* \leq N$ lies in the ellipsoid

$$(\beta-f)'H(\beta-f) \leq c \tag{24}$$

where

$$H = (N-D^*)(D^*-D_*)^{-1}(N-D^*) + N - D^*$$
$$f = (N-D_*)^{-1}(r+(D^*-D_*)(N-D^*)^{-1}r/2)$$
$$= (N-D_*)^{-1}(N-(D^*+D_*)/2)(N-D^*)^{-1}r$$
$$c = r'(N-D^*)^{-1}(D^*-D_*)(N-D_*)^{-1}r/4 \quad .$$

This ellipsoidal region is much easier to work with than the true maximum likelihood region. If interest centers on some linear combination of parameters, say $\psi'\beta$, where ψ is a vector of constants, then the extreme estimates over the ellipsoid (24) are $\psi'f \pm (\psi'H^{-1}\psi c)^{\frac{1}{2}}$. The most serious problem with this interval is that it ignores the constraint (17). The adjusted k+1 regressions when they are in the same orthant do make use of the constraint (17), but set ρ_*^2 to zero. The intersection of these two regions makes use of both sources of information and may give a fairly accurate picture of the true maximum likelihood region.

The maximum likelihood region defined by (7), (17) and (18) is wholly contained by ellipsoid (24). An important condition of Theorem 7 is that $N \geq D^*$. In the event $N > D^*$, ellipsoid (24) is finite and by implication the maximum likelihood region is bounded. This implies:

Corollary. The maximum likelihood region is bounded if $1 - \lambda_1 \leq \rho_*^2$ where λ_1 is the smallest eigenvalue of the correlation matrix $\underset{\sim}{N}$, normalized to $N_{ii} = 1, i = 1, \ldots, k$.

When k=2, the smaller eigenvalue of the correlation matrix equals $1 - |N_{12}|$, where N_{12} is the correlation between x_1 and x_2. Thus, a condition to get a bounded region is that $|N_{12}| < \rho_*^2$, the squared correlation between the true X and its measurement x must be known to exceed the absolute value of the correlation between the measured x's. For this reason, the absolute correlation N_{12} and its generalization $1 - \lambda_1$ serve as useful indicators of one aspect of the collinearity problem: if the measured x's are highly correlated it is difficult to rule out the possibility that the true x's are perfectly correlated.

5. An Example

We present in this section an example. The problem we address is the estimation of the Heckscher-Ohlin-Vanek (1968) model of trade which expresses net exports of a commodity as a linear function of the resource endowments of the country. We analyze 1972 net exports of a machinery aggregate composed of SITC classes 71 (non-electrical machinery), 72 (electrical machinery), 73 (transport equipment) and 86 (professional goods). The 1972 net exports of machinery in thousands of dollars of 47 countries are related to the country's land, labor and capital. Land is the total land area of the country in thousands of hectares, taken from the FAO, Production Yearbook. Labor is the number of economically active individuals in thousands, taken from the ILO, Labor Force Projections 1965-1985. Capital is accumulated domestic investment flows in 1972 dollars assuming a fifteen year average asset life, derived from World Bank, World Tables 1976.

The regression of net exports of machinery on land, labor and capital yields the following equation with t values in parentheses:

$$Y = 84981 \underset{(4.0)}{-6.18}\ \text{Land} + \underset{(.6)}{6.20}\ \text{Labor} + \underset{(6.4)}{12.35}\ \text{Capital}, \quad R^2 = .52.$$

This regression suggests, as might be expected, that countries with relatively large amounts of labor and capital have a comparative advantage in machinery production, and countries with relatively large amounts of land have comparative advantages in other products. However, the variables land, labor, and capital are doubtlessly measured with error and we need to know if these inferences are sensitive to assumptions about the error variances.

This can be probed by computing the three reverse regressions. They are reported along with the direct regression in Table 1. The first column contains the estimated direct regression of net exports on the three explanatory variables. The next three columns contain the three reverse regressions. If minimization of the sum-of-squares is done in the direction of the land variable, the estimated coefficients have the same signs as the direct regression. But the other two reverse regressions have estimated coefficients with different signs. We are accordingly in the situation in which the set of maximum likelihood estimates is unbounded, and these data are useless unless we make use of additional prior information.

