

## 5 Econometric Methods

The estimation of the Heckscher-Ohlin-Vanek equations with linear least-squares regression requires a sequence of incredible assumptions: (1) Unmeasured determinants of trade can be collected into additive error terms, which, conditional on the measured determinants, behave like a sequence of independent, identically distributed normal random variables with zero means and constant, unknown variances. (2) The observed determinants of trade are measured without error. (3) The equation is linear. (4) The data set is so informative that other sources of information can be neglected.

Because none of these claims is believable, a data analysis that makes use of them need not be taken seriously unless the consequent inferences can be demonstrated to be adequately insensitive to changes in these assumptions. In this chapter, we discuss econometric methods that are designed to determine which inferences are excessively fragile and which are adequately robust. In section 5.1 the related problems of heteroscedasticity, gross errors, and nonlinearities are discussed. Cross sections of data with vastly different countries can imply very strong inferences that can be completely reversed if observations are reweighted to adjust for heteroscedasticity, or if subsets of observations are omitted, or if a nonlinear model is used. Systematic heteroscedasticity and nonlinear models are standard fare in econometrics texts and receive brief treatment here. Econometric methods to detect and adjust for gross errors are not standard material and are discussed more intensively. The method used in this research is to determine the effect on the estimates of one-at-a-time deletion of observations. This approach is selected with some discomfort because of its unclear methodological foundation.

Chronic measurement errors in the variables are the subject of section 5.2. The proper study of regression with all variables measured with error has only recently been stated by Klepper and Leamer (1984). Because the observed variability in the explanatory variables is due partly to true variability and partly to errors in measurements, an attempt to purge the observations of measurement error can lead to a perfectly collinear set of true variables. As will be shown, the data sets in this book are so highly correlated that it is impossible to find sensible assumptions about measurement errors that would assure that the true data are not perfectly collinear.

Data sets that are perfectly collinear do not admit inferences about individual regression coefficients unless they are supplemented with additional information. In section 5.3 supplemental information in the

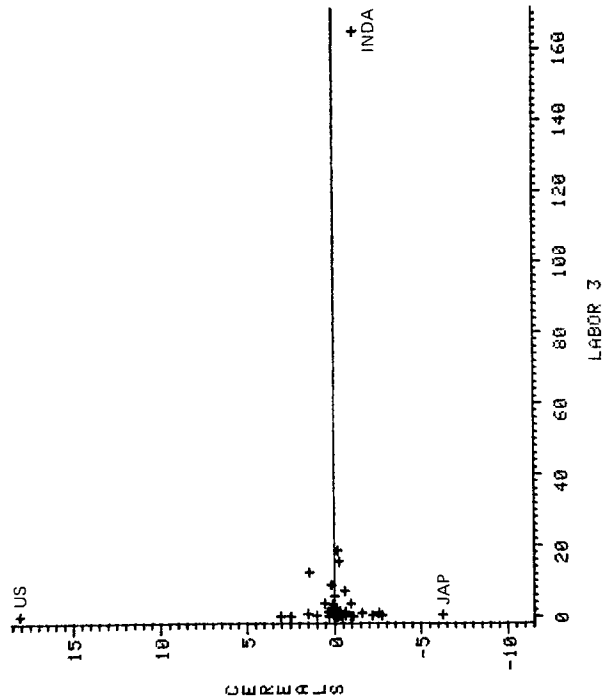


Figure 5.1  
Scatter diagram, CEREALS net exports versus LABOR 3: 1975.

form of a prior distribution is assumed and Bayesian methods are discussed. Because no multivariate prior distribution is particularly believable, a study of the sensitivity of the Bayes estimates to choice of prior distribution is essential. The sensitivity analyses reported in this section allow the prior covariance matrix to vary in certain ways, and they identify the corresponding variability of the Bayes estimates.

Finally, in section 5.4 it is argued that beta values, not  $t$ -values, are the better measure of the importance of a variable.

### 5.1 Heteroscedasticity, Gross Errors, and Nonlinearities

The econometric study of a cross section of countries must confront the problem that individual observations can exert an uncomfortably large influence on the estimates. In that event, weighting the observations or using a nonlinear model may greatly change the inferences. Features of the data set that are cause for concern are illustrated in figure 5.1–5.3. The net export of CEREALS in 1975 is plotted against LABOR 3 in

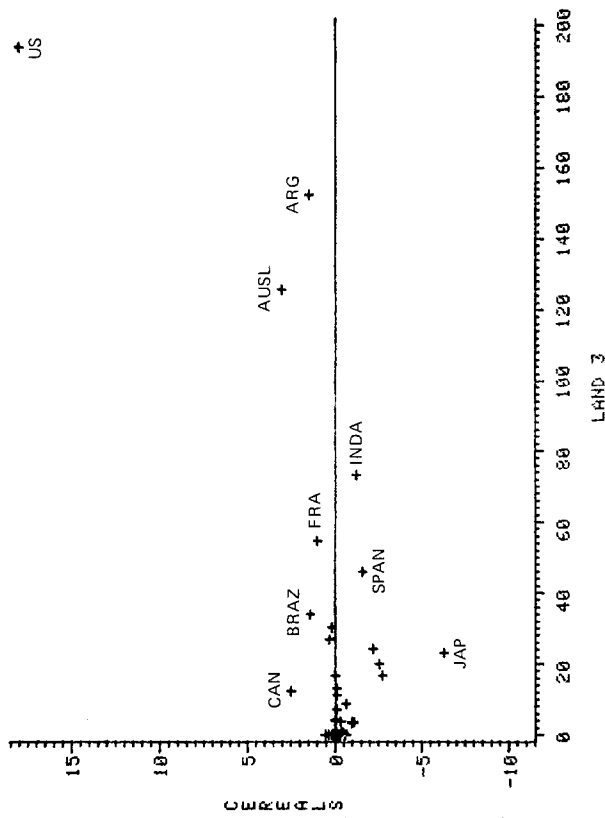


Figure 5.2  
Scatter diagram, CEREALS net exports versus LAND 3: 1975.

figure 5.1. It is difficult to detect any relation between CEREALS and LABOR 3 in this figure, but a computer printout can tell a different story. Estimates and  $t$ -values as a function of the observation subset are as follows: for no excluded observations, estimated slope =  $-7.9$  and  $t$ -value =  $-5$ ; for the United States excluded, estimated slope =  $-5.1$  and  $t$ -value =  $-7$ ; for India excluded, estimated slope =  $-5.6$  and  $t$ -value =  $-0.06$ ; and for the United States and India excluded, estimated slope =  $28.2$  and  $t$ -value =  $.6$ . When all observations are included, there is mild evidence of a negative relation. If the  $t$ -value of  $-5$  were treated as a posterior  $t$ -value, we could say that, given these data, the probability is approximately .8 that LABOR 3 contributes to comparative disadvantage in CEREALS. The conflict between the visual impression left by figure 5.1 and this estimated regression can be used to cast doubt either on human vision or on the method of regression. In this case, it seems clear that the method of regression is highly suspicious because its implications are extremely fragile. If the United States is excluded, the significance actually increases, but if India is excluded, the  $t$ -value falls virtually to zero. And

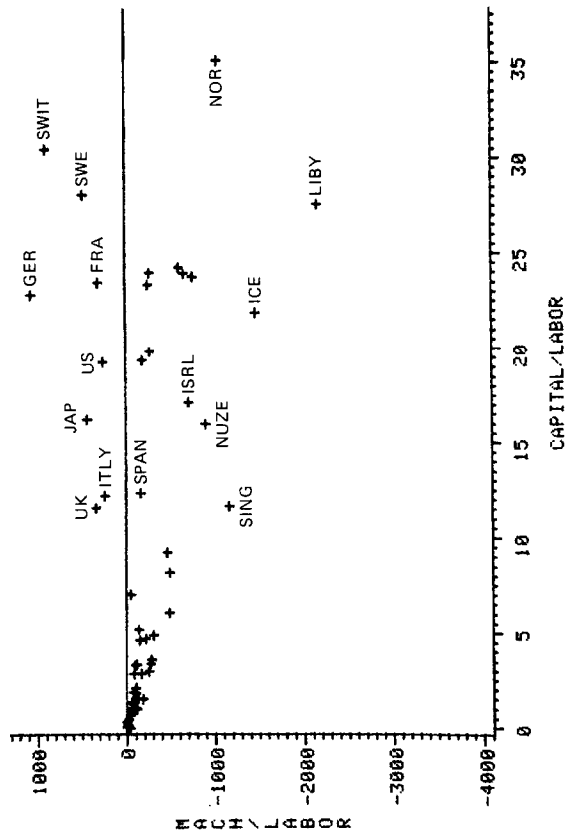


Figure 5.3  
Scatter diagram, MACH net exports/LABOR versus CAPITAL/LABOR: 1975.

if both the United States and India are omitted, the sign is actually reversed.

Figure 5.2 illustrates another case in which the United States is an influential observation. Here when the United States is excluded, the estimated effect of LAND 3 on CEREALS is much less, though still positive. Finally, figure 5.3, which compares 1975 MACHINERY/LABOR with CAPITAL/LABOR, illustrates a situation in which there is no statistically significant regression relation, but nonetheless the scatter has an interesting feature; countries with low levels of capital abundance are modest importers, but countries with higher levels of capital abundance can be either exporters or importers. An ordinary regression analysis would never reveal this feature.

This section reviews econometric methods designed to deal with these problems. Issues concerning the weighting of observations are usually discussed in the econometric literature under the heading "heteroscedasticity," which is the pathology of unequal residual variances, and which requires weights on observations inversely proportional to the residual variances. Gross errors, or outliers, which are associated with nonnormal

error terms, can also be treated by weighted regression, with relatively low weights placed on extreme observations. Because the treatment is the same, it is fair to suspect that the disease is the same. Indeed, it is the case that nonnormal errors can be thought to imply a type of heteroscedasticity.

