KONSTANTINOS T. VELENTZAS

RIDGE REGRESSION

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

Multiple regression analysis has become one of the most widely used econometrical tools for analyzing economic data. The standard approach in regression analysis is to use a sample of data to compute an estimate of the proposed relationship, and in addition, to evaluate the results of estimation using statistics such as t, F and the coefficient of determination R^2.

The general multiple regression model is written:

$$\underset{TX1}{y} = \underset{TXP}{X} \; \underset{PX1}{\beta} + \underset{TX1}{u}$$

where X is a TXP matrix of T observatios on P explanatory variables, y is a TX1 vector of observations on the dependent variable, β is a PX1 vector of unknown parameters, and u is a TX1 vector of unknown disturbances. The usual method for estimating β is the method of least squares which involves minimizing the sum of squares of the residuals under certain assumptions, the method of least squares has some very attractive statistical properties which has made it one of the most powerful methods of regression analysis. The fundamental assumptions in the general multiple linear regression are that: $E(u) = 0$, $E(uu') = \sigma^2 I$, the TXP matrix X is nonstochastic, and the rank of X is P (number of columns in X), and P is less than T, the number of observations. All these assumptions are icrucial for the estimation process. Therefore, one of the basic assumptons of the general linear model is:

$$\text{rank}(X) = P \quad \text{and} \quad P < T$$

This assumption states that no column of the X matrix can be written as a linear combination of other columns of the matrix, so that these columns are linearly independent vectors; that is, there is no exact linear relationship among the X variables. In other words, there is no perfect multicollinearity. In matrix notation this is equivalent to saying that there exists no vector α such that:

$$\alpha' \mathbf{x} = 0$$

where α' is a lxP row vector and $\mathbf{x}$ is Pxl column vector. The reason for this assumption is that the least squares estimator $\hat{\beta} = (X'X)^{-1} X'y$ requires the inversion of $(X'X)$, and under this assumption it follows that $(X'X)$ is nonsingular, so it can be inverted to obtain $\hat{\beta}$.

Unfortunately, in most economic applications it is often found that the rank condition is "relaxed". In empirical econometrics, the more typical situation is not one of perfect multicollinearity, but rather one of a multicollinearity problem. In this case $(X'X)$ is not singular, but is close to singular. We meet this problem when the rank assumption isonly just satisfied, that is when some or all of the explanatory variables are highly, but not perfectly collinear. It is recognized that in this situation, i.e., the determinant of $(X'X)$ is close to zero, a less extreme, but still serious problem arises. It is also known that the problem of multicollinearity is one of the most significant and difficult problems in applied econometrics because when multicollinearity is present in a set of explanatory variables, the least squares estimates of the individual regression coefficients tend to be unstable and can lead to erroneous inferences. Therefore, various problems arise in empirical econometrics when the rank condition is only just satisfied. The question is what should be done when we are sure that a serious problem of multicollinearity exists. We know that several possible methods are suggested for this problem. These include (Maddala, 1977):

1. Dropping one or more variables
2. Getting more or new data
3. Using prior information
4. Using principal components of the explanatory variables
5. Using ratios or first differences
6. Ridge regression

The purpose of this study on the one hand is to demonstrate the theory and the logic of the ridge regression method of estimation, and on the other hand to apply the technique to a standard estimation problem in economic models.

The problem selected is the estimation of an import demand function and a number of specifications are examined. These specifications differ by the amount of multicollinearity.

Chapter Two develops the theoretical side of ridge regression. It begins with the problems that arise when the multicollinearity problem exists. In addition, the ridge estimator is derived and its optimality is shown. This chapter concludes with a discussion of two methods of choosing the ridge parameter.

Chapter Three discusses the critical analyses of ridge regression that have been developed by statisticians outside the classical least squares framework. Specifically, the decision theory of biased estimators and Bayesian statistical inference are included.

Chapter Four presents the ridge regression method in practice. The estimating power of ridge regression is also compared to the estimating power of least squares method in the case of the estimation of import demand functions.

Chapter Five concludes the study with a general summary. Suggestions for further research are also included in this chapter.

CHAPTER II

RIDGE REGRESSION

The objective of this chapter is to introduce and survey the approach and technique of the new method of estimation: Ridge regression, which was introduced by the chemical engineer Hoerl.

This new method of estimation is called ridge regression because the basis of mathematics is similar to the method of ridge analysis that Hoerl used earlier (1959) to describe the behavior of second-order response surfaces.

The first section of this chapter discusses the ordinary least squares estimator and the multicollinearity problem. The deviation of the ridge estimator and its optimality are presented in the next two sections. Methods for choosing the ridge parameter are included in the final section.

Ordinary Least Squares and the Multicollinearity Problem

In the following discussion we consider the general multiple regression model:

$$y = X\beta + u$$

where:

where:	y	is a (Txl) vector of observations
	X	is a (TxP) matrix of T observations on
	P	explanatory variables
	β	is a (Pxl) vector of unknown regression coefficients
	u	is a (Txl) vector of disturbances.

It is assumed that $E(u) = 0$ and $E(uu') = \sigma^2 I$.

We know that by the method of least squares, thepoint estimate $\hat{\beta}$ of the vector β that minimizes the sum of squared residuals (e'e) is:

$$\hat{\beta} = (X'X)^{-1} X'y$$

According to the Gauss-Markov theorem, the least squares estimator $\hat{\beta}$ is linear, unbiased and has minimum variance in the class of unbiased linear estimators. That is, the variance matrix of any other linear unbiased estimator exceeds the variance matrix of the least squares estimator by a positive semidefinite matrix.

There is no quarantee that the variance of the least squares estimator will be small. The variance of the estimator can be shown to be:

$$V(\hat{\beta}) = E(\hat{\beta}-\beta)(\hat{\beta}-\beta)'$$
$$= \sigma^2 AA'$$

where $A \equiv (X'X)^{-1}X'$. This is:

$$V(\hat{\beta}) = \sigma^2(X'X)^{-1}$$

and shows that the magnitudes of the variances of the least squares estimators depend on the X matrix.

A more useful approach to determining the variance is to partition the X matrix as follows:

$$X = [x_1 \ X_2]$$

where x_1 is any column, and X_2 consists of all the other columns. By the theorem on partitioned matrices (Goldberger, 1964) it can be shown that:

$$(X'X)^{-1} = \begin{bmatrix} x_1'x_1 & x_1'X_2 \\ X_2'x_1 & X_2'X_2 \end{bmatrix}^{-1} = \begin{bmatrix} G & -GH' \\ -HG & D^{-1}+HGH' \end{bmatrix}$$

where $G \equiv [x_1'x_1 - x'_1X_2H]^{-1}$, $D \equiv (X_2'X_2)$

$$H \equiv (X_2'X_2)^{-1}X_2'x_1$$

This latter term H is the ordinary least squares estimator of the regression coefficients in

$$x_1 = X_2b + v$$

and G is the estimated residual sum of squares from this regression. This can be written in terms of R^2 from this regression as:

$$G = [nS_1^2 (1-R_1^2)]^{-1}$$

where R^2 is subscripted to distinguish this regression from the complete model. Hence, the first row of $(X'X)^{-1}$ is:

$$\left[\frac{1}{nS_1^2(1-R_1^2)} \quad \frac{-h_0}{nS_1^2(1-R_1^2)} \quad \frac{-h_2}{nS_1^2(1-R_1^2)} \quad \cdots \quad \frac{-h_p}{nS_1^2(1-R_1^2)} \right]$$

This shows that:

$$V(\hat{\beta}_1) = \frac{\sigma^2}{nS_1^2(1-R_1^2)}$$

and in general,

$$V(\hat{\beta}_p) = \frac{\sigma^2}{nS_p^2(1-R_p^2)}$$

The coefficient of determination obtained by regressing $\mathbf{x}_p$ against the other independent variables, which is written R_p^2, is often called the degree of multicollinearity in the matrix X. The above shows that the larger is R_p^2, the longer is $V(\hat{\beta}_p)$.

Multicollinearity also tends to produce least squares estimates, $\hat{\beta}_p$, that are too large in absolute value. It is easy to see this by examining the definition of variance:

$$\begin{aligned} V(\hat{\beta}_p) &= E[\hat{\beta}_p - E(\hat{\beta}_p)]^2 \\ &= E(\hat{\beta}_p)^2 + [E(\hat{\beta}_p)]^2 - 2E(\hat{\beta}_p)E(\hat{\beta}_p) \\ &= E(\hat{\beta}_p)^2 + \beta_p^2 - 2\beta_p^2 \\ &= E(\hat{\beta}_p)^2 - \beta_p^2 \end{aligned}$$

where $E(\hat{\beta}_p) = \beta_p$ follows from unbiasedness of the least squares estimators. Thus, the variance of $\hat{\beta}_p$ is the difference between the average length (squared) of $\hat{\beta}_p$ and the true length (square of β_p), as the $V(\hat{\beta}_p)$ is larger, so the longer will be $\hat{\beta}_p$ relative to β_p.

In sum, the greater the degree of multicollinearity in the X matrix, the longer will be the expected value of the vector of estimated coefficients.

Derivation of the Ridge Estimator

The central idea of ridge regression is to choose an estimator that is similar to ordinary least squares, but has a shorter length. The ridge estimator is defined as that estimator which minimizes the sum of the squared distances of the points from the estimated line subject to a constraint on the length of the estimating vector.

Let the ridge estimator be β^*, and the computed points be defined by:

$$y^* = X^*\beta$$

The residuals are:

$$e^* = y - y^*$$

and the ridge estimator minimizes the sum of the squared e^* values subject to a maximum length of β^*. Let this maximum be denoted by l. Thus, the task of finding β^* is a lagrangean problem. The lagrangean expression is:

$$L = e^{*\prime}e^* + k(\beta^{*\prime}\beta^* - l)$$

Differentiation of L with respect to β^* gives:

$$\frac{\partial L}{\partial \beta^*} = \frac{\partial e^*}{\partial \beta^*} e^* + \frac{\partial e^*}{\partial \beta^*} e^* + k\frac{\partial \beta^*}{\partial \beta^*} \beta^* + k\frac{\partial \beta^*}{\partial \beta^*} \beta^*$$

$$= 2\frac{\partial e^*}{\partial \beta^*} e^* + 2k\beta^*$$

we know that $e^* = y - y^*$

$$= y - X\beta^*$$

therefore, the derivative of e^* with respect to β^* is:

$$\frac{\partial e^*}{\partial \beta^*} = -X'$$

and $\dfrac{\partial L}{\partial \beta^*} = -2X'e^* + 2k\beta^*$

$$= -2X'(y-y^*) + 2k\beta^*$$

$$= -2X'y + 2X'X\beta^* + 2k\beta^*$$

Setting this equal to zero and solving for β^* yields:

$$\beta^* = (X'X + kI)^{-1} X'y$$

This is the ridge regression estimator for the vector of parameters β, which gives us the best fit to the data for any estimator of given length. If $k=0$, this implies $\beta^*=\hat{\beta}$ (we have least squares estimators), when $k \rightarrow \infty$, then $\beta^* \rightarrow 0$. This means as k increases, the ridge estimators get smaller and smaller in absolute size.