First, note that if we were willing to assume that capital is measured correctly, then the set of maximum likelihood estimates would be bounded. This can be seen from Table 1. If the capital column and row are removed (and the intercept row and column are also removed since the constant regressor is measured correctly), the signs of the remaining estimates are the same. However, the capital variable is perhaps the explanatory variable most likely to be measured with error. Labor likewise can be expected to be measured with error because of differences in vacations, work week, overtime, effort, etc. from country to country.

Theorems 6 and 7 indicate that it isn't necessary to assume that some of the explanatory variables are measured correctly in order to bound the set of maximum likelihood estimates. Theorem 6 indicates that if one is willing to select a sufficiently small value for R^{*2}, the maximum R^2 that would result if the measurement error in the explanatory variables were removed, then the set of maximum likelihood estimates will be restricted to the orthant containing the direct regression. The R^2 based on the measured explanatory variables is .52. Using (22) and the estimates reported in

Table 1, R_m^{*2}, the maximum value of R^{*2} for which the set of maximum likelihood estimates is restricted to the orthant containing the the direct regression, is .698. Therefore if one is willing to select a value for R^{*2} less than .698 then the set of maximum likelihood estimates will be bounded. Alternatively stated, g, the fraction of the gap between $R^2 = .52$ and $R^2 = 1$ that can be attributed to measurement error without causing the set of maximum likelihood estimates to be unbounded, is .3717.

Bounds for the estimates as functions (20) of R^{*2} are graphed in Figure 4. If $R^{*2} = .52$ there is no room for measurement error, and these bounds narrow to a point at the left of these figures. Both the labor and capital bounds overlap the origin at $R^{*2} = .698$. At this point some of the adjusted k+1 regressions exit the orthant of the direct regression and the region becomes unbounded. The lines that form the boundaries of these sets of estimates connect the direct regression on the left with the extreme estimates from the k+1 regressions on the right.

A plausible value for R^{*2} is .6. It strikes us as excessively optimistic to expect the R^2 to increase from .52 above .6 merely by the elimination of the measurement error in land, labor and capital. For R^{*2} equal to .6, Figure 4 indicates that the land coefficient and the capital coefficient are tightly bounded, but the labor coefficient bound is quite wide.

An alternative to assuming that R^{*2} is less than or equal to R_m^{*2} is to select a value for ρ_*^2, the minimum squared correlation between the true variables and their measurements. The corollary to Theorem 7 indicates that if $\rho_*^2 > 1-\lambda_1 = .565$, where λ_1 is the minimum eigenvalue of the correlation matrix of the explanatory variables, then the set of estimates is bounded. Bounds for the estimates as a function of ρ_*^2 using Theorem 7 are reported in Table 2. At the right of this table are the direct regression estimates. The bounds get larger as the value of ρ_*^2 gets smaller.

A plausible value for ρ_*^2 is .8. It strikes us as excessively pessimistic to think that the squared correlations between the measurements and the true variables are less than .8. If ρ_*^2 is set equal to .8, Table 2 indicates that the intervals of estimates for the land and capital coefficients are fairly tight, but the interval for the labor coefficient is wide and overlaps the origin.

Thus, although the k+1 regressions are not in the same orthant, sensible prior restrictions on the error variances allow us to make inferences from these data. The general conclusion is that inferences about the land and capital coefficients are reasonably insensitive, but inferences about the labor coefficient are very sensitive, to measurement error issues. This, incidentally, conforms to the t-values in the usual direct regression.

6. Concluding Comments

It is conventional practice to act as if it is possible to measure without error the theoretical constructs of economic models. This assumption is generally imposed for convenience rather than because it is widely shared. Our results suggest that it is possible to loosen this rigid assumption without inordinate sacrifices in tractability. We propose two statistics to supplement the conventional regression summary. One statistic indicates the proportion of the difference between $R^2 = \hat{R}^2$ and $R^2 = 1$ that can be attributed to measurement error without causing the set of maximum likelihood estimates to become unbounded. The other statistic indicates the minimum squared correlation between true variables and their measurements required to bound the set of maximum likelihood estimates.