For example, suppose we assumed that the regression residuals were drawn from a Student distribution that has relatively fat tails compared with normal and is more likely to produce outliers. A Student random variable  $t$  can be generated by selecting a random variance  $\sigma^2$  from an inverse gamma distribution and then, conditional on that value of  $\sigma^2$ , selecting the random variable  $t$  from a normal distribution with mean zero and variance  $\sigma^2$ . Thus the difference between a sequence of independent identically distributed normal random variables and a sequence of independent identically distributed Student random variables is that the Student sequence allows the (conditional) variance to vary whereas the normal sequence requires a constant variance. The word "condition" is important in this sentence since the marginal distribution of the Student random variable has constant variance. Before the Student sequence is observed, there is no information about which values of the variances are largest, but once the sample is selected, it is proper to draw the inference that the most extreme observations came from normals with the largest variances. In estimating a mean, or a regression function, this form of heteroscedasticity requires exactly the same cure as all others, namely, weighted regression with weights on observations inversely proportional to the variances. Consequently, extreme observations receive less weight.

The traditional form of heteroscedasticity in contrast is systematically related to other variables; for example, the variance of residual  $i$  may be modeled as

$$\sigma_i^2 = \alpha x_i^\theta \quad (5.1)$$

where  $x$  is an observable and  $\alpha$  and  $\theta$  are unknown parameters. Models of outliers, on the other hand, assume nonsystematic heteroscedasticity; for example, it may be assumed that  $\sigma_i^2$  is drawn from an inverse gamma distribution.

Our data set is likely to exhibit systematic heteroscedasticity, since the error term is hypothesized to represent unmeasured resource endowments and since countries with relatively large GNPs are likely to have relatively large stocks of the unmeasured resources. In particular, in the next chapter

we hypothesize the model (5.1) with explanatory variable  $x_i$  representing GNP. A preliminary examination of the use of total labor force as the predictor of heteroscedasticity yielded results that were significantly inferior to the GNP variable.

Maximum likelihood estimation of a regression equation with residual variances generated by equation (5.1) would require special computer programming, and I have opted instead to use the three-step approach in which the logarithms of the squared residuals from an unweighted regression of the trade equations are regressed against the logarithm of GNP to estimate the coefficients  $\alpha$  and  $\theta$  in equation (5.1). The predicted values of  $\sigma_i^2$  from this auxiliary regression are then used as weights for reestimating the trade equation. This relatively inexpensive alternative to maximum likelihood estimation has been studied by Harvey (1976), who demonstrates that the estimate of  $\theta$  is consistent and the corresponding weighted least-squares estimator is consistent and efficient. As a partial check of the approach, likelihood ratios of the homoscedastic model versus two heteroscedastic models were computed. This reveals that it is generally much better to weight by GNP than to use the unweighted estimates.

It can also be expected that there will be nonsystematic heteroscedasticity (outliers), possibly due to measurement errors, but also because of unusual countries that are not well described by the 11 basic resources. A formally proper way to detect and discount outlying observations is to hypothesize a family of nonnormal error distributions, say, a Student distribution, and to estimate the parameters of this distribution and the regression parameters jointly with maximum likelihood estimators (for example, see Blattberg and Gonedes, 1975.) This approach suffers from the particularity of the Student assumption as well as from its cost of implementation. Another approach is to study the asymptotic behavior of alternatives to least squares as in Huber (1972, 1973). Papers by Leamer (1981) and Gilstein and Leamer (1983a) have dealt with the issue of particularity by a global sensitivity analysis, but these are rather expensive methods as well. Instead, I adopt the relatively economical approach proposed by Belsley, Kuh, and Welsch (1980) and report the consequences of one-at-a-time deletions from the data set.<sup>1</sup> This informal approach does detect outliers and uncovers some sensitivities of the estimates, but it clearly cannot deal with extremes more than one-at-a-time; nor does it properly yield estimates and standard errors that account for the fact that

the data have been inspected for outliers. My decision to take this path reflects my opinion that there are other, more worrisome problems.

Observations that have extreme values for the explanatory variables as well as for the dependent variable can suggest either nonlinearities, systematic heteroscedasticity, or nonsystematic heteroscedasticity. A fairly typical scatter of observations with a cross-section sample of countries is depicted in figure 5.1, which has most countries with small values of LABOR 3, but one (India) with a very large value. A regression estimate with these data has a fairly significant LABOR 3 variable. But the conclusion that LABOR 3 is a major source of comparative disadvantage is very fragile, and is completely eliminated if India is omitted. It is inappropriate merely to discard such an observation. Rather, it is desirable to identify and to report these sources of fragility and to try to find credible assumptions that eliminate it.

The reason the inference is fragile is that there are many other possibilities why India may be different from the other countries, other than a large supply of LABOR 3. If such a reason can be conjured up, the result is an equally good model for explaining CEREALS trade. One candidate variable is the square of LABOR 3. That is, suppose the relation is quadratic rather than linear. Then the effect of LABOR 3 on CEREALS trade is estimated with great inaccuracy because there are many parabolas that fit the data equally well—namely, all those that go through the center of the cluster of observations and also through the extreme observation. A pattern such as that depicted in figure 5.1 will therefore be revealed either by adding a squared term to the regression or by omitting the extreme observation. The model with systematic heteroscedasticity can also reveal the fragility of the inference, since the extreme observation can be explained by a model in which LABOR 3 has no effect on the mean value of trade, but does significantly increase the residual variance.<sup>2</sup>

The one-at-a-time deletions from the data set make use of a rather peculiar set of weights, with no weight put on one observation and equal weights put on all others. I have studied the sensitivity of point estimates to other choices of the weights (Leamer, 1984). In particular, if the covariance matrix  $\Sigma$  of the residuals is assumed to satisfy  $I \leq \Sigma \leq 2I$ , where the matrix inequality  $A \leq B$  means  $B - A$  is positive semidefinite, then the sign of an estimate does not depend on the choice of  $\Sigma$  in this range if and only if  $|t| \geq ((n - k)/8)^{1/2}$ , where  $n - k$  is the degrees of freedom. For the samples considered below  $n - k = 46$  and the critical

value of the  $t$ -statistic is  $(46/8)^{1/2} = 2.4$ . If the  $t$ -statistic exceeds 2.4, the estimate can be said to be insensitive to reweighting of observations.

### 5.1.1 Example of the Effects of Omitting Observations

Table 5.1 indicates the effect of omitting observations from the 1975 CEREALS net export equation. The first row, labeled OLS, contains the estimates with all data included. The first column contains the  $t$ -value for country dummy variables entered one at a time into the equation. The other columns report estimates of the coefficients when one country is omitted. Consider first the LAND 3 column. The (weighted) least-squares coefficient with all the data included is 16.3. The omission of single observations generally does not have much of an impact on this estimate. The exceptions are, with estimates in parentheses, the United States (13.2), Libya (13.7), France (14.7), and Argentina (22.3). These estimates have to be interpreted with care, since they are the opposite of what you may be thinking. In particular, do not be confused into thinking that the United States data point favors the value 13.2. On the contrary, this is the estimate when the United States is excluded. Consider figure 5.2. Here all ways of excluding observations lead to large positive estimates except the omission of the United States, which produces a smaller estimate. The United States data point should be thought to favor a larger estimate, generated when it is included, not excluded. Referring now to table 5.1, we see that if the United States is included in the CEREALS equation, the estimate is 16.3, but that if the United States is excluded, the estimate is 13.2. Thus the U.S. data point makes LAND 3 appear to have a larger effect on CEREALS than the other data as a whole. Likewise, if Argentina is omitted, the coefficient increases dramatically, and Argentina is revealed to be holding down the estimate. The reason for this is easily understood. The United States and Argentina are the most abundantly endowed in LAND 3 (see the boxplot, figure 4.4) and have roughly the same endowments, but the United States has over 10 times the net exports of CEREALS of Argentina, which was nonetheless the fourth largest exporter. France also is a naturally influential observation, since it has the fifth largest endowment of LAND 3. The influence of Libya is a bit of a mystery—it has no LAND 3 and modest imports of CEREALS—but CEREALS imports per worker are greater than any other country, as can be seen in figure C.6a. It is clear from this figure that the omission of Libya makes land seem less important.

*The 1975 CEREALS equation: 1975*

Country	CAPITAL LABOR 1	LABOR 2	LABOR 3	LAND 1	LAND 2	LAND 3	LAND 4	COAL	MINERALS	OIL
OLS	-4.5	-51.0	-6.8	1.32	-4.83	16.09	-3.61	0.39	0.97	0.24
AFG	-4.4	-45.8	-6.2	1.28	-5.45	16.83	-3.92	0.40	1.03	0.24
ARG	-4.8	-50.5	-9.9	1.19	-5.14	20.93	-3.31	0.35	0.90	0.24
AUSL	-4.3	-31.7	-6.0	1.35	-3.11	16.35	-3.97	0.42	1.08	0.21
AUST	-4.5	-51.1	-7.1	1.32	-4.83	16.13	-3.62	0.39	0.97	0.24
BLUX	-4.4	-64.8	-6.5	1.34	-4.82	16.05	-3.59	0.39	0.96	0.24
BRAZ	-4.9	-42.4	-7.7	-1.07	-4.80	15.66	-4.04	0.36	1.03	0.25
BRMA	-4.0	-67.2	-3.1	1.23	-5.15	16.42	-4.01	0.37	1.03	0.24
CAN	-4.6	-2.5	-45.3	1.60	-4.14	16.33	16.66	0.35	0.89	0.21
CHLE	-4.4	-52.6	-5.0	0.80	-6.26	15.67	-5.61	0.33	1.32	0.25
CLUM	-4.4	-51.1	-6.7	1.28	-4.86	16.12	-3.65	0.39	0.98	0.24
CSTA	-4.5	-50.9	-6.9	1.32	-4.83	16.11	-3.63	0.39	0.97	0.24
CYPR	-4.5	-51.0	-7.1	1.32	-4.81	16.09	-3.61	0.39	0.97	0.24
DEN	-4.4	-50.0	-6.5	1.34	-4.85	16.10	-3.62	0.39	0.97	0.24
DORE	-4.5	-51.3	-7.2	1.31	-4.81	16.11	-3.62	0.38	0.97	0.24
ECUA	-4.5	-51.1	-6.9	1.31	-4.83	16.10	-3.63	0.39	0.97	0.24
EGPT	-4.5	-51.1	-6.8	1.32	-4.82	16.08	-3.61	0.39	0.97	0.24
FIN	-4.5	-56.0	-6.6	1.32	-4.88	16.17	-3.68	0.39	0.98	0.24
FRA	-3.9	-809.1	-6.1	7.8	-4.55	14.70	-2.30	0.72	0.78	0.26
GANA	-4.5	-50.7	-6.9	1.31	-4.83	16.11	-3.62	0.39	0.97	0.24
GER	-3.4	-108.1	-2.4	1.77	-4.60	15.16	-2.89	0.51	0.86	0.22
GRCE	-4.4	-50.8	-6.7	1.31	-4.88	16.17	-3.64	0.39	0.97	0.24
HOKO	-4.4	-50.6	-6.6	1.32	-4.83	16.05	-3.59	0.39	0.96	0.24
HOND	-4.5	-50.9	-7.0	1.32	-4.81	16.10	-3.60	0.39	0.97	0.24
ICE	-4.5	-51.0	-7.1	1.32	-4.81	16.04	-3.62	0.38	0.97	0.24
INDA	-5.5	-49.2	-70.7	2.06	-4.11	14.87	-3.76	0.25	0.91	0.27
INDO	-5.4	-36.8	-13.1	1.13	-4.88	15.77	-3.87	0.34	1.00	0.25
IRE	-4.5	-51.0	-6.8	1.32	-4.84	16.10	-3.61	0.39	0.97	0.24
ISRL	-4.5	-52.0	-7.0	1.30	-4.84	16.05	-3.64	0.38	0.97	0.24
ITLY	-4.5	-50.3	-8.9	1.13	-4.80	16.18	-3.75	0.34	0.98	0.24