In addition to reducing the absolute size of the estimating vector, the ridge approach also produces an estimator with a smaller variance than ordinary least squares. By definition, the variance-covariance matrix of β^* is given by:

$$V(\beta^*) = E\{[\beta^* - E(\beta^*)][\beta^* - E(\beta^*)]'\}$$

Substituting $y = X\beta + u$ into the formula of ridge regression estimator yields:

$$\beta^* = (X'X+kI)^{-1} X'(X\beta+u)$$

$$= (X'X+kI)^{-1} X'X\beta+(X'X+kI)^{-1}X'u$$

taking expected values of both sides gives:

$$E(\beta^*) = E[(X'X+kI)^{-1} X'X\beta + (X'X+kI)^{-1} X'u]$$

$$= E[X'X+kI)^{-1} X'X\beta] + E[(X'X+kI)^{-1} X'u]$$

$$= (X'X+kI)^{-1}X'X\beta + (X'X+kI)^{-1}X'Eu$$

$$= (X'X+kI)^{-1}X'X\beta$$

Therefore:

$$\beta^* - E(\beta^*) = [X'X+kI)^{-1}X'X\beta + (X'X+kI)^{-1}X'u] - [(X'X+kI)^{-1}X'X\beta]$$

$$= (X'X+kI)^{-1}X'u$$

This can be substituted into the expression for the variance of β^* to yield:

$$
\begin{aligned}
V(\beta^*) &= E\{[(X'X+kI)^{-1}X'u]\,[(X'X+kI)^{-1}X'u]'\} \\
&= E[(X'X+kI)^{-1}X'uu'X\,(X'X+kI)^{-1}] \\
&= (X'X+kI)^{-1}X'E(uu')\,X\,(X'X+kI)^{-1} \\
&= (X'X+kI)^{-1}X'\sigma^2 IX(X'X+kI)^{-1}
\end{aligned}
$$

Therefore, the variance-covariance matrix of the ridge regression estimator is:

$$V(\beta^*) = \sigma^2(X'X+kI)^{-1}X'X(X'X+kI)^{-1}$$

If $k=0$ this implies that

$$V(\beta^*) = \sigma^2\,(X'X)^{-1} = V(\hat{\beta})$$

As $k \rightarrow \infty$, then $V(\beta^*) \rightarrow 0$. This means that the variance of the ridge regression estimator is a decreasing function of k.

The Optimality of the Ridge Estimator

The optimality of the least squares estimator stems from the fact that it has minimum variance in the class of linear, unbiased estimators. Because the estimator is unbiased, minimum variance implies minimum mean square error. That is, the variance $E[\hat{\beta}_p - E(\hat{\beta}_p)]^2$ is the same as the mean square error $E[\hat{\beta}_p - \beta_p]^2$.

However, the ridge estimator is biased. This is easily seen by noting that the ridge estimator is a linear transformation of the least squares estimator, as shown by the following argument:

$$
\begin{aligned}
\beta^* &= (X'X+kI)^{-1}\,X'y \\
&= (X'X+kI)^{-1}\,X'X\hat{\beta}
\end{aligned}
$$

by the definition of the least squares estimator $\hat{\beta}$. Adding and subtracting kI to the $X'X$ term which premultiplies $\hat{\beta}$ yields:

$$
\begin{aligned}
\beta^* &= (X'X+kI)^{-1}\,[(X'X+kI) - kI]\hat{\beta} \\
&= [I - k(X'X+kI)^{-1}]\hat{\beta}
\end{aligned}
$$

and the expected value of β^* is therefore given by:

$$E(\beta^*) = [I - k(X'X+kI)^{-1}]\,\beta$$

Hence, the bias is given by:

$$E(\beta^*) - \beta = -k(X'X+kI)^{-1}\,\beta$$

when dealing with a biased estimator, such as β^*, the appropriate optimality concept is mean square error. In order to examine the mean square error of β^* it is useful to reparametrize the regression model into canonical form as follows:

$$y = X\beta + u = Z\Theta + u$$

where

$$\begin{aligned} Z &= PX \\ \Theta &= P'\beta \end{aligned}$$

and

$$\begin{aligned} P'P &= PP' = I \\ P' &= P^{-1} \end{aligned}$$

The matrix P is the matrix of normalized characteristic vectors of $X'X$ such that:

$$Z'Z = P'X'XP = D = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_p)$$

where $\lambda_1, \lambda_2, \ldots, \lambda_p$ are the P characteristic roots of $X'X$.

The least squares estimator of the model in canonical form is:

$$\hat{\Theta} = (Z'Z)^{-1}\,Z'y$$

and this is a linear transformation of $\hat{\beta}$ because:

$$\begin{aligned} \hat{\Theta} &= (Z'Z)^{-1}Z'y = (P'X'XP)^{-1}P'X'y \\ &= (P^{-1}X'XP)^{-1}P'X'y = P^{-1}(X'X)^{-1}PP'X'y \\ &= P'\,\hat{\beta} \end{aligned}$$

Similarly, the ridge estimators bear the same relationship

$$\Theta^* = P'\,\beta^*$$

The mean square errors of $\hat{\Theta}$ and $\hat{\beta}$, and of Θ^* and β^* are equal:

$$\begin{aligned} MSE(\Theta^*) &= E[(\Theta^*-\Theta)'(\Theta^*-\Theta)] \\ &= E[(\Theta^*-\Theta)'P'P(\Theta^*-\Theta)] \\ &= E[(P\Theta^*-P\Theta)'(P\Theta^*-P\Theta)] \\ &= E[(\beta^*-\beta)'(\beta^*-\beta)] \\ &= MSE(\beta^*) \end{aligned}$$

Thus, the comparison of $MSE(\beta^*)$ with $MSE(\hat{\beta})$ is the same as the comparison of $MSE(\Theta^*)$ with $MSE(\hat{\Theta})$. By definition, the mean square error can be decomposed into variance and bias components:

$$\begin{aligned} MSE(\Theta^*) &= E[(\Theta^*-E\Theta^*)(\Theta^*-E\Theta^*)'] + E[(\Theta-E\Theta^*)(\Theta-E\Theta^*)'] \\ &= V(\Theta^*) + E[(\Theta-E\Theta^*)(\Theta-E\Theta^*)'] \end{aligned}$$

Using the results for variance and bias derived previously we have:

$$\begin{aligned} MSE(\Theta^*) &= \sigma^2 [(Z'Z+kI)^{-1} Z'Z (Z'Z+kI)^{-1}] + \\ &+ [k(Z'Z+kI)^{-1} \Theta . k (Z'Z+kI)^{-1} \Theta'] \\ &= (Z'Z+kI)^{-1} (\sigma^2 Z'Z+k\Theta\Theta') (Z'Z+kI)^{-1} \\ &= (D+kI)^{-1} (\sigma^2 D + k\Theta\Theta') (D+kI)^{-1} \end{aligned}$$

The trace of this $MSE(\Theta^*)$ is the sum of the mean square errors of the individual estimators:

$$\begin{aligned} tr\ MSE(\Theta^*) &= tr[(D+kI)^{-1} (\sigma^2 D+k\Theta\Theta') (D+kI)^{-1}] \\ &= \sum_{p=1}^{p} \frac{1}{(\lambda_p+k)^2} (\sigma^2\lambda_p + k\Theta_p{}^2) \end{aligned}$$

Differentiation with respect to k yields:

$$\frac{\partial tr\ MSE(\Theta^*)}{\partial k} = 2 \sum_{p=1}^{p} \frac{\lambda_p (k\Theta_p{}^2-\sigma^2)}{(\lambda_p+k)^3}$$

This derivative will be negative if $(k\Theta_p^2 - \sigma^2) < 0$ for each p, since k is always nonnegative, and since the characteristic roots of X'X are positive since the matrix is positive definite. Thus, if k is chosen between zero and σ^2/Θ^2_{max}, the derivative will be negative and the ridge mean square error will be less than the mean square of the ordinary least squares estimator $(k=0)$. The result arises because the reduction in variance exceeds the increase in bias. The contribution of bias is given by:

$$B(k) = \sum_{p=1}^{p} \frac{k^2\Theta_p^2}{(\lambda_p+k)^2} \text{ and } \frac{dB(k)}{dk} = 2k \sum_{p=1}^{p} \frac{\Theta_p^2\lambda_p}{(\lambda_p+k)^3} > 0$$

and the contribution of variance is given by:

$$\sum_{p=1}^{p} \frac{\lambda_p\sigma^2}{(\lambda_p+k)^2}$$

and the latter decreases as k increases.

Methods for Choosing the Ridge Parameter

Although there always exists a positive value of k such that the ridge regression estimator has a smaller mean square error than the least squares estimator, the best method for selecting a particular value of k is not obvious. The question is how to choose the value of the unknown ridge coefficient $k>0$ and consequently, a unique β^* without using information other than the sample information.

There are a number of alternative suggestions which have been proposed for selecting the particular value of k. In this section the methods of choosing k that are easy to compute and used in practice by researchers are discussed.

The earliest method for choosing the unknown coefficient k is the graphic technique which Hoerl and Kennard (1970) have suggested based on the "Ridge Trace".

The ridge trace is a simple graph of the values of the ridge regression estimators on the vertical axis plotted against the small corresponding values of k in the interval zero to one. The trace includes one curve for each coefficient. The purpose of the ridge trace is to give the analyst a picture of the effect of the multicollinearity and to assist further to choose the lowest possible value of k for which the estimate coefficients have stabilized. By stable we mean that the coefficients are not sensitive to small changes in the estimation data. The value of k at which the coefficients are stable yields the desired set of coefficients.