These two statistics are easy to compute, and, as we demonstrate in in Section 5, they are simple to use. This should make it easier for researchers to deal seriously with the possibility of errors of measurement.

Endnotes

1. An analysis that also makes use of information of this form is Leonard (1979).

2. The assumption that y and $\underset{\sim}{x}$ are normally distributed enables us to interpret the set of values of $\underset{\sim}{\beta}$ satisfying (5), (6), and (7) as the set of maximum likelihood estimates. However, this set is of interest independent of the normality assumption. Note that in the population, (1), (2), and (3) must be satisified for any distributions of y and $\underset{\sim}{x}$. Consequently, for any distributions of y and $\underset{\sim}{x}$, if the sample size is large enough, the set of estimates of $\underset{\sim}{\beta}$ satisfying (5), (6), and (7) almost surely contains the true value of $\underset{\sim}{\beta}$.

3. A version of the following proof (in Lemmas 1-3) is in the original manuscript by Klepper (1980). An error in an alternative proof was kindly pointed out to us by Rudolf Kalman. Kalman (1981) uses the Perron-Frobenius Theorem cited in Lemma 1 in much the same way as we do to establish results similar to Lemmas 1-3.

References

Dhondt, A. (1960), "Sur une généralisation d'un theorème de R. Frisch en analyse de confluence," Cahiers du Centre d'Etudes de Recherche Operationnelle, Vol. 2, No. 1.

Frisch, R. (1934), Statistical Confluence Analysis by Means of Complete Regression Systems, University Institute of Economics, Oslo, Norway.

Gantmacher, F.R. (1959), Applications of the Theory of Matrices, Interscience Publishers Inc., New York.

_______________ (1960), The Theory of Matrices, Volume One, Chelsea Publishing Company, New York.

Goldberger, A.A. (1972), "Structural Equation Methods in the Social Sciences," Econometrica, 40, 979-1001.

Kalman, R.G. (1981), "System Identification From Noisy Data," presented at the International Symposium of Dynamical Systems, Gainesville, FL.

Klepper, S. (1980), "Summarizing the Data for the Classical Normal Errors-in-Variables Model," mimeo.

Koopmans, T. (1937), Linear, Regression Analysis of Economic Time Series, Netherlands Econometric Institute, Harrlem-de Erwen F. Bohn N. V.

Leamer, E.E. (1981), "Sets of Posterior Means with Bounded Variance Priors," Econometrica, 50, 725-736.

_______________ (1978), Specification Searches: Ad Hoc Inference with Non-experimental Data, Wiley, New York.

Leonard, H. (1979), Inference in Partially Identified Models, unpublished doctoral dissertation, Harvard University.

Malinvaud, E. (1970), Statistical Methods of Econometrics, North-Holland, Amsterdam.

Patefield, W.M. (1981), "Multivariate Linear Relationships: Maximum Likelihood Estimation and Regression Bounds," J. Royal Statistical Society B, Vol. 43, 342-52.

Reiersol, O. (1945), "Confluence Analysis by Means of Instrumental Sets of Variables," Arkiv for Mathematik, Astronomi och Fysik, Vol. 32, 1-119.

Vanek, J. (1968), "The Factor Proportions Theory: The N-Factor Case," Kyklos, 21, 749-54.