hypotheticals. God (and very few others) knows how all these countries collect information on GNP, savings flows, labor force and so on. In a world of changing relative prices the measurement of capital as an accumulation of savings flows is not even conceptually correct. Endowments of oil, coal, and minerals are proxied by production data, but for various reasons some countries may exhaust their known supplies more quickly and/or look for new supplies more intensively than other countries.

The study of nonsystematic heteroscedasticity described in the previous section is intended partly to identify and to treat gross errors in the data by determining the effect of reducing the weights on observations that may be subject to gross measurement errors. This would not adequately deal with chronic errors. There is, however, an econometric tradition for dealing with chronic measurement errors. If it is assumed that there is a linear relation between two variables  $\eta_i$  and  $\chi_i$ ,  $\eta_i = \alpha + \beta\chi_i$ , and if it also assumed that there are independent measures of  $\eta_i$  and  $\chi_i$ ,  $y_i = \eta_i + \varepsilon_{yi}$ ,  $x_i = \chi_i + \varepsilon_{xi}$ , where the measurement errors  $\varepsilon_{yi}$  and  $\varepsilon_{xi}$  and the true variables  $\chi_i$  are independent serially uncorrelated random variables with means  $(0, 0, \bar{x})$  and variances  $(\sigma_y^2, \sigma_x^2, \sigma_x^2)$ , then the joint distribution of the observables  $y$  and  $x$  has mean vector and covariance matrix

$$E(y, x) = (\alpha + \beta\bar{x}, \bar{x}),$$

$$V(y, x) = \begin{bmatrix} \beta^2\sigma_x^2 + \sigma_y^2 & \beta\sigma_x^2 \\ \beta\sigma_x^2 & \sigma_x^2 + \sigma_x^2 \end{bmatrix}.$$

Estimates of the parameters may be found by setting these theoretical moments equal to observed sample moments:

$$\hat{\alpha} + \beta\hat{\bar{x}} = \bar{y},$$

$$\hat{\bar{x}} = \bar{x},$$

$$\beta^2\hat{\sigma}_x^2 + \hat{\sigma}_y^2 = s_y^2,$$

$$\beta\hat{\sigma}_x^2 = s_{xy},$$

$$\hat{\sigma}_x^2 + \hat{\sigma}_x^2 = s_x^2.$$

There are, however, seven unknown parameters, but only six equations. In a word, the model is underidentified. An arbitrary selection of one parameter may be made, and we may then solve for the others. If  $\hat{\beta}$  is arbitrarily selected, then estimates of the other parameters are

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x},$$

$$\hat{\bar{x}} = \bar{x},$$

$$\hat{\sigma}_x^2 = s_{xy}/\hat{\beta},$$

$$\hat{\sigma}_y^2 = s_y^2 - \hat{\beta}s_{xy},$$

$$\hat{\sigma}_x^2 = s_x^2 - (s_{xy}/\hat{\beta}).$$

Of course, it is known that the variances are positive, and values of  $\hat{\beta}$  must be selected that imply positive estimates of these parameters. The restriction  $\hat{\sigma}_x^2 > 0$  implies that  $\hat{\beta}$  and  $s_{xy}$  must have the same sign. The other two inequalities may be written either as  $s_{xy}/s_x^2 \leq \hat{\beta} \leq s_y^2/s_{xy}$  or as  $s_y^2/s_{xy} \leq \hat{\beta} \leq s_{xy}/s_x^2$ , depending on the sign of  $s_{xy}$ .

The ratio  $s_{xy}/s_x^2$  is the "direct" regression, formed by regressing  $y$  on  $x$ . The ratio  $s_y^2/s_{xy}$  is the "reverse" regression, formed by first regressing  $x$  on  $y$  and then by inverting the estimated equation to express  $y$  as a function of  $x$ . Alternatively, the direct regression can be found by minimizing the sum of squared residuals with distance between an observation point and the regression line measured in the direction of the  $y$  axis (vertically), and the reverse regression may be found by measuring the distance in the direction of the  $x$  axis (horizontally).

The inequalities for  $\hat{\beta}$  in the previous paragraph can then be summarized by the statement that the set of estimates of  $\beta$  compatible with the observed means, variances, and covariances is the interval between the direct regression  $\hat{\beta}^D = s_{xy}/s_x^2$  and the reverse regression  $\hat{\beta}^R = s_y^2/s_{xy}$ . This interval of estimates is also the set of maximum likelihood estimates if normal distributions are assumed, and it will capture the true value of  $\beta$  with probability approaching one as sample size increases. The length of the errors-in-variable interval is a decreasing function of the squared correlation; in particular, the ratio  $\hat{\beta}^D/\hat{\beta}^R$  is the squared correlation  $s_{xy}^2/s_x^2 s_y^2$ . This is one of the few cases in econometric theory in which the size of the squared correlation really matters.

The treatment of chronic errors in variables therefore begins with the computation of the direct and the reverse regressions. If this interval of estimates is narrow enough to be useful, then it can be reported that chronic measurement errors are revealed by the data to be inconsequential. If the interval of estimates is uselessly large, then an attempt may be made to narrow the interval by appeal to prior information about either of the

In the typical case when the  $k + 1$  regressions are not all in the same orthant and the set of maximum likelihood estimates is unbounded, a restricted set of estimates will be bounded if one of the following restrictions is imposed:

- a. The squared correlation between each true explanatory variable and its measurement exceeds  $1 - \lambda_1$ , where  $\lambda_1$  is the minimum eigenvalue of the correlation matrix of the measured explanatory variable.
- b. The squared multiple correlation between the dependent variable and the true explanatory variable does not exceed

$$R_m^2 = R^2 + (1 - R^2) \left( \max_{i,j} (1 - \hat{\beta}_{ij}/b_j) \right)^{-1},$$

where  $R^2$  is the squared multiple correlation based on the measured variables,  $b_j$  is the usual least-squares estimate of coefficient  $j$ , and  $\hat{\beta}_{ij}$  is the reverse regression estimate of coefficient  $j$  with residuals minimized in the direction of variable  $i$ . For comparison across equations with different  $R^2$  values, it is convenient to report the proportion of the gap between  $R^2$  and 1 that can be attributed to measurement errors without destroying the usefulness of the data:

$$g = (R_m^2 - R^2)/(1 - R^2) = \left( \max_{i,j} (1 - \hat{\beta}_{ij}/b_j) \right)^{-1}.$$

These two diagnostics,  $1 - \lambda_1$  and  $R_m^2$ , are constructive in the sense that they indicate assumptions that are sufficient to imply that the true data are not perfectly collinear. Leamer (1983) suggests destructive diagnostics, which are sets of measurement error variances that would imply that the true explanatory variables are perfectly collinear. If error variances in the neighborhoods of these cannot be ruled out on an a priori basis, then the data alone are useless, and prior information that restricts the values of the coefficients either absolutely or probabilistically would have to be employed.

The preceding discussion is slightly modified if a subset of variables is known to be measured perfectly. Then only the reverse regressions corresponding to the mismeasured variables need to be computed, and only the coefficients on the mismeasured variables need to be inspected for sign changes.

Parenthetically, it may be noted that the Hecksher-Ohlin-Vanek model defines a system of equations with the same set of hypothetical variables

two measurement error variances. If credible restrictions on the error variances do not produce usefully narrow sets of estimates, then we are forced to conclude that measurement errors may be so severe that these data by themselves are useless.

This approach to the treatment of chronic measurement errors has been extended to the multiple regression setting by Klepper and Leamer (1984). If there are  $k$  explanatory variables, then there are  $k + 1$  ways of estimating the regression, depending on the direction in which the residuals are measured, or, equivalently, depending on which variable is used as the left-hand-side variable. If these  $k + 1$  regressions have the same sign pattern (after normalization) then any weighted average of these  $k + 1$  regressions is a maximum likelihood estimate, and any point that is not a weighted average is not a maximum likelihood estimate. This is the straightforward generalization of the bivariate errors-in-variables bound. But if these  $k + 1$  regressions have sign changes, then the set of maximum likelihood estimates is unbounded. The reason the set is unbounded is the following. The measured covariance matrix is equal to the true covariance matrix plus the (diagonal) measurement error covariance matrix. To recover the true covariance matrix from the measured covariance matrix, we must subtract out the measurement error covariance matrix. The result can be a singular matrix. If the result is a singular matrix, the normal equations based on this estimated covariance matrix do not admit a unique solution, the set of all least-squares solutions is unbounded, and standard errors may be thought to be infinite. If the  $k + 1$  regressions all have the same sign patterns, no choice of the measurement error covariance matrix that is compatible with the first two observed moments can imply a singular true covariance matrix, and the set of possible estimates is consequently bounded. But when the observed variables are sufficiently collinear that the signs of the estimates depend on the direction in which the sum-of-squared residuals is minimized, then, using the data alone, we cannot rule out the possibility that the true variables are perfectly collinear.