Criteria for choosing k have been outlined by Hoerl and Kennard (1970). The criteria are reproduced as follows:

1. At a certain value of k the system will stablize and have the general characteristics of an orthogonal system.
2. Coefficients will not have unreasonable absolute values with respect to the factors for which they represent rates of change.
3. Coefficients with apparently incorrect signs at $k=0$ will have changed to have the proper sign.
4. The residual sum of squares will not have been inflated to an unreasonable value. It will not be large relative to what would be a reasonable variance for the process generating the data. (p. 65).

An alternative procedure uses the fact that in theory $k=\hat{\sigma}^2/\hat{\Theta}_{max}$. In this procedure, an initial ordinary least squares regression is estimated on the transformed model $y = Z\Theta + u$ and the maximum $\hat{\Theta}$ is used as an initial value of k. Using this value, an initial ridge regression is estimated and this yields a second estimate Θ_{max} which in turn yields another estimate of k. This procedure is repeated until k converges. Finally, given the selected estimate of the Θ vector, the β vector is estimated from:

$$\beta^* = P'\Theta^*$$

It should be noted that these procedures make k a function of the sample data and, therefore, k becomes stochastic. The properties of the ridge estimator when k is stochastic are not well known, but are discussed in the next chapter.

CHAPTER III

CRITICISMS OF RIDGE REGRESSION

The purpose of this chapter is to survey the critical analyses of ridge regression that have been developed by statisticians outside the classical least squares framework. Two major approaches are considered, namely, the decision theory of biased estimators and Bayesian statistical inference.

This chapter is divided into five sections. The first section represents the decision theory approach to estimation. The minimax analysis of regression is discussed in the second section. A discussion of William G. Brown and Bruce R. Beattie's critique of the ridge regression method is included in section three. The two final sections discuss the Bayesian statistical inference, and Bayesian inference and ridge regression.

The Decision Theory Approach to Estimation

In econometrics we are concerned with the use of sample data to learn about the unknown economic parameters. Our interest is in finding good point estimates of economic parameters. Typically, a "best" estimator is defined to be one that predominates over any other when comparative criteria had been used.

It is well known that for comparison of unbiased estimators, the variance criterion is used and the estimator is chosen that has the smallest variance. When the estimator is biased, as in the case of the ridge estimator, a better comparison of the precision of estimators would be obtained by comparing their mean square errors.

Furthermore, the unbiasedness property plays a more important role in the theory of interval estimation than it does in point estimation. This is because confidence intervals are centered on an unbiased estimate and because the width of the interval is equal to the estimated square root of the variance of the estimator. The smaller variance for the biased estimator implies that it is a more stable estimator than is the unbiased estimator.

The justification given for ridge regression in the last chapter was that, for a range of values of k, the sum of the mean square errors of the ridge estimators of the regression coefficients is less than the sum of the mean square errors of the ordinary least squares estimators. That is:

$$\sum_{p=1}^{p} E(\beta_p^* - \beta_p)^2 < \sum_{p=1}^{p} E(\hat{\beta}_p - \beta_p)^2$$

This reduction is obtained because the variances of the ordinary least squares estimators exceeds the sum of the variances of the ridge estimators plus their biases squared.

The preceding justification weights the squared errors for both estimation techniques equally. More generally, the performance of an estimator might be evaluated using a weighted sum of squared error such as:

$$\sum_{p=1}^{p} W_p\,(\beta_p^* - \beta_p)^2$$

In decision theory, such a function is called a loss function and, in general, it would be written (Greenberg and Webster, 1983, p. 160):

$$L(\beta^*, \beta) = (\beta^* - \beta)'\,W\,(\beta^* - \beta)$$

where β^* is a vector of estimates of the parameters contained in the β vector, and W is a positive semidefinite matrix of weights. The particular example used here is the case of a quadratic loss function. The problem, in terms of decision theory, is to decide on an estimating formula that will minimize the loss function.

However, the loss function is stochastic, and thus the size of the loss associated with any estimator will vary depending on the sample data. Decision theorists, therefore, consider the expected value of the loss function, which is called the risk function:

$$R\,(\beta^*, \beta) = E[L(\beta^*, \beta)]$$

Our objective is to decide on an estimating formula β^* which minimizes, in some sense, the risk function.

Usually, it is not possible to find an estimator which globally minimizes the risk. Thus, what is undertaken is a comparison of existing estimators in terms of their risk. In cases where one estimator has the lowest risk for one range of parameter values, and another estimator has the lowest risk for the remaining range of parameter values, a frequently used criterion is the minimax criterion.

According to the minimax criterion, we find the maximum risk of each estimator and we choose that estimator which minimizes this maximum risk. An estimator β^* is minimax if

$$\max R(\beta^*, \beta) \leqslant \max R(\beta^*, \beta)$$

for all β^*.

We note that another criterion for choosing a "good" decision is the criterion which is based on the idea of admissibility. For this case we eliminate the "bad" decisions.

If we have the estimators β^* and $\beta^{*\prime}$, then β^* is said to dominate $\beta^{*\prime}$, when the following relationships hold (Maddala, 1977, p. 54):

$$R(\beta^*, \beta) \leqslant R(\beta^{*\prime}\, \beta) \qquad \text{for all } \beta$$

and

$$R(\beta^*, \beta) \leqslant R(\beta^{*\prime}\, \beta) \qquad \text{for some } \beta.$$

Minimax Analysis of Regression

It was noted in the previous section that when we drop the assumption of unbiasedness, we have a wider class of estimators which includes biased estimators. To analyze this class, decision theory is used which starts with the loss function.

Now to discuss the minimax analysis of regression, we assume again the standard regression model:

$$y = X\beta + u$$

where $E(u)=0$ and $E(uu')=\sigma^2 I$. We want to find a linear estimator β^* such that:

$$\beta^* = Wy$$

and from the assumptions above $E\beta^* = WX\beta$ since $Eu=0$. We know from the previous section that the risk function for a quadratic loss is defined as:

$$R(\beta^*, \beta) = E[(\beta^* - \beta)'(\beta^* - \beta)]$$

From this relationship we have (Greenberg and Webster, 1983, p. 166):

$$E[(\beta^* - \beta)'(\beta^* - \beta)]$$

$$= \{[(\beta^* - E\beta^*) - (\beta - E\beta^*)]'[(\beta^* - E\beta^*) - (\beta - E\beta^*)\}$$

$$= E[(\beta^* - E\beta^*)'(\beta^* - E\beta^*)] + [(E\beta^* - \beta)'(E\beta^* - \beta)]$$

Substituting $\beta^* = Wy$ and $E\beta^* = WX\beta$ yields:

$$E[(\beta^* - \beta)'(\beta^* - \beta)]$$

$$= E[(Wy - WX\beta)'(Wy - WX\beta)] + [(WX\beta - \beta)'(WX\beta - \beta)]$$

$$= E[(y - X\beta)'W'W(y - X\beta)] + [\beta'(WX - I)'(WX - I)\beta]$$

$$= \sigma^2 \operatorname{tr} WW' + \beta'(WX - I)'(WX - I)\beta.$$

Investigating the last expression shows that the first term $\sigma^2 \operatorname{tr} W'W$ is independent of β while the second is a positive semidefinite quadratic form in β and it does not involve β^*. If $\beta \to \infty$, then this term will grow without bound.

All the above implies that the risk of an estimator that is dependent of β will be affected by large values of β. When $WX = I$ then:

$$R(\beta^*, \beta) = E[(\beta^* - \beta)'(\beta^* - \beta)]$$

$$= \sigma^2 \operatorname{tr} WW'$$

This means that we have a linear estimator that is minimax in the class of linear estimators. When $W = (X'X)^{-1}X'$ then $WW' = (X'X)^{-1}$. This is the case of the least squares estimator and shows that it is minimax in the class of linear estimators.

Regarding the ridge regression estimator we can see that this estimator is dependent on β because k is defined as:

$$k = \frac{\sigma^2}{\Theta^2_{max}}$$

and Θ is a function of the coefficient vector β. Thus, for large values of at least one Θ_i the risk of the ridge estimator $R(\Theta^*,\Theta)$ will be greater than least squares estimator $R(\hat{\Theta},\hat{\Theta})$ because the latter does not depend on the Θ_i's. Therefore, choosing k without reference to the data will not yield a minimax estimator.

Hoerl, Kennard and Baldwin (1975) suggested for choosing **k** based on data:

$$k = \frac{P\hat{\sigma}^2}{\hat{\Theta}'\hat{\Theta}}$$

Thisted (1976) has shown that this formula is minimax if and only if:

$$\frac{\sum_{i=1}^{p} \lambda_i^{-2}}{\lambda_p^{-2}} \geqslant 2 + \frac{P}{2}, \qquad P \geqslant 3$$

where λ_p is the smallest root of $X'X$. To meet this condition λ_p must be large. For the case of the multicollinearity problem λ_p is small, so the condition of minimaxity is unlikely to be met. Therefore, people who believe the minimax criterion do not suggest the ridge regression method as a method to solve the multicollinearity problem. Other criteria and approaches such as admissibility might imply different conclusions. For example, those who use in their analysis the criterion of admissibility suggest that we can use the ridge regression method to solve multicollinearity because the ridge estimator dominates the least squares estimator in terms of mean square error.

Prior Information and Ridge Regression

Brown and Beattie (1975) suggest that the ridge estimator may be most appropriate in the case where the researcher has prior information about the signs of the unknown regression coefficients. Their statement is as follows:

> . . . ridge estimates can also be unreliable and misleading under certain conditions. To avoid erroneous conclusions

> from ridge regression, some prior knowledge about the true regression coefficients is helpful. A theorem on expected bias implies that ridge regression will give much better results for some economic models such as certain production function, than for others because of smaller expected bias ... on the other hand ridge estimation of other economic models, such as certain demand functions, could give very poor or misleading results . . . .(p. 21).

Brown and Beattie's assertions and suggestions are products of the following analysis. Consider again the formula of bias:

$$E(\beta^*) - \beta = -k(X'X+kI)^{-1}\beta \tag{3.1}$$

which was defined in thc last chapter, third section.