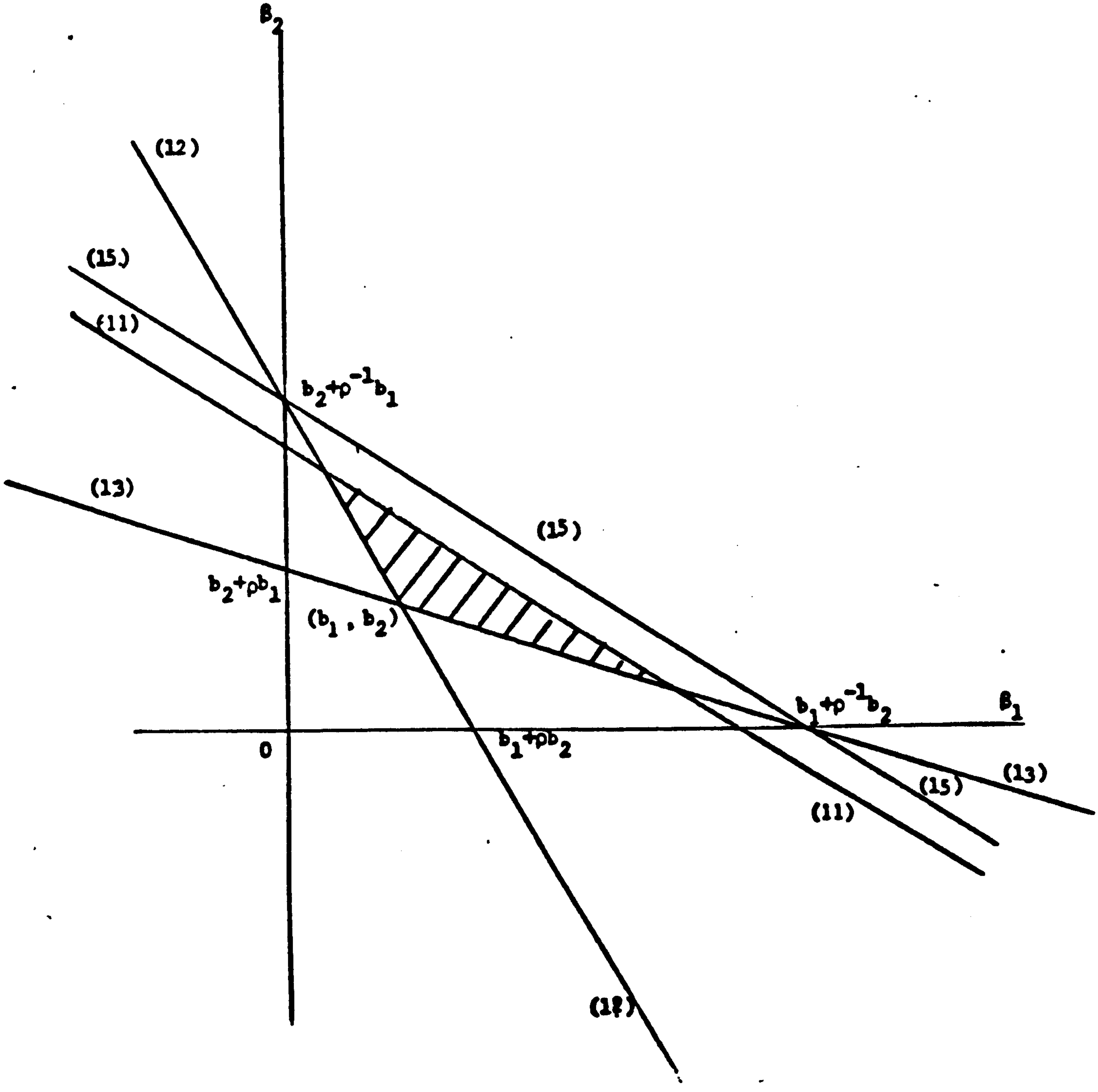

Figure 1

The Set of Maximum Likelihood Estimates

$b_1 > 0,\ b_2 > 0,\ \rho > 0,\ r_1 r_2 - \rho > 0$