If the set of estimates compatible with the first two moments of the data is too large to be useful, then some prior information on the measurement error variances may prove helpful. Klepper and Leamer (1984) consider in particular two forms of prior restrictions: (a) the minimum squared correlation between the true variable and its measurement and (b) the maximum value for the squared multiple correlation if the measurement errors in the explanatory variables were removed.



in each. The foregoing discussion is applicable to a single equation, but the evidence can be much stronger if the whole system is analyzed because the same hypothetical variables enter each equation. The proper treatment of underidentified systems of equations with errors in variables has not been worked out, and we consequently study the data one equation at a time. Another shortcoming of this treatment is that the measurement error variance is assumed to be the same for all countries—large and small. A better model or errors in measurement might be the logarithmic model  $\log x_i = \log \xi_i + \varepsilon_{xi}$ , where  $x_i$  is the measurement and  $\xi_i$  is the true variable. This, too, creates econometric theory problems that have not been fully worked out.

**5.2.1 Example of Estimation with Chronic Measurement Errors**

Direct and reverse regressions for the 1975 machinery equation are reported in table 5.2. The first column in table 5.2 contains the usual least-squares estimates. The second column contains a reverse regression estimate formed by regressing CAPITAL on machinery net exports and the other resource variables and then solving the estimated equation for machinery in terms of all the resource endowments. Equivalently, the size of each residual is measured as the difference between an observation point and the regression plane in the direction of the CAPITAL axis. A feature of this regression is that the coefficient of the CAPITAL variable is enlarged; in fact, it is multiplied by the inverse of the squared partial correlation between MACH and CAPITAL. If it is thought that CAPITAL is the only variable (other than MACH) that is measured with error, then only these first two regressions need to be considered, and the set of maximum likelihood estimates is just the interval of estimates between these two. It could then be concluded that chronic measurement errors in CAPITAL would greatly limit the inferences that can be made about the effect of LAND 4, since either sign is possible. The other coefficients do retain their signs and many change little in magnitude.

If CAPITAL and LABOR 1 are both suspected to be contaminated by measurement errors, then the first three columns in this table must be considered. Because neither the CAPITAL coefficient nor the LABOR 1 coefficient changes sign in the first three columns, the set of maximum likelihood estimates remains bounded. The set of estimates is still bounded even if CAPITAL and all three LABOR variables are possibly measured with error, since the subset of coefficients corresponding to these variables

Table 5.2  
Direct and reverse regressions, MACHINERY equation: 1975<sup>a</sup>

	CAPITAL	LABOR 1	LABOR 2	LABOR 3	LAND 1	LAND 2	LAND 3	LAND 4	COAL	MINERALS	OIL
CAPITAL	32.2	67.6	130.2	33.1	443.7	2.6	31.4	-12.1	38.8	33.6	13.0
LABOR 1	-2,118	-8,850	-42,548	-37,154	-77,848	-81,277	14,686	2,321	3,158	-10,321	-2,572
LABOR 2	107	110	1,874	3,452	405	-847	-229	-53	600	524	205
LABOR 3	25	285	905	93	3,261	2,453	-514	-252	-642	115	-41
LAND 1	.6	8.5	24	-5	817	41	2.7	105	13	26	-182
LAND 2	7.1	.6	-49	-15	-147	466	382	70	403	4	100
LAND 3	-33	-33	37	16	34	-145	-330	-299	-51	-26	-39
LAND 4	3.4	-1.3	-5.0	19	-87	569	191	5.1	507	27.6	88.9
COAL	7.9	9.5	39	39	37	165	4.5	6.1	65	31	17
MINERALS	-4.2	-4	-5	-8	7	-177	-60	-5	-111	-9	-30
OIL	-2.6	-1	5	3	20	4	-9	-4	-16	-5	-9

<sup>a</sup> Each column contains a set of estimates formed by minimization of the residuals in the direction indicated.  $R^2 = .91, \rho = .037, 1 - \lambda_1 = .9934$ .

do not change sign in the first five columns. If, however, any one of the other variables is thought also to be measured with error, then the set of feasible estimates becomes unbounded, since relevant estimates change in sign.

Prior information may be useful in restricting these unbounded sets of estimates. The smallest eigenvalue of the correlation matrix is rather small, and in order for a bounded set of estimates to be obtained by appeal to information about the error variances, it would have to be assumed that the squared correlation between true variables and measured variables exceeds .9934. This seems to me to be an unreasonable assumption. Another form of prior information that may be entertained is the maximum value that the multiple correlation coefficient can be expected to attain if the explanatory variables were perfectly measured. If this number is not much higher than the  $R^2$  based on the mismeasured variables, then the set of maximum likelihood estimates is also bounded. The statistic  $g$  indicates the proportion of the difference between the measured  $R^2$  and 1 that could be attributed to measurement errors without causing the sets to become unbounded. This number is only .037, which means that if it is thought that the elimination of the measurement error could not increase the  $R^2$  by more than  $(1 - .91) \times .037 = .003$ , then the set is bounded. That seems to be an unreasonable assumption.

Destructive diagnostics are reported in table 5.3. Entries in this table indicate pairs of measurement error variances that would imply that the true data are perfectly collinear. As discussed in Leamer (1983), this form of collinearity affects all of the coefficients, not just the two measured with error. The numbers in table 5.3 are the destructive error variances divided by the observed sample variances, reported in pairs, with the row variance first and the column variance second. (The squared correlation between a measured variable and the true variable is equal to one minus this variance ratio.) Variance ratios less than .0005 are rounded off to .000. Pairs of asterisks indicate that perfect collinearity cannot be induced by assuming measurement error only in those two variables (for example, CAPITAL and LABOR 1). The numbers in this table are disturbingly small, and the possibilities for perfect collinearity seem abundant. For example, if the noise variance were .1% of the total variance for CAPITAL and LABOR 2, or for LABOR 1 and LABOR 2, or for LABOR 3 and LAND 2 . . . , the true variables would be perfectly collinear.

The conclusion that is suggested by this example is that sensible assump-

Table 5.3 Destructive diagnostics, MACHINERY equation: 1975<sup>a</sup>

	CAPITAL	LABOR 1	LABOR 2	LABOR 3	LAND 1	LAND 2	LAND 3	LAND 4	COAL	MIN-ERALS	OIL
CAPITAL	(*)	(*)	(*)	(*)	(*)	(*)	(*)	(*)	(*)	(*)	(*)
LABOR 1	(**)	(*)	(*)	(*)	(*)	(*)	(*)	(*)	(*)	(*)	(*)
LABOR 2	(**)	(*)	(*)	(*)	(*)	(*)	(*)	(*)	(*)	(*)	(*)
LABOR 3	(**)	(*)	(*)	(*)	(*)	(*)	(*)	(*)	(*)	(*)	(*)
LAND 1	(**)	(*)	(*)	(*)	(*)	(*)	(*)	(*)	(*)	(*)	(*)
LAND 2	(**)	(*)	(*)	(*)	(*)	(*)	(*)	(*)	(*)	(*)	(*)
LAND 3	(**)	(*)	(*)	(*)	(*)	(*)	(*)	(*)	(*)	(*)	(*)
LAND 4	(**)	(*)	(*)	(*)	(*)	(*)	(*)	(*)	(*)	(*)	(*)
COAL	(**)	(*)	(*)	(*)	(*)	(*)	(*)	(*)	(*)	(*)	(*)
MIN-ERALS	(**)	(*)	(*)	(*)	(*)	(*)	(*)	(*)	(*)	(*)	(*)
OIL	(**)	(*)	(*)	(*)	(*)	(*)	(*)	(*)	(*)	(*)	(*)

a. Noise variance divided by total variance. If these noise variances are subtracted from the observed variances, the covariance matrix becomes singular. (\*\*, \*) indicates that no singular matrix is producible.

tions about the errors of measurement preclude estimation from these data alone, since we cannot rule out the possibility that the true explanatory variables are perfectly collinear. This same conclusion holds for every one of the estimated regressions.

### 5.3 Prior Information and the Collinearity Problem

Data sets with perfectly collinear explanatory variables cannot be used to estimate individual regression coefficients unless credible restrictions on the coefficients of some form are imposed. A fairly typical approach is to identify a set of "doubtful" variables and to see what happens to the regression equation as subsets of these doubtful variables are omitted. To say that a variable is doubtful is equivalent to announcing the opinion that the corresponding coefficient is small, and this approach thus implicitly supplements the given data set with another set of information about the coefficients. A Bayesian analysis is a formally proper way of supplementing the information in the given data set. The critical defect of a Bayesian analysis is that it presupposes that users can characterize their opinions about the regression coefficients in the form of probability distributions. The impossibility of making a credible selection of such a distribution forces practicing Bayesians to perform extensive sensitivity analyses to determine whether the precise choice of prior distribution matters very much.\*

As an introduction, consider a regression model in which it is generally agreed that a set of variables "belongs" in the equation, but there exists also a set of variables that "may or may not belong." (Quotation marks surround ambiguous words.) The variables that are certainly included in the equation will be symbolized by  $x$ ; the doubtful variables will be denoted  $z$ . A three-variable linear regression model with one certain and two doubtful variables can be written as

$$y_t = \beta x_t + \gamma_1 z_{1t} + \gamma_2 z_{2t} + u_t, \quad (5.2)$$

where  $t$  indexes a set of  $T$  observations,  $u_t$  is assumed to be a sequence of independent normal random variables with mean zero and unknown variances  $\sigma^2$ ,  $(y_t, x_t, z_{1t}, z_{2t})$  is a sequence of observable vectors, and

\*The following material, up to section 5.3.1, is taken in part from Leamer and Leonard (1983). For a more extensive treatment consult Leamer (1978).

$(\beta, \gamma_1, \gamma_2)$  is an unobservable vector of regression coefficients. Inferences are to be drawn from a data set about the effect  $\beta$  of the variable  $x$  on the dependent variable  $y$ . In an ideal experiment, the variables  $z_1$  and  $z_2$  would have been controlled at some constant level. As a substitute for experimental control, the variables  $z_1$  and  $z_2$  are included in the equation.