Now we consider the following regression model for the simple case of two standardized explanatory variables:

$$y = \beta_1 x_1 + \beta_2 x_2 + u$$

and the ridge regression normal equations are:

$$(x'x+kI)\beta^* = x'y$$

or

$$\begin{bmatrix} (1+k) & (r_{12}) \\ r_{12} & (1+k) \end{bmatrix} \begin{bmatrix} \beta_1^* \\ \beta_2^* \end{bmatrix} = \begin{bmatrix} r_{1y} \\ r_{2y} \end{bmatrix}$$

where r_{12} is the simple correlation coefficient between x_1 and x_2 and r_{jy} is the simple correlation coefficient between x_j and y, $j=1, 2$. Now the inverse of $(x'x+kI)$ is:

$$(x'x+kI)^{-1} = \begin{bmatrix} \dfrac{(1+k)}{(1+k)^2 - r^2_{12}} & \dfrac{-r_{12}}{(1+k)^2 - r^2_{12}} \\ \dfrac{-r_{12}}{(1+k)^2 - r^2_{12}} & \dfrac{(1+k)}{(1+k)^2 - r^2_{12}} \end{bmatrix}$$

and consequently:

$$(x'x+kI)^{-1}\beta = \begin{bmatrix} \frac{(1+k)}{(1+k)^2 - r^2_{12}} & \frac{-r_{12}}{(1+k)^2 - r^2_{12}} \\ \\ \frac{-r_{12}}{(1+k)^2 - r^2_{12}} & \frac{(1+k)}{(1+k)^2 - r^2_{12}} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \\ \beta_2 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{(1+k)\beta_1 - r_{12}\beta_2}{(1+k)^2 - r^2_{12}} \\ \\ \frac{-r_{12}\beta_1 + (1+k)\beta_2}{(1+k)^2 - r^2_{12}} \end{bmatrix}$$

according to the last expression for $\beta_1{}^*$ and $\beta_2{}^*$ equation (3.1) yields:

$$E(\beta_1{}^*) - \beta_1 = \frac{-k}{(1+k)^2 - r^2_{12}} \left[(1+k)\beta_1 - r_{12}\beta_2\right]$$

$$E(\beta_2{}^*) - \beta_2 = \frac{-k}{(1+k)^2 - r^2_{12}} \left[(1+k)\beta_2 - r_{12}\beta_1\right]$$

Investigation of the above equation implies:

1. When X_1 and X_2 are positively correlated the expected bias of $\beta_1{}^*$ and $\beta_2{}^*$ will be smallest if β_1 and β_2 have the same sign and they are also of about equal magnitude.
2. If β_1 and β_2 have opposite signs and $r_{12} > 0$ then the expected bias of $\beta_1{}^*$ and $\beta_2{}^*$ will be greatly increased.
3. If X_1 and X_2 are negatively correlated then the bias squared will be smallest when β_1 and β_2 differ in sign and they are about equal in absolute magnitude.

Also, Brown and Beattie (1975, p. 31) show that all of the above results can be generalized to the case of P standardized explanatory variables. They proved that the bias of the ridge estimate of the jth standardized coefficient can be expressed as (1975, p. 31):

$$E(\beta_j^*) - \beta_j = \frac{kc_{jj}}{|A^*|} \sum_{i=1}^{p} \beta^*_{ji}\, \beta_i$$

where $A^* = (x'x+kI)$, c_{jj} is a cofactor of A^*, $\beta_{ji}^* = -1.0$ if $i=j$ and $i \neq j$, β^*_{ji} denotes the ridge estimate of the coefficient for the ith variable, where x_j has been regressed on the P—1 remaining explanatory variables. Investigation of this general case gives the same result as previously discussed, i.e., the bias and mean square error will be smaller as most of the β's have the same sign and β_j is approximately equal to the average of the other P—1 explanatory variables. Thus, Brown and Beattie advocate the use of ridge regression only when the signs of β's are the same such as in the Cobb-Douglas production function, because the ridge regression for other cases yields a larger bias and thus poor results.

I would like to take this opportunity to emphasize that unbiasedness plays no important role in the theory of point estimation. Therefore, the bias by itself is not an important criterion for checking the results of the estimation. Also, the purpose of ridge regression is to reduce the high variance of the estimating coefficients by adding some bias. Its focal point is the mean square error and not the bias. Therefore, the question is: To what extent is the magnitude of the mean square error and not the magnitude of the bias. For solving multicollinearity the ridge regression method of estimation can be used in spite of their analysis because there always exists a value of $k > 0$ such that:

$$MSE(\beta^*) < V(\hat{\beta})$$

and the coefficients are stable.

Bayesian Statistical Inference

A more general approach to the incorporation of prior information in regression is to use Bayesian statistical inference. Bayesian inference is based on Bayes theorem which states that the posterior probabilities density function (pdf) is proportional to the prior probabilities density function times the likelihood function:

posterior (pdf) $\propto$ prior (pdf) x likelihood function

where $\propto$ sign denotes "proportional to". The prior probabilities density function incorporates all prior information, while the likelihood function incorporates all the sample information. Therefore, the Bayesian approach combines a prior distribution with sample information to form a posterior distribution. The mean of the posterior distribution of the parameter vector β gives the Bayes estimator $\bar{\bar{\beta}}$.

Applying the Bayesian approach to the problem of estimation of the linear regression model:

$$y = X\beta + u \tag{3.2}$$

the posterior distibution for β, which is conditional on σ, is proportional to the product of the prior distribution for β and the likelihood function. The sample information for the linear regression model is given by the data on the dependent and explanatory variables. The data are given by the matrix D defined as:

$$\underset{TX(PX1)}{D} = (\underset{TX1}{y} \mid \underset{TXP}{X})$$

It is well known that if x has normal distribution with mean μ and variance σ^2, or $x \sim N(\mu, \sigma^2)$, then the probability density function of x is defined as:

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right]$$

and by definition, the likelihood function is given by the following formula:

$$L(\mu, \sigma^2) = \prod_{i=1}^{T} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x_i-\mu)^2\right] =$$

$$= \frac{1}{(2\pi\sigma^2)^{T/2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{T}(x_i-\mu)^2\right]$$

To find the Bayes estimator for β, if we assume for the linear regression model (3.2) that $u \sim N(0, \sigma^2 I)$ and u_i are independent random variables, then the likelihood function for the sample value is:

$$L(\beta, \sigma / y) = \frac{1}{(2\pi\sigma^2)^{T/2}} \exp\left[-\frac{1}{2\sigma^2} u'u\right]$$

or (Judge, 1982, p. 228):

$$L(\beta, \sigma / y) \propto \frac{1}{\sigma^T} \exp\left[-\frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)\right]$$

The exponential term in the likelihood function can be written in terms of the ordinary least squares estimator $\hat{\beta}$ using the following argument. Expanding the quadratic form yields:

$$(y - X\beta)'(y - X\beta) = y'y - (\beta'X'y)' - \beta'X'y + \beta'X'X\beta$$

Substitute $X'X\hat{\beta}$ for $X'y$ (the normal equations of ordinary least squares) yields:

$$(y - X\beta)'(y - X\beta) = y'y - \hat{\beta}'X'X\beta - \beta'X'X\hat{\beta} + \beta'X'X\beta$$

Add the two terms $\hat{\beta}'X'X\hat{\beta} - \hat{\beta}'X'y$ and $\hat{\beta}'X'X\hat{\beta} - y'X\hat{\beta}$, both of which are zero by the normal equations. Combining terms we obtain:

$$(y - X\beta)'(y - X\beta) = (y - X\hat{\beta})'(y - X\hat{\beta}) + (\beta - \hat{\beta})'X'X(\beta - \hat{\beta})$$

but from ordinary least squares theory:

$$S^2 = (y - X\hat{\beta})'(y - X\hat{\beta}) / n$$

where $n = T - P$. Hence:

$$(y - X\beta)'(y - X\beta) = ns^2 + (\beta - \hat{\beta})'X'X(\beta - \hat{\beta})$$

and the likelihood function is:

$$L(\beta, \sigma / y) = \propto \frac{1}{\sigma^T} \exp\left\{-\frac{1}{2\sigma^2}\left[ns^2 + (\beta - \hat{\beta})'X'X(\beta - \hat{\beta})\right]\right\}$$

In order to obtain the posterior distribution we must specify the prior density function. If prior information about β can be described by a normal distribution with mean vector β_0, and covariance matrix v^2A^{-1} or $\beta \sim N(\beta_0, v^2A^{-1})$, then the prior density is (Greenberg and Webster, 1983, p. 194):

$$g(\beta,\sigma) \propto \frac{1}{v^k} \exp \left[- \frac{1}{2} (\beta-\beta_0)' A / v^2 (\beta-\beta_0) \right]$$

The product of this prior density and likelihood function is:

$$p(\beta / \sigma) \propto v^{-k}\sigma^{-T} \exp \left\{ - \frac{1}{2} \left[\frac{ns^2 + (\beta-\hat{\beta})' X'X(\beta-\hat{\beta})}{\sigma^2} + \right. \right.$$

$$\left. \left. + (\beta-\beta_0)' \frac{A}{v^2} (\beta-\beta_0) \right] \right\}$$

The quadratic form can be expanded as:

$$- \frac{1}{2} \left[ns^2 + \beta' \left(\frac{X'X}{\sigma^2} + \frac{A}{v^2} \right) \beta + \text{other terms}\right]$$

Judge (1982, p. 228) shows that this may be written:

$$(3.3) \qquad - \frac{1}{2} \left[ns^2 + (\beta-\bar{\bar{\beta}})' \left(\frac{X'X}{\sigma^2} + \frac{A}{v^2} \right) (\beta-\bar{\bar{\beta}})\right]$$

where

$$(3.4) \qquad \bar{\bar{\beta}} = \left[\frac{X'X}{\sigma^2} + \frac{A}{v^2} \right]^{-1} \left[\frac{X'X}{\sigma^2} \hat{\beta} + \frac{A}{v^2} \beta_0 \right]$$

Therefore, according to the above analysis the posterior density of β can be written:

$$p(\beta / \sigma) \propto v^{-k}\sigma^{-T} \exp - \frac{1}{2} \left\{ - \frac{1}{2} \left[ns^2 + (\beta-\bar{\bar{\beta}})' \left(\frac{X'X}{\sigma^2} + \right.\right.\right.$$

$$\left.\left.\left. + \frac{A}{v^2} \right) (\beta-\bar{\bar{\beta}}) + \text{terms not involving } \beta\right] \right\}$$

This is a normal distribution, with mean $\bar{\bar{\beta}}$, and covariance-variance matrix $[X'X/\sigma^2 + A/v^2]^{-1}$. Thus, $\bar{\bar{\beta}}$ is the Bayes estimator of β.