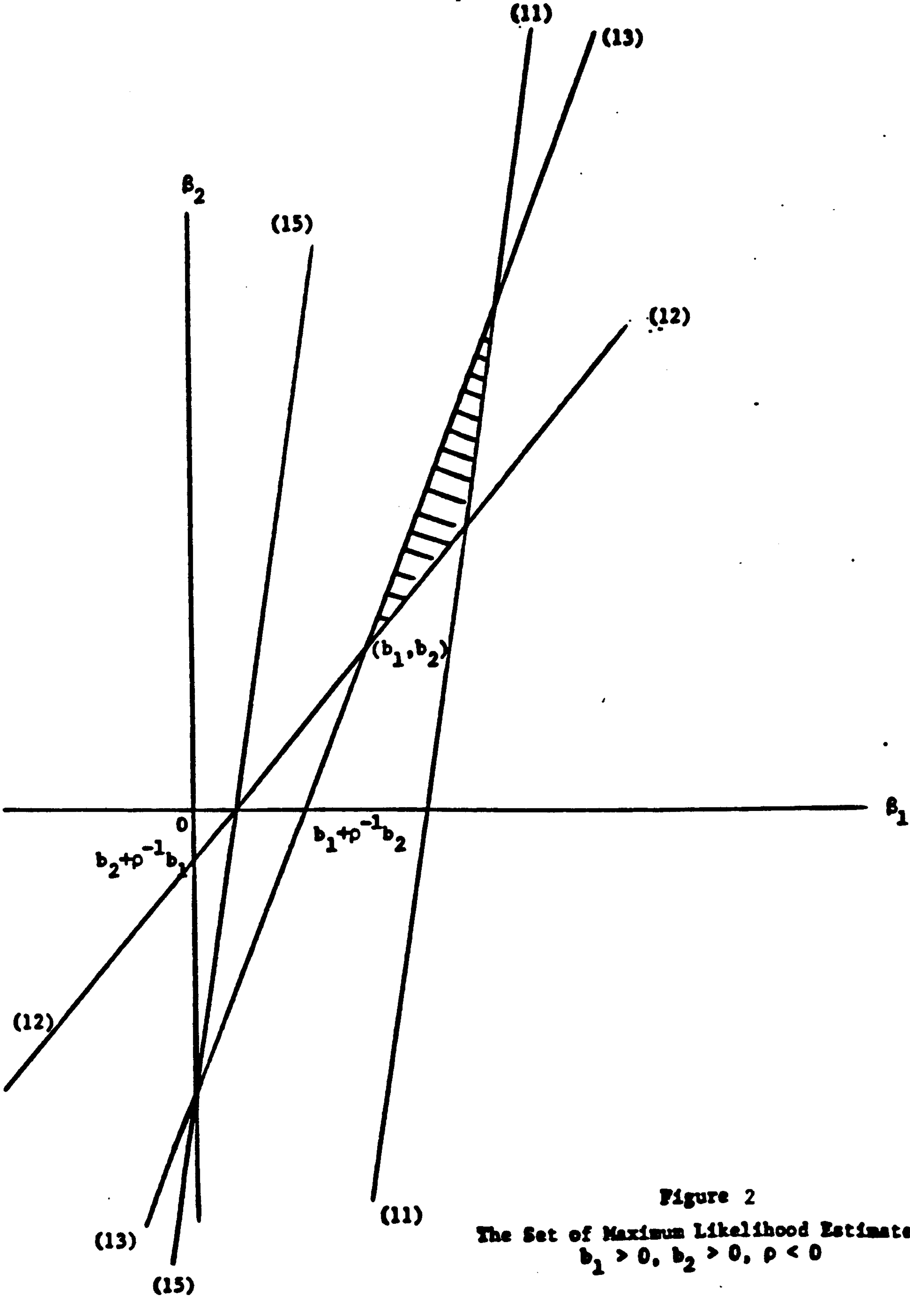

Figure 2

The Set of Maximum Likelihood Estimates
$b_1 > 0$, $b_2 > 0$, $\rho < 0$