A researcher wishing to show that  $\beta$  is large or finding it difficult to estimate  $\gamma_1$  and  $\gamma_2$  accurately might estimate the four different regressions using different subsets of the "control" variables  $(z_1, z_2)$  and select for reporting purposes the most favorable result. The alternative procedure, which is advocated here, is, first, to enlarge the search and, second, to require reporting of both the most favorable and the least favorable outcomes. The search may be enlarged by defining a composite control variable

$$w_t(\theta) = z_{1t} + \theta z_{2t},$$

where  $\theta$  is a number to be selected by the researcher. The regression model is now

$$y_t = \beta x_t + \eta w_t(\theta) + u_t. \quad (5.3)$$

Each value of  $\theta$  selects a different constraint of the form  $\gamma_2 = \theta\gamma_1$  and consequently a different method for estimating  $\beta$ . Allowing  $\theta$  to take any value contrasts with the usual search procedure, in which  $\theta$  is implicitly permitted to take one of only four values. The usual procedure has the virtue of historical acceptance and the additional merit that it is comparatively easy to carry out computationally in the context of the existing econometric technology. As is discussed subsequently, there are settings in which the method is completely justifiable. This is not always the case, and we now expand the search to include all values of  $\theta$ .

To each value of  $\theta$  there is a least-squares estimate of  $\beta$ ,  $\hat{\beta}(\theta)$ . The most "favorable" value of  $\theta$ , for the researcher who wishes to show  $\beta$  is large, is found by maximizing  $\hat{\beta}(\theta)$  with respect to  $\theta$ , and the least favorable value is found by minimizing  $\hat{\beta}(\theta)$ . These extreme values,  $\hat{\beta}_{\min}$  and  $\hat{\beta}_{\max}$ , delineate the ambiguity in the inferences about  $\beta$  induced by the ambiguity in choice of model. If the interval  $[\hat{\beta}_{\min}, \hat{\beta}_{\max}]$  is short in comparison with the sampling uncertainty, or if all estimates in the interval imply the same decision, the ambiguity in the model may be considered irrelevant, since all models lead to essentially the same inferences. But if the bound is wide,

an effort should be made to narrow the family of models, and, it is hoped, to sharpen the inferences. One way to narrow the family of models is to constrain the parameters  $\gamma_1$  and  $\gamma_2$  to lie within the ellipse

$$\gamma_1^2 a^2 + \gamma_2^2 \leq r^2, \quad (5.4)$$

where  $a$  is the relative length of the two principal axes and  $r$  is the radius. This may seem to be a peculiar thing to do, but this constraint is the foundation of the voluminous literature on "biased estimation."<sup>3</sup> It can be justified in the following way. The only compelling reason for the omission of the  $z$  variables is that they are thought to be doubtful. If they truly do not belong in the equation, then a better estimate of  $\beta$  can be produced by an equation with the  $z$  variables omitted. To say that the  $z$  variables are doubtful is to say that the parameters  $\gamma_1$  and  $\gamma_2$  are small. One precise definition of smallness is given by equation (5.4), and a natural way to estimate the parameter  $\beta$  is to use least squares subject to the constraint (5.4). Henceforth, this constraint will be called the prior ellipse in reference to the fact that it represents information about  $\gamma_1$  and  $\gamma_2$  that is available prior to the data analysis.

If the parameters of the prior ellipse,  $a^2$  and  $r^2$ , were known, this procedure would generate a unique estimate, but we are unaware of any real data analysis in which the values  $a^2$  and  $r^2$  could sensibly be taken as given. For any value of  $a^2$  and  $r^2$ , there is a constrained least-squares estimate,  $\hat{\beta}(a^2, r^2)$ , computed by minimizing the sum of squared residuals subject to the constraint (5.4). We now turn to an examination of the function  $\hat{\beta}(a^2, r^2)$ .

Consider first the case when  $a^2$  is known, taken without loss of generality to be equal to one. For each value of  $r^2$ , equation (5.4) defines a circle located at the origin, depicted in figure 5.4. Also depicted in figure 5.4 are the unconstrained least-squares estimates of  $\gamma_1$  and  $\gamma_2$ ,  $(g_1, g_2)$ , and the contours of equal residual sums of squares around  $(g_1, g_2)$ . It should be noted that the sum-of-squared residuals depends on the estimate of  $\beta$  as well as on the estimates of  $\gamma_1$  and  $\gamma_2$ . But  $\hat{\beta}$  is selected by regressing  $y - \gamma_1 z_1 - \gamma_2 z_2$  on  $x$ , which implies that  $\hat{\beta}$  is linear function of  $\hat{\gamma}_1$  and  $\hat{\gamma}_2$ . With this estimate of  $\beta$ , the residual sum of squares can be expressed as a function of  $\hat{\gamma}_1$  and  $\hat{\gamma}_2$  only.

The estimation problem of minimizing the residual sum of squares subject to the constraint (5.4) can be described graphically in terms of a tangency point between a sum-of-squares ellipse and the prior circle

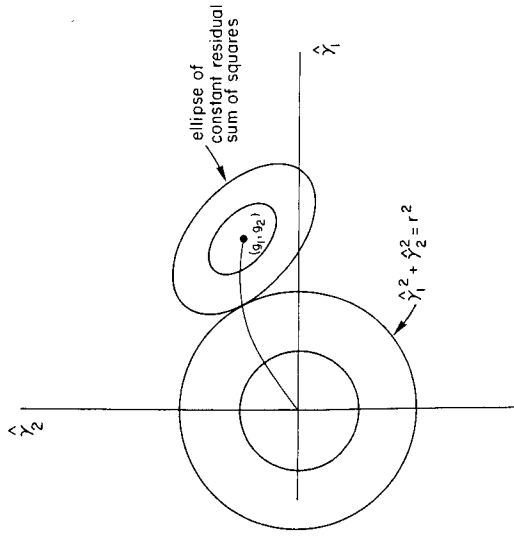
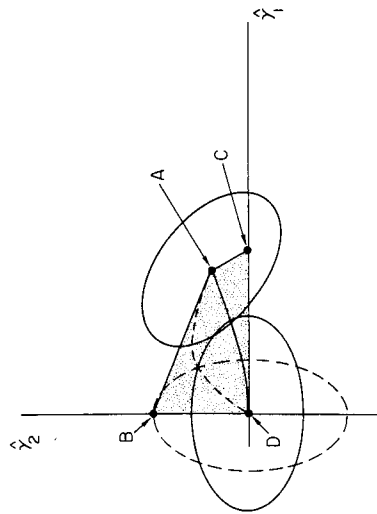


Figure 5.4  
An information contract curve.

located at the origin. As the radius of the circle is varied, a curve of estimates is formed that we call an *information contract curve*.<sup>4</sup> This language is selected to suggest the Edgeworth-Bowley analysis of trade between a pair of consumers, a setting that is analogous to our own problem both mathematically and substantively. In the Edgeworth-Bowley analysis, a contract curve represents the Pareto efficient allocation of a pair of commodities to a pair of consumers with conflicting desires. Here, the information contract curve represents the "Pareto efficient" set of estimates given two conflicting sources of information.

The choice of a point on the contract curve in the Edgeworth-Bowley analysis requires cardinal utility and a social welfare criterion. To put it differently, there has to be a way of comparing the utilities of the two consumers. Analogously, the choice of a point on the information contract curve requires us to compare the strength of the two information sources, a problem to which we shall return.

Next consider the case when neither  $a^2$  nor  $r^2$  can be taken as known. For any  $a^2$  there will be a contract curve, two of which are depicted in figure 5.5. The hull of all such curves is the shaded area in figure 5.5, which has been shown by Leamer and Chamberlain (1976) to be a subset of the



**Figure 5.5**  
Contract curves in the hull of 2<sup>d</sup> regressions: A = unconstrained least-squares estimate; B = least squares given  $\hat{\gamma}_1 = 0$ ; C = least squares given  $\hat{\gamma}_2 = 0$ ; D = least squares given  $\hat{\gamma}_1 = 0$  and  $\hat{\gamma}_2 = 0$ .

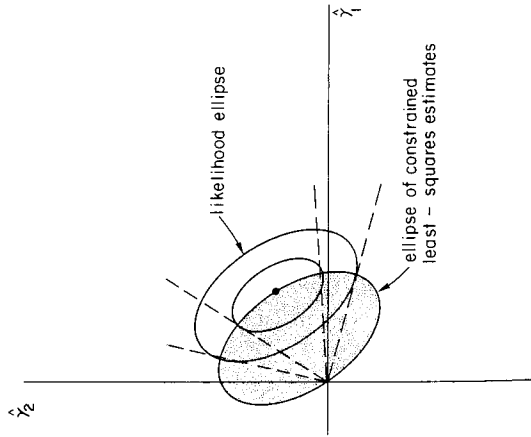
set of all weighted averages of the four regressions formed by omitting or including the two  $z$  variables. This brings us back to the traditional procedure. Now we have a justification for it: If, in the researchers' opinion,  $\gamma_1$  and  $\gamma_2$  are thought to be small in the sense of the ellipse (5.4), but neither  $a$  nor  $r$  is known, then the extreme estimates that can be generated from the sample are the four regressions formed by omitting the  $z$  variables. We would then recommend, in fact require, that both the minimum and the maximum estimate of  $\beta$  from among this set of four regressions be reported.

The widest bound for  $\beta$  swept out by the parameter  $\theta$  can also be depicted graphically. The prior ellipses so far considered all have axes in the coordinate directions. If the quadratic form were, more generally,

$$\gamma_1^2 a^2 + \gamma_2^2 + c\gamma_1\gamma_2 \leq r^2,$$

then the ellipse is tilted, and estimates can lie outside the shaded area in figure 5.5. The hull of all contract curves, with all families of prior ellipses, is the shaded area in figure 5.6. The boundary of this region, which is an ellipse, is the set of constrained least-squares points subject to constraints of the form  $\gamma_2 = \theta\gamma_1$ , where  $\theta$  varies from  $-\infty$  to  $\infty$ .