Add and subtract β_0 to the right hand side of (3.4) and write the $-\beta_0$ term as:

$$-\left[\frac{X'X}{\sigma^2} + \frac{A}{v^2}\right]^{-1} \left[\frac{X'X}{\sigma^2} + \frac{A}{v^2}\right]\beta_0$$

This enables the term to be brought inside the second matrix expression on the right hand side of (3.4) and the A/v^2 terms cancel. The result is:

$$\bar{\bar{\beta}} = \beta_0 + \left[\frac{X'X}{\sigma^2} + \frac{A}{v^2}\right]^{-1} \frac{X'X}{\sigma^2}(\hat{\beta}-\beta_0) \equiv \beta_0 + W(\hat{\beta}-\beta_0)$$

where the matrix $W \equiv \left(\frac{X'X}{\sigma^2} + \frac{A}{v^2}\right)^{-1} \frac{X'X}{\sigma^2}$

Finally we have:

$$\bar{\bar{\beta}} = (I-W)\beta_0 + W\hat{\beta}$$

Therefore, $\bar{\bar{\beta}}$ is a weighted average of β_0 and $\hat{\beta}$.

Bayesian Inference and Ridge Regression

It was shown in the previous section that the Bayes estimator of β is:

$$\bar{\bar{\beta}} = \left[\frac{X'X}{\sigma^2} + \frac{A}{v^2}\right]^{-1} \left[\frac{X'X}{\sigma^2}\hat{\beta} + \frac{A}{v^2}\beta_0\right] \tag{3.5}$$

If we choose the prior mean $\beta_0 = 0$ and $A = I$; this yields:

$$\bar{\bar{\beta}} = \left[\frac{X'X}{\sigma^2} + \frac{1}{v^2}I\right]^{-1} \left[\frac{X'y}{\sigma^2}\right]$$

$$= \left[X'X + \frac{\sigma^2}{v^2}I\right]^{-1} X'y$$

We know that v^2 is the common variance of the prior distribution of β. When $k = \sigma^2/v^2$, we obtain:

$$\bar{\bar{\beta}} = \beta^* = (X'X+KI)^{-1} X'y$$

This is the ridge estimator, therefore, for $k=\sigma^2/v^2$ the ridge estimator is a Bayes estimator. The question which arises here is if we can assume that the prior mean $\beta_0 = 0$ and that all regression coefficients have a common variance. This seems an unlikely characterization of prior beliefs and thus forms the basis of the Bayesian criticism of ridge regression.

The close relationship between the ridge estimator and the Bayesian estimator shows that the ridge estimator is an attempt to incorporate prior information about the unknown parameters β.

In the case of multicollinearity, the researcher usually doubts the accuracy of the sample information. That is, the researcher obtains sample results that conflict with some prior information. In particular, the researcher believes that the true parameters are closer to zero (smaller) than the sample estimates. Thus, the ridge estimator which is an average of a zero vector and the sample estimates, represents a reasonable approach to reducing the size of the regression coefficients.

We can be less restrictive in the specification of v^2 and β_0. If we choose a different prior variance for each coefficient $(v_1^2 + v_2^2 + + \dots + v_p^2)$, we obtain:

$$\begin{aligned} k_1 &= \sigma^2/v_1^2 \\ k_2 &= \sigma^2/v_2^2 \\ &\vdots \end{aligned}$$

and so on. This yields a "generalized" ridge estimator:

$$B_G^* = [X'X + \Omega]^{-1} X'y$$

where

$$\Omega = \begin{bmatrix} k_1 & 0 & \dots & 0 \\ 0 & k_2 & \dots & 0 \\ 0 & 0 & \dots & k_p \end{bmatrix}$$

Further, we can select β_0 to be nonzero and obtain the Bayes estimator using formula (3.5).

CHAPTER IV

AN APPLICATION OF RIDGE REGRESSION: DEMAND FOR GREEK IMPORTS

The aim of this chapter is to demonstrate an example where we can see ridge regression at work in data analysis in a realistic setting.

Although empirical results are presented here, and real-world data is used, no statement is made about the estimating model or results significant to economic theory, because the objective is demonstration, and not model determination. For this purpose we also assume that none of the other problems of empirical econometrics, such as heteroscedasticity or autocorrelation are present.

This chapter is divided into four sections. The first discusses the models and data used in the estimation procedure. The second section develops the least squares method and the measures of multicollinearity. The ridge regression estimates and ridge trace method of choosing k are presented in the third section. The final section is devoted to choosing k using prior information and to employing the minimax criterion for the ridge estimates.

Model and Data

In our discussion, we have chosen an example based on aggregate data concerning import activity in the Greek economy. It is assumed that a linear relationship exists between the dependent variable imports (RIMP) and two explanatory variables income (RGDP), and relative prices (RP) and a disturbance term u. For the term u it also is assumed that it is a random quantity, independently distributed with zero mean and constant variance σ^2. The description of the variables in the study is given in Table 4.1.

TABLE 4.1

Description of Variables in the Study

Variables	*Symbol*	*Description*
Imports	RIMP	Imports, cif. Billions of 1975 Drachmas
Income	RGDP	Gross Domestic Product. Billions of 1975 Drachmas
Relative Prices	RP	Import Unit Value/GDP Deflator 1975 = 100

Note: The variable RP is defined by the formula:

$$RP = \frac{\text{Unit Value of Imports (UVIMP)}}{\text{Greek Prices}}$$

We consider the following specifications for the unstandardized variables:

$$\text{Model I: } \log RIMP = \beta_0+\beta_1 \log RGDP + \beta_2 \log RP + u$$

$$\text{Model II: } \log RIMP = \beta_0+\beta_1 \log RGDP + \beta_2 \log RGDP_{-1} + \beta_3 \log RP + u$$

$$\text{Model III: } \log RIMP = \beta_0+\beta_1 \log RGDP + \beta_2 \log RGDP_{-1} + \beta_3 \log RP + \beta_4 \log RP_{-1} + u$$

and, for the standardized variables:

$$\text{Model IV: } \log SRIMP = \beta_1 \log SRGDP + \beta_2 \log SRP + u$$

$$\text{Model V: } \log SRIMP = \beta_1 \log SRGDP + \beta_2 \log SRGDP_{-1} + \beta_3 \log SRP + u$$

$$\text{Model VI: } \log SRIMP = \beta_1 \log SRGDP + \beta_2 \log SRGDP_{-1} + \beta_3 \log SRP + \beta_4 \log SRP_{-1} + u$$

The ith observation of each standardized variables Z is derived from the corresponding unstandardized variable X as follows:

$$Z_i = \frac{X_i - \bar{X}}{\sqrt{\frac{1}{n-1} \Sigma \ (X_i - \bar{X})^2}}$$

According to economic theory, the explanatory variable gross domestic product should positively influence imports. Contrarily, imports should increase as relative prices decrease. In mathematical notation, the above can be expressed as follows:

$$\frac{\partial RIMP}{\partial RGDP} > 0 \ , \ \frac{\partial RIMP}{\partial RGDP_{-1}} > 0 \ , \ \frac{\partial RIMP}{\partial RP} < 0 \ , \ \frac{\partial RIMP}{\partial RP_{-1}} < 0$$

Therefore, according to economic theory we expect from the estimation of the coefficients of the models positive signs for the coefficients of RGDP and negative signs for coefficients of RP.

The coefficients β_1 and β_2 of Model I are the elasticities with respect to income and price, respectively. In mathematical notation we have:

$$E_{RIMP,RGDP} = \frac{\partial \log \ RIMP}{\partial \log \ RGDP} \ , \quad E_{RIMP,RP} = \frac{\partial \log \ RIMP}{\partial \log \ RP}$$

The calculation of elasticities for all the models is given in Table 4.2.

TABLE 4.2

Calculation of Elasticities

	Income		*Price*	
Model	*Short Run*	*Long Run*	*Short Run*	*Long Run*
I	β_1	β_1	β_2	β_2
II	β_1	$\beta_1+\beta_2$	β_3	β_3
III	β_1	$\beta_1+\beta_2$	β_3	$\beta_3+\beta_4$
IV	$\breve{\beta}_1$	$\breve{\beta}_1$	$\breve{\beta}_2$	β_2
V	$\breve{\beta}_1$	$\breve{\beta}_1+\breve{\beta}_2$	$\breve{\beta}_3$	$\breve{\beta}_3$
VI	$\breve{\beta}_1$	$\breve{\beta}_1+\breve{\beta}_2$	$\breve{\beta}_3$	$\breve{\beta}_3+\breve{\beta}_4$

Note: The coefficients, $\breve{\beta}_i$'s, of the explanatory variables, X_i's, are defined as follows:

$$\breve{\beta}_i = \beta_i \ \frac{S_{RIM}}{S_{X_i}}$$

The data to be used in the analysis is given in Table 4.3, and was obtained from *International Financial Statistics, English Year Book 1981.* This is an historical data set with observations indexed by time, and consists of thirty observations on the dependent variable IMP and three explanatory variables, GDP, RGDP and UVIMP for the years 1951 through 1980.

TABLE 4.3

Annual Data on Greek Economy; Import Data (Billions of Greek Drachmas)

Year	*IMP*	*UVIMP*	*GDP*	*RGDP*
1951	5.98	23.2	39.4	145.5
1952	5.19	23.7	41.3	146.8
1953	7.16	35.6	54.1	167.5
1954	9.90	42.0	62.4	172.6
1955	11.47	42.3	72.1	185.8
1956	13.91	44.2	83.1	200.7
1957	15.73	45.1	89.5	215.0
1958	16.95	41.4	93.8	233.1
1959	17.01	41.1	97.5	241.5
1960	21.06	39.9	105.2	202.1
1961	21.42	39.3	118.6	280.2
1962	21.04	38.9	126.0	284.4
1963	24.13	38.8	140.8	313.2
1964	26.55	40.1	158.0	339.3
1965	34.01	40.6	179.8	370.9
1966	36.69	41.2	200.0	393.6
1967	35.59	40.8	216.1	415.2
1968	41.80	41.1	234.6	442.8
1969	47.83	41.1	266.5	486.7
1970	58.75	42.9	298.9	525.4
1971	62.94	44.4	330.3	562.9
1972	70.44	48.3	377.7	612.8
1973	102.75	57.6	484.1	657.6
1974	131.56	84.8	564.2	633.7
1975	172.02	100.0	672.2	672.2
1976	221.82	108.5	825.0	714.9
1977	252.15	116.6	963.7	739.5
1978	287.73	127.0	1,161.4	788.9
1979	356.82	155.2	1,430.9	818.8
1980	452.88	209.4	1,722.1	832.3

Source: International Financial Statistics, English Year Book 1981.