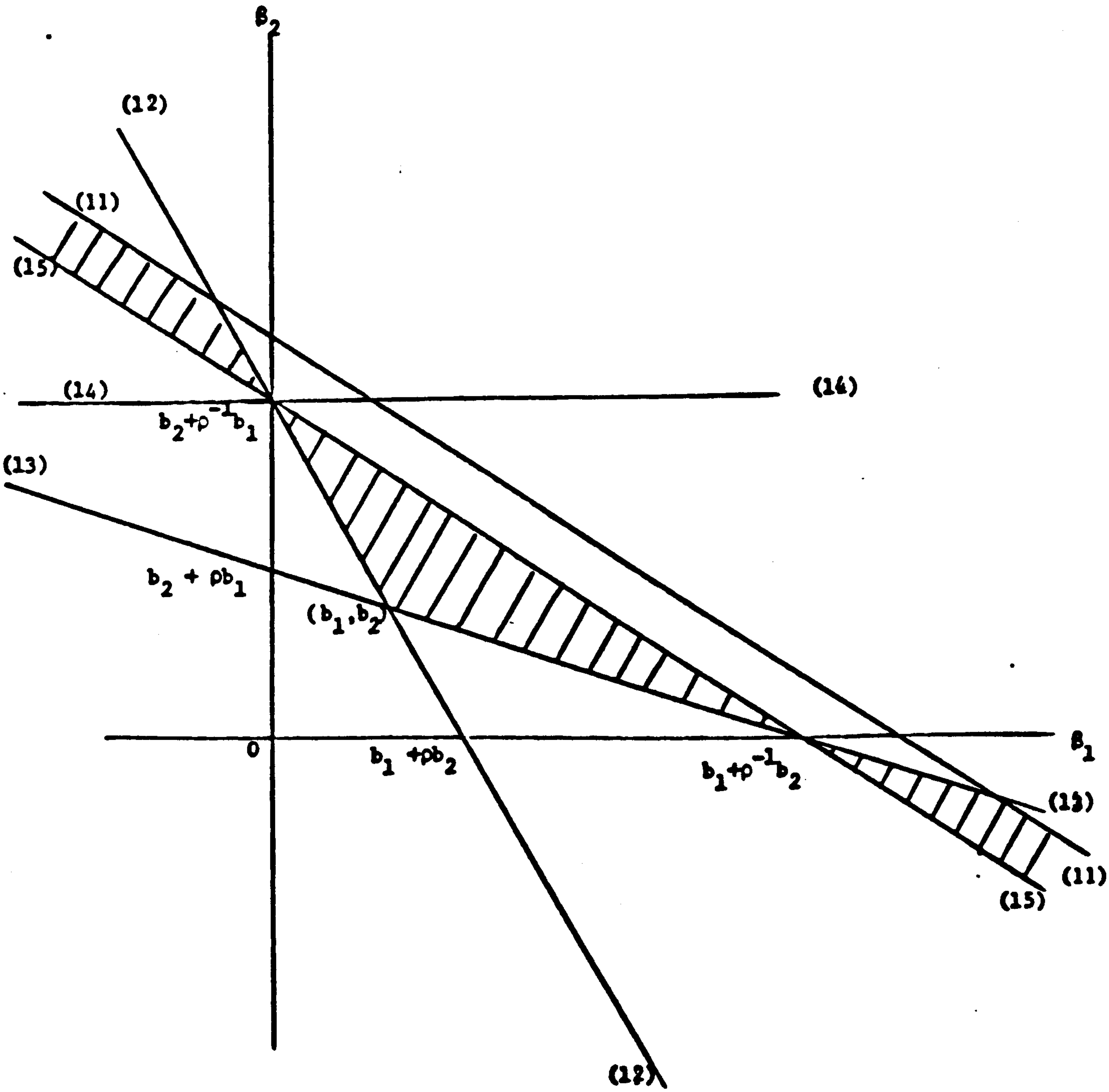

Figure 3

The Set of Maximum Likelihood Estimates
$b_1 > 0$, $b_2 > 0$, $\rho > 0$, $r_1r_2-\rho < 0$