We have now considered three bounds for the estimates of  $\beta$  that can be generated from a given data set. The choice among these bounds depends on how precisely the researcher is willing to define the vague



**Figure 5.6**  
Ellipse of constrained estimates.

notion that the  $z$  variables are doubtful. If it can only be said that the  $z$  variables are doubtful, the widest set of estimates, depicted in figure 5.6, applies. If it can be agreed that the definition of doubtful should be restricted to mean that  $\gamma_1^2 a^2 + \gamma_2^2$  is likely to be small, where  $a^2$  is left undetermined, then the shaded area in figure 5.5 is the bound. Finally, if doubtful is even more precisely defined to mean that  $\gamma_1^2 + \gamma_2^2$  is small, then the contract curve in figure 5.4 defines the bound.

To narrow the bounds further, it will be necessary to devise a method for choosing points from a contract curve. As in the Edgeworth-Bowley analysis, this will require a comparison of the strengths of the two sources of information. If the sample information is relatively precise, it will be better to select points relatively close to the least-squares point. Conversely, if the prior information is relatively precise, we shall prefer points in the neighborhood of the prior point, the origin in the example. One way to make the prior information comparable to the sample information is to act as if the prior information came from a previous set of observations. Suppose we observed the process

$$y_t^* = \gamma_1 z_t^* + \gamma_2 z_{2t}^* + u_t^*,$$

with  $u_i^*$  normally distributed with variance  $\sigma_1^2$ . Suppose also that

$$\sum_t z_{1t}^* z_{2t}^* = 0, \quad \sum_t z_{it}^{*2} = 1, \quad \text{and} \quad \sum_t z_{it}^* y_t^* = 0. \quad (5.5)$$

Then the least-squares estimate of  $\gamma_1$  and  $\gamma_2$  based on these data would be  $(0, 0)$  with variance matrix  $\sigma_1^2 I$ .

Now consider pooling this "prior" sample with the sample generated by equation (5.2). The pooled estimates are computed using the usual formula with pooled sample moments:

$$\begin{bmatrix} \hat{\beta} \\ \hat{\gamma}_1 \\ \hat{\gamma}_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ \sigma_1^{-2} & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} + \sigma^{-2} \begin{bmatrix} x'x & z_1'x & z_2'x \\ z_1'x & z_1'z_1 & z_1'z_2 \\ z_2'x & z_2'z_1 & z_2'z_2 \end{bmatrix}^{-1} \begin{bmatrix} x'y \\ z_1'y \\ z_2'y \end{bmatrix} \cdot \sigma^{-2}. \quad (5.6)$$

Using the partitioned inverse rule, we can obtain

$$\begin{bmatrix} \hat{\gamma}_1 \\ \hat{\gamma}_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \sigma_1^{-2} & 0 \\ 0 & 1 \end{bmatrix} + \sigma^{-2} \begin{bmatrix} z_1'Mz_1 & z_1'Mz_2 \\ z_2'Mz_1 & z_2'Mz_2 \end{bmatrix}^{-1} \begin{bmatrix} z_1'My \\ z_2'My \end{bmatrix} \sigma^{-2},$$

where  $M = I - x(x'x)^{-1}x'$ . This equation describes the pooled estimates of  $\gamma_1$  and  $\gamma_2$  as a function of the variance ratio  $\sigma_1/\sigma^2$ . If this variance ratio is small, the estimates will be close to the "prior" estimates  $(0, 0)$ , and if this variance ratio is large, the estimates will be close to the least-squares estimates from the second data set  $(g_1, g_2)$ . Moreover, as the variance ratio is varied from zero to infinity, the estimates  $(\hat{\gamma}_1, \hat{\gamma}_2)$  will sweep out exactly the contract curve depicted in figure 5.4, where the contours (5.4) of the prior ellipse are the likelihood contours implied by the prior sample  $y^*, z_1^*, z_2^*$ .

To the extent that the prior information can be considered as coming from a hypothetical normal experiment, and to the extent that we can select the variance ratio  $\sigma_1^2/\sigma^2$ , we are now able to pick a particular point from the contract curve. The variance  $\sigma^2$  can be estimated from the data  $y, x, z_1, z_2$ . The prior variance  $\sigma_1^2$  presents greater difficulty. This number determines the size of the prior confidence intervals for  $\gamma_i$  since a 95% interval, for example, is  $|\gamma_i| \leq 1.96\sigma_1$ . In selecting a value of  $\sigma_1$ , it is therefore necessary to indicate the degree of confidence in the statement that  $\gamma_i$  is small. If  $|\gamma_i| \leq 1.96$  is believed with considerable confidence, then  $\sigma_1 = 1$  may be a useful starting point.

It seems unlikely that a precise number for  $\sigma_1$  could ever be selected. A sensitivity analysis in which several different values of  $\sigma_1$  are selected

is therefore desirable. A researcher might sensibly constrain  $\sigma_1$  to an interval such as  $.5 \leq \sigma_1 \leq 2$ . As  $\sigma_1$  is varied in this interval, a subset of points on the contract curve is selected. Often this subset of points will be so narrow that the ambiguity that remains will be for practical purposes irrelevant.

A sensitivity analysis can be applied to other features of the prior covariance matrix. Let  $\Omega$  denote the  $3 \times 3$  prior covariance matrix that above has been taken to be the  $3 \times 3$  diagonal matrix  $\sigma_1^2 \cdot \text{diag}\{\infty, 1, 1\}$ . Then  $\Omega^{-1}$  replaces the  $3 \times 3$  diagonal matrix  $\sigma_1^{-2} \cdot \text{diag}\{0, 1, 1\}$  in equation (5.6). Leamer (1982) has shown how restrictions on  $\Omega$  imply restrictions on the pooled estimate  $(\hat{\beta}, \hat{\gamma}_1, \hat{\gamma}_2)$ . It is required that  $\Omega$  be bounded from above,  $\Omega \leq \Omega_U$ , or from below,  $\Omega_L \leq \Omega$ , where  $A \leq B$  means  $B - A$  is positive semidefinite. An alternative to the interval of values for  $\sigma_1, .5 \leq \sigma_1 \leq 2$ , is the interval of matrices,  $\text{diag}(\infty, .25, .25) \leq \Omega \leq \text{diag}(\infty, 4, 4)$ . The interval of matrices restricts the prior standard error for  $\gamma_i$  to be between .5 and 2, but does not restrict the prior standard error of  $\gamma_1$  to be equal to the prior standard error of  $\gamma_2$ , nor does it restrict the covariances to be zero.

Though this example has made use of the prior information that some of the coefficients are close to zero, merely by choice of origin it applies equally well to prior information that some or all of the coefficients are close to selected nonzero values. It needs to be emphasized that a credible selection of a precise and complete multivariate prior distribution cannot be made, and the choice of some particular distribution to characterize prior information is inherently whimsical. The sensitivity analyses that have just been discussed are designed to control the whimsy. In principle, the sensitivity analysis should encompass all arbitrary assumptions about the prior distribution, including (1) the form of the distribution, (2) the location of the distribution, and (3) the dispersion of the distribution. I do not possess computer software that can do a sensitivity analysis of the choice of distributional form, but the effect of the choice of location and the choice of variance-covariance matrix given a normal distribution can be studied. In this book the location and the form of the prior will be taken as known, and the effect of changing the prior variance-covariance matrix on the point estimates will be studied. A particular hypothetical prior covariance matrix  $\Omega_0$  will be selected. First, this prior covariance matrix is multiplied by a scalar  $\sigma_1$  and the point estimates of the coefficients are expressed as functions of  $\sigma_1$ . This shows how the estimates

a real rate of interest of, say, 3–4%, a coefficient of around .17 is obtained. However, there are two reasons why it would be unsurprising to obtain an estimate higher than .17. First, because we assume only a 15-year life, and because the CAPITAL figure excludes certain forms of human capital that may not be adequately captured by the LABOR 1/LABOR 2/LABOR 3 distinction, the capital stock number is likely to be a serious underestimate. Second, unmeasured resources that contribute to GNP cannot sensibly be assumed to be uncorrelated with capital because by generating product they allow greater total savings and consequently contribute to capital formation. For these reasons, we set the prior estimate of the capital coefficient to .5, with a large standard error of .5, thereby indicating not much more than the opinion that the coefficient is likely to be between zero and one.

We take information about the annual earnings of the other factors from U.S. sources, remembering that factor price equalization is presently assumed. Farm workers in the United States received annual earnings of \$1,700 in 1958 and \$5,200 in 1975. Workers in radio and television broadcasting earned \$7,100 in 1958 and \$13,000 in 1975.<sup>5</sup> These numbers serve as a basis for the prior means for the labor coefficients. Most agricultural workers in the United States fall into the LABOR 2 category: literate but not professional. Accordingly I use the U.S. agricultural earnings as the prior mean for the LABOR 2 coefficients. The radio and television broadcasting workers are the highest paid among the categories reported in the U.S. National Income and Products Accounts. In the United States there are technical and professional workers who are paid more, but the LABOR 1 categories also include lower-paying occupations such as teachers. Thus I will take the prior mean of LABOR 1 coefficients to be the earnings of U.S. broadcast workers. Table 1.1 indicates that agricultural workers in developing countries earn about a tenth of U.S. workers, and I will take the prior mean for the illiterate workers to be a tenth of the U.S. agricultural annual earnings. The prior standard deviations are all set equal to half the prior means to represent a high degree of confidence in the signs of the coefficients.