Measures of Multicollinearity

It was demonstrated in the previous theoretical discussion that when there are strong linear relationships among the explanatory variables, the least squares estimates of the individual regression coefficients tend to be unstable. The extent of multicollinearity and the estimating power of least squares method for the models can be shown by the regression results. The least squares fits are:

$$\log RIMP = \underset{(-12.076)}{-7.063} + \underset{(53.38)}{1.39}\log RGDP - \underset{(-2.91)}{0.31}\log RP$$

$$\log RIMP = \underset{(-11.26)}{-6.91} + \underset{(2.73)}{1.069}\log RGDP + \underset{(0.82)}{0.31}\log RGDP_{-1} - \underset{(-3.0076)}{0.33}\log RP$$

$$\log RIMP = \underset{(-16.58)}{-7.77} + \underset{(3.87)}{1.077}\log RGDP + \underset{(1.21)}{0.33}\log RGDP_{-1} - \underset{(-6.57)}{0.81}\log RP + \underset{(5.037)}{0.63}\log RP_{-1}$$

$$\log SRIMP = \underset{(54.40)}{0.96}\log SRGDP - \underset{(-2.97)}{0.052}\log SRP$$

$$\log SRIMP = \underset{(2.78)}{0.74}\log SRGDP + \underset{(0.84)}{0.22}\log SRGDP_{-1} - \underset{(-3.067)}{0.056}\log SRP$$

$$\log SRIMP = \underset{(3.95)}{0.75}\log SRGDP + \underset{(1.24)}{0.23}\log SRGDP_{-1} - \underset{(-6.71)}{0.13}\log SRP + \underset{(5.15)}{0.104}\log SRP_{-1}$$

The number in parenthesis below each coefficient is a t-ratio.

All the regression results are displayed in Tables 4.4, 4.5, and 4.6.

TABLE 4.4

Eigenvalues and Condition Number for Each Model

Model	*I*	*II*	*III*	*IV*	*V*	*VI*
Eigenvalues						
λ_1	0.013	0.0095	0.0088	16.18	0.048	0.016
λ_2	4.650	0.0190	0.0180	41.82	18.180	0.081
λ_3	1692.730	5.4100	0.0810		65.770	8.640
λ_4		2659.5500	8.7100			3214.910
λ_5			3243.8300			
Condition Number $\lambda_{max}/\lambda_{min}$	130,210	279,952.63	368,617	2.58	1,370.208	200,931.87

TABLE 4.5

Coefficient of Determination R_p^2 Against the Explanatory Variables and R^2 from Regression

Model	*I*	*II*	*III*
Dependent Variable			
logRGDP		0.9966	0.9966
logRGDP_{-1}		0.9965	0.9965
logRP	0.2427	0.2776	0.6962
logRP_{-1}			0.7020
R^2	0.9935	0.9937	0.9969

Note: The results of Table 4.5 are the same for Models IV, V, and VI.

TABLE 4.6

Correlation Matrix of Coefficients

	Variables	*Intercept*	*logRGDP*	*logRGDP$_{-1}$*	*logRP*
Model I	logRGDP	—0.6800			
	logRP	—0.9727	0.4926		
Model II	logRGDP	—0.3280			
	logRGDP$_{-1}$	0.2849	—0.9977		
	logRP	—0.9719	0.2466	—0.2148	
Model III	logRGDP	—0.3079			
	logRGDP$_{-1}$	0.2618	—0.9976		
	logRP	—0.3106	0.1551	—0.1469	
	logRP$_{-1}$	—0.3633	0.0063	0.0099	—0.7611

Note: The correlation matrix of coefficients is symmetric.
The results of Table 4.6 are the same for Models IV, V and VI (without including intercept).

Examining the least squares fits shows that the algebraic signs of the estimated coefficients for Models I, II, IV, and V agree with the economic theory. The sign of the estimated coefficient $\hat{\beta}_1$ for Models III, and VI is positive while it was expected, according to economic theory, to be negative. Also t-ratios for the coefficient $\hat{\beta}_2$ in the Models II, III, V, and VI appear very low, so that the null hypothesis $H_0 : \beta_2=0$, is accepted for level of significance $\alpha=0.05$. On the other hand, the resultant F values are high for all the models, therefore, the test statistic F for the explanatory variables as a group is significantly different from its tabulated value.

The regression results that are included in Table 4.4 show us that the condition number for Model I is high, while for Model IV it is very low. Also the condition number for Models II, III, V, and VI is very high. We note that this number for Models V, and VI is smaller than for the Models II, and III, respectively, but it remains high.

Investigating the results that are reported in Table 4.5 shows that a high estimated value R^2, for all the models is obtained. Also the value of the coefficients of determination R_p^2 among the explanatory variables are high for Models II, III, V, and VI, and for the case in which the explanatory variables logRGDP, and $\log RGDP_{-1}$ are considered as dependents.

Examining the simple correlations r_{ij} between the estimated coefficients shows that $r_{12}=-0.9977$ and $r_{12}=-0.9976$ for the Models II, V, and III, V, respectively. This means that there is very strong negative correlation between the coefficients $\hat{\beta}_1$ and $\hat{\beta}_2$. The correlation matrix for the coefficients is presented in Table 4.6.

All the "symptoms" which are discussed in the previous analysis for Models II, III, V, and VI, namely, high values of R^2, highly significant F combined with highly insignificant t, high condition number, high correlation between the coefficients $\hat{\beta}_1$ and $\hat{\beta}_2$ and also for Models III and VI the sign of the estimated coefficient $\hat{\beta}_4$ did not conform to prior expectation, indicate that a substantial multicollinearity problem is present.

Ridge Trace Estimates

It was noted previously the one relatively new approach to multicollinearity is to use a biased linear estimator in place of least squares. The purpose of this section is to provide the alternative estimation method of ridge regression for the imports data. It was shown, in chapter two, that ridge regression amounts to adding a scalar $k>0$ to the diagonal of the cross product or correlation matrices of the regressors before inverting them for least squares estimation. That is, we must solve the equation:

$$\beta^* = (X'X+kI)^{-1}X'y$$

for several values of the ridge coefficient k. In this section, we report results for k in the interval (0, 0.02).

The numerical results for the Models I, II, III, IV, V, and VI are shown in the Tables 4.7, 4.8, 4.9, 4.10, 4.11, and 4.12, respectively, for twenty-one different values of k, including the least squares solution $k=0$.

The ridge results that are included in Table 4.7 show that the estimate coefficients $\hat{\beta}_1$ and $\hat{\beta}_2$ of the Model I did not change as k was increased from the value 0.0 to 0.02.

Examining the ridge estimates for Model II that are reported in Table 4.8 show that the estimate coefficients of logRGDP and log $RGDP_{-1}$ changed as k was increased. Specifically, the estimated coefficient $\hat{\beta}_1$ was decreased as k was increased and $\hat{\beta}_2$ was increased, while the coefficient $\hat{\beta}_3$ did not change.

Table 4.9 shows that as k was increased the coefficients of log RGDP and logRGDP$_{-1}$ changed as in the case of the Model II.

To find the value of k which tends to stabilize the estimated coefficients of logRGDP and logRGDP$_{-1}$ we follow the ridge trace method whose maps for all the coefficients and Models I, II, and III, are portayed in Figures 4.1, 4.2, 4.3, and 4.4. Figures 4.1 and 4.2 illustrate the instability of the least squares solution. We can see that at value of k in the interval [0.001, 0.002] reasonable stability of the coefficients is achieved. If we choose k=0.001 the ridge regression fits for the Models I, II, and III are respectively:

TABLE 4.7

Ridge Estimates for Model I

k	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	R^2	F
0.000	—7.063	1.392	—0.318	0.9935	1988.270
0.001	—7.037	1.390	—0.322	0.9935	1987.901
0.002	—7.012	1.388	—0.325	0.9935	1986.820
0.003	—6.986	1.387	—0.328	0.9935	1985.039
0.004	—6.961	1.385	—0.332	0.9935	1982.551
0.005	—6.935	1.383	—0.335	0.9935	1979.376
0.006	—6.910	1.381	—0.338	0.9935	1975.528
0.007	—6.885	1.380	—0.341	0.9934	1971.021
0.008	—6.860	1.378	—0.345	0.9934	1965.867
0.009	—6.835	1.376	—0.348	0.9934	1960.087
0.010	—6.810	1.374	—0.351	0.9934	1953.697
0.011	—6.786	1.373	—0.354	0.9934	1946.714
0.012	—6.761	1.371	—0.357	0.9933	1939.157
0.013	—6.737	1.369	—0.361	0.9933	1931.046
0.014	—6.712	1.368	—0.364	0.9933	1922.402
0.015	—6.688	1.366	—0.367	0.9933	1913.245
0.016	—6.664	1.364	—0.370	0.9932	1903.595
0.017	—6.640	1.362	—0.373	0.9932	1893.475
0.018	—6.616	1.361	—0.371	0.9931	1882.905
0.019	—6.592	1.359	—0.379	0.9931	1871.908
0.020	—6.569	1.357	—0.382	0.9931	1860.505

TABLE 4.8

Ridge Estimates for Model II

k	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	R^2	F
0.000	—6.918	1.069	0.317	—0.338	0.9937	1309.818
0.001	—6.840	0.930	0.452	—0.349	0.9936	1303.162
0.002	—6.798	0.865	0.515	—0.355	0.9936	1295.559
0.003	—6.769	0.828	0.550	—0.358	0.9936	1289.786
0.004	—6.746	0.803	0.574	—0.361	0.9936	1285.367
0.005	—6.727	0.786	0.590	—0.364	0.9935	1281.832
0.006	—6.710	0.773	0.601	—0.366	0.9935	1228.873
0.007	—6.694	0.763	0.610	—0.368	0.9935	1276.292
0.008	—6.679	0.755	0.618	—0.370	0.9935	1273.959
0.009	—6.665	0.748	0.623	—0.372	0.9935	1271.788
0.010	—6.652	0.743	0.628	—0.373	0.9935	1269.719
0.011	—6.639	0.738	0.632	—0.375	0.9935	1267.709
0.012	—6.626	0.734	0.635	—0.376	0.9935	1265.728
0.013	—6.614	0.730	0.638	—0.378	0.9934	1263.752
0.014	—6.602	0.727	0.640	—0.379	0.9934	1261.764
0.015	—6.590	0.724	0.642	—0.381	0.9934	1259.751
0.016	—6.578	0.721	0.644	—0.382	0.9934	1257.704
0.017	—6.567	0.719	0.645	—0.384	0.9934	1255.614
0.018	—6.555	0.717	0.647	—0.385	0.9934	1253.475
0.019	—6.544	0.715	0.648	—0.386	0.9934	1251.284
0.020	—6.532	0.713	0.649	—0.388	0.9934	1249.035