Figure 4

BOUNDS FOR ESTIMATES AS A FUNCTION OF R^{*2}

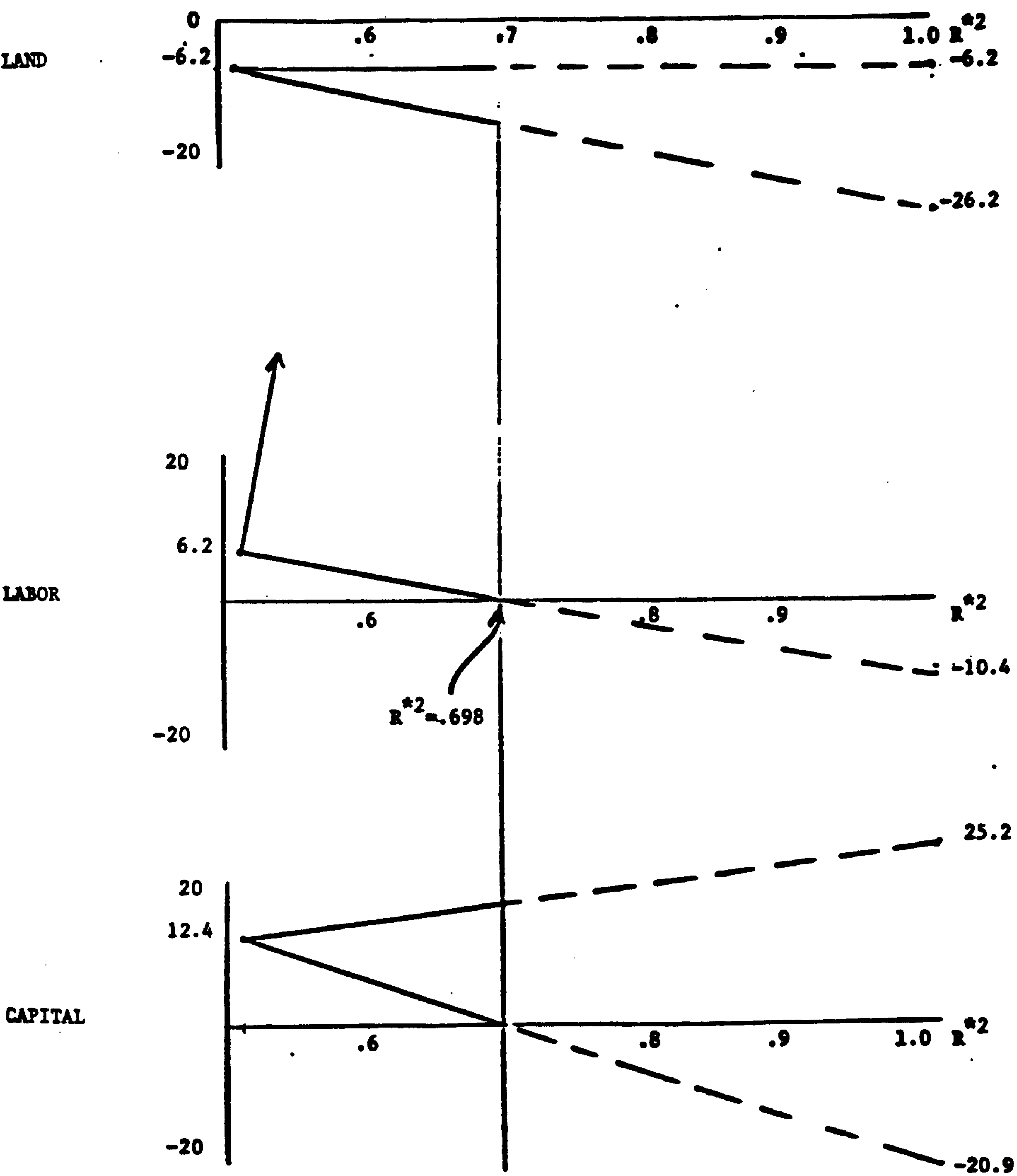

Table 1

Direct and Reverse Regressions

Coefficients	Y	Direction of minimization Land	Labor	Capital
Intercept/10^3	85.0	1062	-3861	-182
Land	-6.2	-22.7	-26.2	-10.3
Labor	6.2	26.3	708.2	-10.4
Capital	12.4	20.6	-20.9	25.2

$R^2 = .52$, $R_m^{*2} = .698$, $g = .3717$

Table 2

Bounds for Estimates Given
Minimum Squared Correlation Between True Variable and Measured Variable
$1-\lambda_1 = .565$

		Minimum Squared Correlation (ρ_k^2)					
		.5	.6	.7	.8	.9	1.0
Intercept/10^3:	Min	$-\infty$	-5283	-1937	-868.	-287.	85.0
	Max	∞	10813	2745	1202	498.	85.0
Land:	Min	$-\infty$	-114.7	-28.0	-14.7	-9.2	-6.2
	Max	∞	14.8	-1.2	-4.0	-5.4	-6.2
Labor:	Min	$-\infty$	-256.2	-65.6	-24.4	-5.1	6.2
	Max	∞	181.0	66.6	33.6	16.7	6.2
Capital:	Min	$-\infty$	-11.3	7.1	10.2	11.5	12.4
	Max	∞	143.7	39.5	23.1	16.2	12.4