The USDA (1978) reports figures on rent per acre of cropland and pasture. In 1975 these numbers varied for cropland from a high of \$70 an acre for irrigated land in Nebraska to \$14.9 for land in Texas. Pastureland varied from a high of \$26.70 per acre in Iowa to \$5.30 in Texas. A hectare of land (100 square meters) is 2.47 acres, and the highest rental value that

Table 5.4  
Prior means and standard errors for the GNP regressions<sup>a</sup>

	1958		1975	
	Mean	Standard error	Mean	Standard error
CAPITAL	500	500	500	500
LABOR 1	7,100	3,550	13,000	6,500
LABOR 2	1,700	850	5,200	2,600
LABOR 3	170	85	520	260
LAND 1	3.3	6.7	10	20
LAND 2	1.7	3.3	5	10
LAND 3	33.3	8.3	100	25
LAND 4	16.7	8.3	50	25
COAL	1.0	.1	1.0	.1
MINERALS	1.0	.1	1.0	.1
OIL	1.0	.1	1.0	.1

a. All variables are measured in thousands—dollars, workers or hectares—except CAPITAL, which is measured in millions of dollars.

change as the prior is diluted or concentrated in one special way. Next it is assumed that the prior variance matrix  $\Omega$  lies in an interval of matrices  $\lambda_1 \Omega_0 \leq \Omega \leq \lambda_2 \Omega_0$ , for some values of the scalars  $\lambda_1$  and  $\lambda_2$ . Corresponding to this interval of prior covariance matrices is an ellipsoid of pooled (posterior) estimates. Extreme estimates of parameters of interest are then selected from the ellipsoid of possible estimates. If this interval of estimates is small, it is reported that the choice of estimates is insensitive to choice of  $\Omega$  in the interval  $\lambda_1 \Omega_0 \leq \Omega \leq \lambda_2 \Omega_0$ . If the interval of estimates is too large to be useful, then a smaller interval of  $\lambda$  values is selected. If the  $\lambda$  interval is so small already that it cannot credibly be narrowed, then it is reported that the point estimates are so sensitive to the choice of prior covariance matrix that the data cannot be used to draw useful inferences about this parameter. For further discussion of these methods, consult Leamer (1978, 1982) or Leamer and Leonard (1983).

### 5.3.1 Choice of Priors for the GNP Regression

Prior means and standard errors for the coefficients in the GNP regressions are reported in table 5.4. These are meant to be guesses about the annual earnings of each of the factors. An explanation of the choice of these numbers follows.

The coefficient of the CAPITAL variable in principle is the real interest rate plus the rate of depreciation. The capital stock figures we are using assume a 15-year life and a 13.3% rate of depreciation. If to this is added

could be expected for either LAND 3 or LAND 4 is  $\$70 \times (2.47) = \$173$ . Keeping in mind that much of the land is not high quality Nebraska farmland, my expectation is that LAND 3 (California) on the average would yield around \$100 per hectare, but that LAND 4 (Michigan) is much less productive, with rental value of \$50 per hectare. The standard errors for these numbers mean that, roughly speaking, I think that there is a 50% chance that the LAND 3 coefficient is between 75 and 125 and that there is a 50% chance that the LAND 4 coefficient is between 25 and 75. For no especially compelling reasons, the prior means for LAND 1 and LAND 2 are set to 10 and 5, respectively, with relatively large standard errors to reflect my relative ignorance of these coefficients. I was unable to find comparable rental values for 1958 U.S. farmland, but indexes of asset values increased by a factor of approximately three over the period from 1958 to 1975. The 1958 prior means and standard deviations therefore are one-third of the 1975 means and standard deviations.

Every dollar in coal, minerals, and oil produced by the economy can be expected to add approximately \$1 to GNP, if these resources are freely traded and if the production of these items does not use other inputs. There are several reasons why the coefficients of these variables might not be exactly equal to one. First, these are not homogeneous products, and the prices used to convert physical to value units may not be completely appropriate. Second, if international transportation costs are significant and cannot be avoided by the exchange of finished products (that is, factor prices are not equalized) then exporters will have access to cheaper resources on the margin than importers (who pay the transportation costs), and the marginal value of the resource will be lower in the exporting countries than the importing country. Third, a dollar in resource extracted does not contribute exactly a dollar to GNP because extraction requires labor and capital inputs, which are consequently double counted in the GNP equation. Nonetheless, the prior means are selected as if these natural resources were costlessly extracted and freely traded, and are set equal to one. The standard errors on these numbers are selected to reflect considerable confidence in these prior means.

### 5.3.2 Choice of Prior Distribution for the Net Export Equations

The choice of a prior distribution for the net export equations is much more difficult than the choice for the GNP equations because the coefficients do not have observable counterparts and because they are in theory

a consequence of a rather complex computation. The GNP coefficients do determine the rough order of magnitude of the net export coefficients, since it is fair to expect a resource to contribute less to net exports than to GNP. This kind of thinking leads to the prior means for the net export equations reported in Tables 5.5 and 5.6. Prior standard errors are set equal to these prior means to represent a fairly high degree of uncertainty. In order to select these prior means it is necessary first to consider the effect that nontraded goods have on the derivatives of net exports with respect to factor endowments. As discussed in section 1.3.3, the even model with nontraded goods and factor price equalization can be written either as  $T = A^{-1}(V - sV_w^*)$  or as  $T = A^{-1}(V^* - sV_w^*)$ , where the asterisks indicate resources devoted to the traded goods sector. The second form is the easier one to think about, since the consumption effect is  $sA^{-1}V_w^* = A^{-1}V_w^*Y/Y_w$ , which is GNP times the share of the commodity in final output. The derivative of this consumption effect with respect to a factor endowment is just the consumption share times the effect of the resource on GNP, the latter predicted by the prior means in table 5.4. To find the prior means for the net export equations, the consumption shares have all been set to 1/20, on the assumption that roughly half of final output is nontraded goods and services and that each of the 10 aggregates comprise roughly the same proportion of final output. If there were no production effect, then the expected effect of a resource on net exports would be the negative of its effect on GNP divided by 20. This number is used as the prior mean in cases when the production effect of a resource is expected to be small. The means for the other coefficients are set equal to each other and large enough that the coefficients sum to zero across equations, as implied by balanced trade. For example, \$1 of COAL is expected to contribute \$1 to GNP. Half of that dollar in earnings is expected to be spent on nontraded goods, and of the remaining half, one-tenth is expected to be spent on each of the 10 traded aggregates. Thus the pure consumption effect is  $1 \times .5 \times .1 = .05$ . The negative of this number serves as the prior mean for all of the COAL coefficients except the coefficient in the MAT equation, on the expectation that COAL does not contribute to the production of any product except coal itself, which is included in the MAT aggregate. The prior mean of .45 in the MAT equation assures that the means add to one across equations. To put this differently, the production effect of COAL on MAT net exports is .5 offset by a consumption effect of .05, the latter a consequence of the



a. Standard errors equal to means.

	PETRO	MAT	FOR	TROP	ANL	CER	LAB	CAP	MACH	CHEM
CAPITAL	-.25	-.25	-.25	-.25	-.25	-.25	-.25	-.25	-.25	-.25
LABOR 1	-.650	-.650	-.650	-.650	-.650	-.650	-.650	-.650	-.650	-.650
LABOR 2	-.260	-.260	-.26	-.26	-.26	-.26	-.26	-.26	-.26	-.26
LABOR 3	-.5	-.5	-.5	-.5	-.5	-.5	-.5	-.5	-.5	-.5
LAND 1	-.25	-.25	-.25	-.25	-.25	-.25	-.25	-.25	-.25	-.25
LAND 2	-.5	-.5	-.5	-.5	-.5	-.5	-.5	-.5	-.5	-.5
LAND 3	-.5	-.5	-.5	-.5	-.5	-.5	-.5	-.5	-.5	-.5
LAND 4	-.25	-.25	-.25	-.25	-.25	-.25	-.25	-.25	-.25	-.25
COAL	-.05	-.05	-.05	-.05	-.05	-.05	-.05	-.05	-.05	-.05
MINERALS	-.05	.45	-.05	-.05	-.05	-.05	-.05	-.05	-.05	-.05
OIL	.20	.20	-.05	-.05	-.05	-.05	-.05	-.05	-.05	-.05

Table 5.6  
Prior means, net export equations: 1975\*

a. Standard errors equal to means.

	PETRO	MAT	FOR	TROP	ANL	CER	LAB	CAP	MACH	CHEM
CAPITAL	-.25	-.25	-.25	-.25	-.25	-.25	-.25	-.25	-.25	-.25
LABOR 1	-.355	-.355	-.355	-.355	-.355	-.355	-.355	-.355	-.355	-.355
LABOR 2	-.85	-.85	-.85	-.85	-.85	-.85	-.85	-.85	-.85	-.85
LABOR 3	-.17	-.17	-.17	-.17	-.17	-.17	-.17	-.17	-.17	-.17
LAND 1	-.09	-.09	-.09	-.09	-.09	-.09	-.09	-.09	-.09	-.09
LAND 2	-.17	-.17	-.17	-.17	-.17	-.17	-.17	-.17	-.17	-.17
LAND 3	-.84	-.84	-.84	-.84	-.84	-.84	-.84	-.84	-.84	-.84
LAND 4	-.05	-.05	-.05	-.05	-.05	-.05	-.05	-.05	-.05	-.05
COAL	-.05	-.05	-.05	-.05	-.05	-.05	-.05	-.05	-.05	-.05
MINERALS	-.05	.45	-.05	-.05	-.05	-.05	-.05	-.05	-.05	-.05
OIL	.20	.20	-.05	-.05	-.05	-.05	-.05	-.05	-.05	-.05

Table 5.5  
Prior means, net export equations: 1958\*

Table 5.7  
Bayes estimates, MACHINERY equation, 1975

$\sigma_1$	0	.5	1	2	$\infty$
$R^2$	-2.2	.79	.83	.86	.91
Confidence <sup>a</sup>	1.0	.999	.999	.98	0
CAPITAL	37.5	24.5	29.1	32.3	37.2
LABOR 1	2,600	-821	-1,177	-1,008	-3,519
LABOR 2	65	77	78	37	142
LABOR 3	-26	-13	8	25	65
LAND 1	-5	-6	-7	-1.1	1.9
LAND 2	-25	-27	-30	-37	5.2
LAND 3	-5.0	-7.2	-11.7	-21.1	-30.2
LAND 4	-2.5	-4.5	-7.5	-10.6	5.5
COAL	-.05	-.05	-.05	.00	2.1
MINERALS	-.05	-.06	-.08	-.12	-3.0
OIL	-.05	-.10	-.17	-.25	-.49

a. One minus the significance level at which the hypothesis that the coefficients equal their indicated values is rejected.

increased GNP induced by increases in COAL resources. Similar calculations apply to all of the other coefficients.

From tables 5.5 and 5.6 it will be observed that I expect LABOR 1 to contribute to comparative advantage in the last two manufactured aggregates—the most skill intensive products. LABOR 2 contributes to all the manufactured commodities plus all the agricultural commodities. LABOR 3, the illiterate work force, has a positive effect on the labor intensive manufactured product and on the agricultural products. Tropical land (LAND 1) is expected to contribute to the tropical agricultural products and to cereals (including rice). The dry land (LAND 2) is associated with the agricultural products, but not forest products or cereals. COAL, MINERALS, and OIL contribute positively to their corresponding trade categories. (OIL includes natural gas, which is included in MAT.)