TABLE 4.9

Ridge Estimates for Model III

k	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	R^2	F
0.000	—7.776	1.077	0.331	—0.812	0.635	0.9969	1946.560
0.001	—7.694	0.940	0.465	—0.818	0.629	0.9969	1926.587
0.002	—7.647	0.876	0.526	—0.819	0.624	0.9969	1903.748
0.003	—7.614	0.838	0.562	—0.818	0.618	0.9968	1886.196
0.004	—7.586	0.814	0.585	—0.816	0.613	0.9968	1872.433
0.005	—7.563	0.797	0.601	—0.814	0.607	0.9968	1861.043
0.006	—7.541	0.784	0.612	—0.811	0.602	0.9968	1851.115
0.007	—7.520	0.774	0.621	—0.809	0.596	0.9968	1842.076
0.008	—7.501	0.766	0.628	—0.806	0.591	0.9967	1833.558
0.009	—7.482	0.759	0.634	—0.803	0.586	0.9967	1825.320
0.010	—7.464	0.754	0.638	—0.801	0.580	0.9967	1817.201
0.011	—7.447	0.749	0.642	—0.798	0.575	0.9967	1809.088
0.012	—7.429	0.745	0.645	—0.795	0.570	0.9967	1800.905
0.013	—7.412	0.741	0.648	—0.792	0.565	0.9967	1792.598
0.014	—7.396	0.738	0.650	—0.790	0.560	0.9966	1784.132
0.015	—7.379	0.735	0.652	—0.787	0.555	0.9966	1775.481
0.016	—7.363	0.732	0.653	—0.784	0.550	0.9966	1766.630
0.017	—7.347	0.730	0.655	—0.782	0.545	0.9966	1757.569
0.018	—7.331	0.727	0.656	—0.779	0.540	0.9966	1748.293
0.019	—7.315	0.725	0.657	—0.776	0.535	0.9966	1738.802
0.020	—7.299	0.723	0.658	—0.774	0.530	0.9965	1729.097

TABLE 4.10

Ridge Estimates for Model IV

k	$\tilde{\beta}_1$	$\tilde{\beta}_2$	R^2	F
0.000	0.969	—0.0529	0.9935	4129.444
0.001	0.968	—0.0534	0.9935	4128.708
0.002	0.967	—0.0540	0.9935	4126.442
0.003	0.965	—0.0546	0.9935	4122.625
0.004	0.964	—0.0551	0.9935	4117.647
0.005	0.963	—0.0557	0.9935	4111.074
0.006	0.962	—0.0562	0.9935	4102.983
0.007	0.960	—0.0567	0.9934	4093.685
0.008	0.959	—0.0573	0.9934	4082.902
0.009	0.958	—0.0578	0.9934	4071.011
0.010	0.957	—0.0583	0.9934	4057.622
0.011	0.956	—0.0589	0.9934	4043.230
0.012	0.954	—0.0594	0.9933	4027.512
0.013	0.953	—0.0599	0.9933	4010.641
0.014	0.952	—0.0649	0.9933	3992.649
0.015	0.951	—0.0610	0.9933	3973.624
0.016	0.950	—0.0615	0.9932	3953.597
0.017	0.949	—0.0620	0.9932	3932.546
0.018	0.947	—0.0625	0.9931	3910.616
0.019	0.946	—0.0630	0.9931	3887.785
0.020	0.945	—0.0635	0.9931	3864.144

TABLE 4.11

Ridge Estimates for Model V

k	$\tilde{\beta}_1$	$\tilde{\beta}_2$	$\breve{\beta}_3$	R^2	F
0.000	0.744	0.224	—0.0562	0.9937	2043.358
0.001	0.647	0.319	—0.0580	0.9936	2032.910
0.002	0.602	0.363	—0.0589	0.9936	2021.125
0.003	0.576	0.388	—0.0596	0.9936	2012.034
0.004	0.559	0.405	—0.0601	0.9936	2005.189
0.005	0.547	0.416	—0.0605	0.9935	1999.712
0.006	0.538	0.424	—0.0608	0.9935	1994.979
0.007	0.531	0.431	—0.0612	0.9935	1990.981
0.008	0.525	0.436	—0.0615	0.9935	1987.369
0.009	0.521	0.440	—0.0618	0.9935	1983.995
0.010	0.517	0.443	—0.0620	0.9935	1980.774
0.011	0.514	0.446	—0.0623	0.9935	1977.676
0.012	0.511	0.448	—0.0626	0.9935	1974.588
0.013	0.508	0.450	—0.0628	0.9934	1971.256
0.014	0.506	0.451	—0.0631	0.9934	1968.326
0.015	0.504	0.453	—0.0633	0.9934	1965.183
0.016	0.502	0.454	—0.0635	0.9934	1962.023
0.017	0.500	0.455	—0.0638	0.9934	1958.790
0.018	0.499	0.456	—0.0640	0.9934	1955.458
0.019	0.498	0.452	—0.0642	0.9934	1952.081
0.020	0.496	0.458	—0.0645	0.9934	1948.441

TABLE 4.12

Ridge Estimates for Model VI

k	$\ddot{\beta}_1$	$\ddot{\beta}_2$	$\ddot{\beta}_3$	$\bar{\beta}_4$	R^2	F
0.000	0.750	0.233	—0.134	0.1044	0.9969	2703.518
0.001	0.654	0.328	—0.135	0.1031	0.9969	2675.831
0.002	0.609	0.371	—0.136	0.1027	0.9969	2644.111
0.003	0.584	0.396	—0.135	0.1016	0.9968	2620.244
0.004	0.567	0.412	—0.135	0.1007	0.9968	2600.587
0.005	0.555	0.424	—0.135	0.0998	0.9968	2584.780
0.006	0.546	0.432	—0.134	0.0989	0.9968	2521.037
0.007	0.539	0.438	—0.134	0.0980	0.9968	2558.467
0.008	0.533	0.332	—0.134	0.0971	0.9967	2546.604
0.009	0.528	0.447	—0.133	0.0962	0.9967	2535.168
0.010	0.525	0.450	—0.133	0.0954	0.9967	2523.861
0.011	0.521	0.453	—0.132	0.0945	0.9967	2512.654
0.012	0.518	0.455	—0.132	0.0937	0.9967	2501.250
0.013	0.516	0.457	—0.131	0.0928	0.9967	2487.709
0.014	0.513	0.458	—0.131	0.0920	0.9966	2477.9828
0.015	0.511	0.460	—0.130	0.0912	0.9966	2465.934
0.016	0.510	0.461	—0.130	0.0903	0.9966	2453.653
0.017	0.508	0.462	—0.129	0.0895	0.9966	2441.070
0.018	0.506	0.463	—0.129	0.0887	0.9966	2428.209
0.019	0.505	0.463	—0.129	0.0879	0.9966	2414.981
0.020	0.504	0.464	—0.128	0.0871	0.9965	2401.523

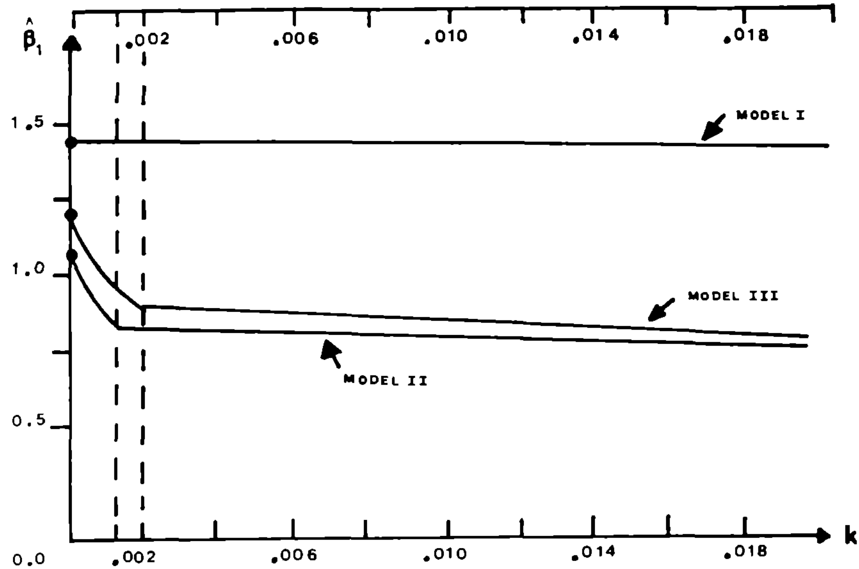

Figure 4.1. Ridge Trace of the Coefficient of logRGDP for the Models I, II, and III

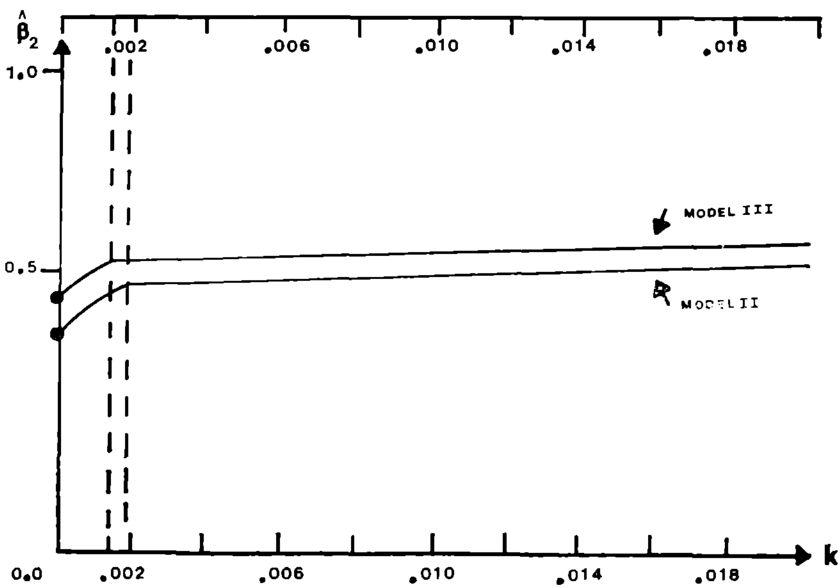

Figure 4.2. Ridge Trace of the Coefficient of logRGDP$_{-1}$ for the Models II and III.