### 5.3.3 Example of the Use of Prior Information

Estimates of the 1975 MACHINERY regression using the prior information described in table 5.6 are recorded in table 5.7. The first row of this table contains values for  $\sigma_1$ , the scalar that multiplies all the prior standard errors. If  $\sigma_1 = \infty$ , the prior standard errors are infinite and the estimates are based on the data alone. These numbers are recorded in the last column of the table. If  $\sigma_1 = 0$ , the prior standard errors are all set to

zero and the coefficients are constrained to equal their prior means. These numbers are recorded in the first column. The column headed by  $\sigma_1 = 1$  contains the compromise estimates built on the prior defined in table 5.6. The value  $\sigma_1 = 2$  uses a prior that is twice as diffuse, and the value  $\sigma_1 = .5$  uses a prior that is twice as dogmatic.

There is little difference between the least-squares estimate of the CAPITAL coefficient (37.2) and the prior estimate (37.5). The pooled estimate, with  $\sigma_1 = 1$  is similar, though it is smaller than both. Similarly, there is relatively little disagreement over the estimates of the coefficients of LABOR 2, LAND 3, MINERALS, and OIL, in the sense that the signs of the prior and data estimates conform. For the other coefficients there is apparent conflict between the prior and the data. Highly skilled workers (LABOR 1) are estimated by the data to have a negative effect on MACHINERY net exports. This negative sign is retained even if  $\sigma_1$  is as small as .5. The LABOR 3 coefficient is estimated with the data alone to be positive, though this turns negative if the prior is sharpened and  $\sigma_1$  is set to .5. The LAND 1, LAND 2, LAND 4, and COAL coefficients, which are positive at the least-squares estimate, all turn negative even with the fairly small dose of prior information with  $\sigma_1 = 2$ .

A reasonable interval of values for  $\sigma_1$  is  $.5 < \sigma_1 < 2$ . This allows the prior to be either twice as informed or twice as uninformed. The estimates corresponding to this interval generally conform in sign to their prior values and any a priori implausible least-squares estimates are eliminated. An example is the least-squares estimate of the MINERALS coefficient, which suggests that each extra dollar in MINERALS output generates \$3 of MACHINERY imports. A little bit of prior information is enough to produce a sensibly small coefficient. The coefficient for LABOR 3 changes sign within the  $.5 < \sigma_1 < 2$  interval, and in this case we may conclude that the data and prior are jointly inconclusive about the sign of the effect of illiterate workers on the net exports of MACHINERY. The major finding suggested by table 5.7 is that LABOR 1 has a negative effect, contrary to the sign of the prior, regardless of the value of  $\sigma_1$  in the interval  $.5 < \sigma_1 < 2$ . The surprising influence of human capital on comparative advantage is discussed further in the next chapter.

This sensitivity analysis considers the effect of simultaneous, proportional increases or reductions of all elements in the prior covariance matrix; the next form of sensitivity analysis allows all elements of the prior variance matrix to change independently. The intervals of estimates

**Table 5.8**  
Estimates and bounds,<sup>a</sup> MACHINERY equation, 1975<sup>b</sup>

	Prior estimate	Least squares	Bayes estimate <sup>c</sup>	Lower bound <sup>b</sup>	Upper bound <sup>b</sup>
CAPITAL	37.5(37.5)	37.2(5.5)	29.1(3.2)	18.1	38.9
LABOR 1	2,600(2,600)	-3,519(1,460)	-1,177(582)	-3,250	1,420
LABOR 2	65(65)	142(82)	78(45)	-117	231
LABOR 3	-26(26)	65.1(45)	8(20)	-82.8	95.1
LAND 1	-5(5)	1.9(3.4)	-7(5)	-3.8	2.1
LAND 2	-25(25)	5.2(7.7)	-3(2)	-1.9	1.3
LAND 3	-5(5)	-30(14)	-12(5)	-37	8
LAND 4	-2.5(2.5)	5.5(6.9)	-7.5(1.9)	-15	2
COAL	-.05(.05)	2.1(4.3)	-.05(.05)	-.33	.28
MINERALS	-.05(.05)	-3.0(1.1)	-.08(.05)	-.40	.22
OIL	-.05(.05)	-.49(.11)	-.17(.04)	-.38	.02

a. This is the Bayes estimate implied by the prior covariance matrix defined in table 5.6. The same estimates may be found in table 5.7 in the column headed  $\sigma_1 = 1$ .

b. Standard errors in parentheses.  
c. These bounds let the prior covariance matrix lie between  $\Omega_0/4$  and  $\Omega_0$ , where  $\Omega_0$  is the covariance matrix defined in table 5.6. If the interval of prior variance is made smaller,  $\Omega_0/(1.5)^2$  to  $1.5^2 \Omega_0$ , the LAND 4 interval becomes -12 to -3 and the OIL interval becomes -.29 to -.06. Except for these and CAPITAL, the sensitivity intervals overlap zero.

implied by intervals of prior covariance matrices  $.25\Omega_0 \leq \Omega \leq 4\Omega_0$  are reported in table 5.8. In this interval, only the CAPITAL coefficient takes on a single sign. If one is unable or unwilling to define the prior more sharply, then it must be concluded that there is little to be learned from these data about the determinants of MACHINERY net exports. But if the interval of prior covariance matrices is narrowed to  $\Omega_0/(1.5)^2 \leq \Omega \leq (1.5)^2 \Omega_0$ , then the LAND 4 coefficient and the OIL coefficient are necessarily negative.

#### 5.4 Beta Coefficients as Measures of Significance

Beta coefficients are used in chapter 6 to rank variables. A beta coefficient is the estimated regression parameter after all variables have been standardized to have unit variance. If  $b_i$  is the ordinary least-squares estimate of the  $i$ th coefficient, then the corresponding beta coefficient is  $b_i \text{sd}(x_i) / \text{sd}(y)$ , where  $\text{sd}(x_i)$  is the standard deviation of variable  $i$  and  $\text{sd}(y)$  is the standard deviation of the dependent variable. A beta coefficient indicates the number of standard deviation changes in  $y$  induced by a one standard deviation change in  $x_i$ .

In applied econometric studies it is much more common to rank

variables by their  $t$ -values rather than their beta values. A large  $t$ -value does indicate four things:

1. If all the standard assumptions hold, the sign of the coefficient is determined with high confidence by the data. Formally speaking, a  $t$  of, say, 2.2 or more, with degrees of freedom in excess of 10, indicates that there is less than a 2.5% chance that the coefficient has a sign opposite its estimate, assuming the researcher is not otherwise informed (that is, the prior is diffuse).
  2. The estimated coefficient is insensitive to the omission of variables. Leamer (1975) has shown that if you omit a variable, then there will be no sign changes to any coefficient with a  $t$ -value higher than the coefficient of the omitted variable.
  3. The estimated coefficient is insensitive to the reweighting of observations. Leamer (1983) demonstrates that the sign of a coefficient is insensitive to the doubling of weights on selected observations if  $|t| > ((n-k)/8)^{1/2}$ , where  $n-k$  is the degrees of freedom.
  4. If all of the standard assumptions hold, then the data cast doubt on the hypothesis that the coefficient is exactly zero. A weighted likelihood ratio (or Bayes factor) against the hypothesis that the coefficient is zero is  $(1+t^2/(n-k))^{n/2} n^{1/2}$  (see Leamer, 1978, p. 114.)
- Although there are settings in which it is appropriate to identify the coefficients that are (1) most likely to be positive, (2) most insensitive to selection of variables, (3) most insensitive to the reweighting of observations, or (4) most likely not to be exactly zero, our desire here is to select the most important variables. It is perfectly possible for a coefficient to have a large  $t$ -value but to take on an infinitesimal value, so small that it surely is insignificant in any sense but statistical.

If it is wished to select the most important variables, then the purposes for which a model is being constructed have to be identified. Forecasting and control are the usual decision problems, neither of which is apparently pertinent to the present study. What is sought is a simple model that offers a reasonably complete "explanation." This rather vague objective can be given many interpretations. Here I imagine the problem is to "predict" a randomly selected country's net exports (or GNP) given its endowments. The best "prediction" is offered by the least-squares regression with all the endowment variables included. The unimportant endowment variables are those that can be expected not to have much impact

on this prediction. Suppose an estimated trade equation were  $\hat{T} = a + b_1 V_1 + b_2 V_2$ . If  $V_2$  were not observed, then a prediction of  $T$  would be  $\bar{T} = a + b_1 \bar{V}_1 + b_2 \bar{V}_2$ , where  $\bar{V}_2$  is the mean value of  $V_2$ . The expected squared prediction error due to not observing  $V_2$  is then  $E(\hat{T} - \bar{T})^2 = E(b_2 V_2 - b_2 \bar{V}_2)^2 = b_2^2 \text{var}(V_2)$ , which is just the squared beta value times  $\text{var}(T)$ . Therefore a large beta value indicates that the variable ought to be used for prediction. This prediction does not use information about the correlation between  $V_1$  and  $V_2$ . Given suitable assumptions, the optimal prediction of  $V_2$  given  $V_1$  is formed by regressing  $V_2$  on  $V_1$ :  $\hat{V}_2 = r_1 + r_2 V_1$ . Then the optimal predictor of  $T$  based on  $V_1$  alone is  $\hat{T}^* = a + b_1 V_1 + b_1(r_1 + r_2 V_1)$ , and the expected squared increment to prediction error is  $b_2^2 \text{var}(V_2|V_1)$ , which uses the conditional variance. This criterion as shown in Leamer (1978) ranks variables the same as  $t$ -statistics, and the choice between beta values and  $t$ -statistics as indicators of importance comes down to the way in which omitted variables are predicted. I prefer beta values because the theory associated with the model  $\hat{T}^*$  is complex:  $V_1$  has a direct and an indirect effect on  $T$ . The model  $\bar{T}$  is simpler since it includes only the direct effect. For further discussion, see Leamer (1978, chapter 6).