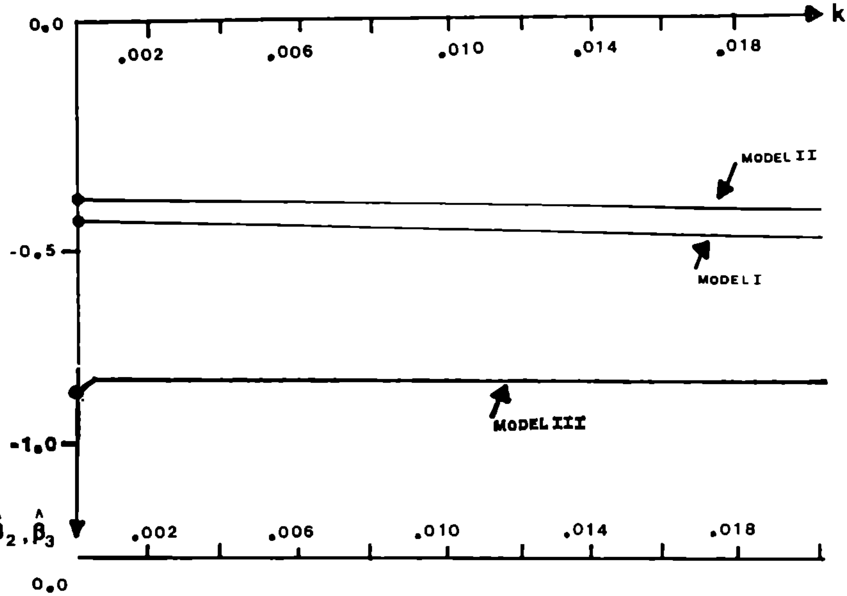

Figure 4.3. Ridge Trace of the Coefficient of logRP for the Models I, II, and III.

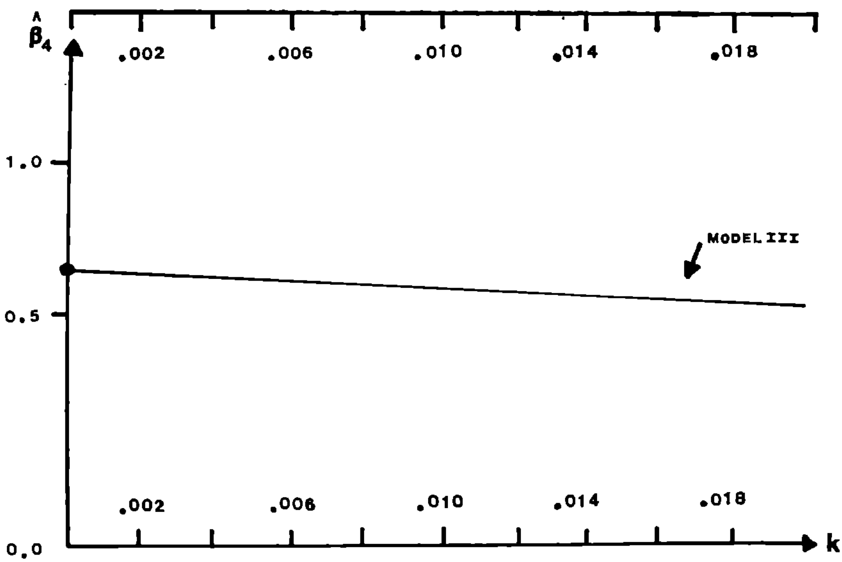

Figure 4.4. Ridge Trace of the Coefficient of logRP$_{-1}$ for the Model III.

$$\log RIM = \underset{(-12.053)}{-7.037} + \underset{(53.4)}{1.39} \log RGDP - \underset{(-2.95)}{0.318} \log RP$$

$$\log RIM = \underset{(-11.46)}{-6.84} + \underset{(3.74)}{0.93} \log RGDP + \underset{(1.9)}{0.452} \log RGDP_{-1}$$

$$- \underset{(-3.14)}{0.349} \log RP$$

$$\log RIM = \underset{(-16.75)}{-7.694} + \underset{(5.31)}{0.94} \log RGDP + \underset{(2.68)}{0.465} \log RGDP_{-1}$$

$$- \underset{(-6.66)}{0.818} \log RP + \underset{(4.99)}{0.629} \log RP_{-1}$$

Note: The number in parenthesis below each coefficient is a t-ratio.

Minimaxity and Bayesian Estimators

To find the value of k which tends to stablize the estimated coefficients, the ridge trace method was used in the previous section, and $k = 0.001$ was chosen. The purpose of this current section is to employ two alternative procedures for choosing k and to calculate also the Bayesian estimators for import data.

One popular ridge estimator chooses $k = \hat{\sigma}^2/\hat{\Theta}^2_{max}$. Although this may be performed iteratively, the initial calculation of k from this formula is given in Table 4.13. Examining the numerical results that are included in this table, shows that the k's are close to interval (0.01, 0.025).

Another popular ridge estimator chooses $k = P\hat{\sigma}^2/\hat{\Theta}'\hat{\Theta}$. The calculation of k from this formula, that is reported in Table 4.14, shows that k's are close to interval (0.0002, 0.054).

Thisted has computed the condition for minimaxity of $P\hat{\sigma}^2/\hat{\Theta}'\hat{\Theta}$ as:

$$\frac{\sum_{i=1}^{P} \lambda_i^2}{\lambda_p^{-2}} \geqslant 2 + \frac{P}{2} \quad , \quad P \geqslant 3$$

Investigating the results of Thisted's condition for minimaxity for the import data which are given in Table 4.15, shows that for all the models, Thisted's condition does not hold. Thus, the ridge estimator is not minimax for our data.

All the above results suggest a value of k close to the value .001 selected on the basis of the ridge trace. Thus, if ridge regression is used to estimate a Greek import function, it is clear that the value of k should be small.

In order to obtain Bayesian estimates of the demand elasticities for imports with respect to income and price for the Greek economy, we follow the theoretical background of chapter four. First, we must specify the prior density function for the regression parameters. We believe that the demand elasticity for imports with respect to income can be described by a normal distribution with mean equal to 1.5 (ela-

TABLE 4.13

Calculation $k = \hat{\sigma}^2/\hat{\Theta}^2_{max}$; *Import Data*

Model	$\hat{\sigma}^2$	$\hat{\Theta}^2_{max}$	*k*
I	0.0043	3.92	0.0010
II	0.0044	28.18	0.0001
III	0.0022	1.96	0.0010
IV	0.0067	0.56	0.0110
V	0.0068	0.25	0.0270
VI	0.0035	1.96	0.0010

TABLE 4.14

Calculation $k = P\hat{\sigma}^2/\hat{\Theta}'\hat{\Theta}$; *Import Data*

Model	$\hat{\sigma}^2$	*P*	$\hat{\Theta}'\hat{\Theta}$	*k*
I	0.0043	3	45.13	0.00020
II	0.0044	4	52.10	0.00033
III	0.0022	5	62.65	0.00018
IV	0.0067	2	1.05	0.01300
V	0.0068	3	0.38	0.05368
VI	0.0035	4	4.54	0.00300

TABLE 4.15

Thisted's Condition for Minimaxity; Import Data

Model	$\Sigma\lambda^{-2}_i$	$\Sigma\lambda^{-2}_p$	$\Sigma\lambda_i^{-2}/\lambda^{-2}_p$
I	5917.2058	5917.1596	1.00
II	13850.4470	11080.3300	1.25
III	16152.0690	12913.2220	1.25
IV	0.0044	0.0038	1.15
V	434.0309	434.0277	1.00
VI	4062.4800	3906.2500	1.04

Note: λ_p is the smallest eigenvalue.

stic). We believe also that the smallest value the income elasticity is likely to have is 1.0 and the largest is 2.0. This implies that the standard deviation S will be equal to 0.25 for the confidence coefficient $\alpha = 0.05$. Therefore, it can be written for Model I that $\beta_1 \sim N(1.5,\ 0.25)$. We believe also that the price elasticity has a normal distribution with mean —0.5 (inelastic) and standard deviation 0.25 because we consider that the confidence limits are —1.0 and 0. Thus, $\beta_2 \sim N(-0.5,\ 0.25)$. These prior beliefs were obtained on the basis of the survey of the literature in Houthakker and Magee (1969). The third parameter of Model I is the intercept. For simplicity, we make our prior mean for the intercept approximately equal to ordinary least squares estimate. This effectively ensures that the Baysian estimate is equal to the ordinary least squares estimate. We again assume that the prior distribution is normal with a standard deviation equal to 0.25.

For Models II and III, the income elasticity is divided between immediate and one prior delayed values. We continue to assume that the long run elasticity is 1.5, and we assume that most of this elasticity occurs within the current period. Thus, we set means for β_1 at 1.0 and for β_2 at 0.5 with standard deviations equal to 0.25.

In Model III, the price elasticity is also divided between immediate and one period delayed values. Although we continue to assume the same prior mean for the long run elasticity, we assume that the one period delayed elasticity is larger than the immediate elasticity. Thus we specify the prior mean for β_3 at —0.2, and the prior mean for β_4 at —0.3.

The results of the Bayesian estimators are given in Tables 4.61 through 4.19. Examining the reported tables shows that the prior variance is large relative to the ordinary least squares variance. That is, $S^2/\hat{\sigma}^2 \rightarrow \infty$ which means that the Bayesian estimator is approximately equal to ordinary least squares estimator. Therefore, Bayesian estimation does not deal with the multicollinearity problem in this data set.

TABLE 4.16

Ordinary Least Squares, and Bayes Estimates, for Model I (Including Prior Information)

Coefficient	*Prior*	*OLS*	*Bayes*
β_0	—7.0	—7.061	—7.011
β_1	1.5	1.390	1.660
β_2	—0.5	—0.320	—0.590

$\hat{\sigma}^2 = 0.0043$, $S^2 = 0.0625$

TABLE 4.17

Ordinary Least Squares, and Bayes Estimates for Model II (Including Prior Information)

Coefficient	*Prior*	*OLS*	*Bayes*
β_0	—7.0	—6.920	—6.9700
β_1	1.0	1.071	1.0041
β_2	0.5	0.380	0.4700
β_3	—0.5	—0.390	—0.3900

$\hat{\sigma}^2 = 0.0043$, $S^2 = 0.0625$

TABLE 4.18

Ordinary Squares, and Bayes Estimates, for Model III (Including Prior Information)

Coefficient	*Prior*	*OLS*	*Bayes*
β_0	—7.0	—7.800	—7.045
β_1	1.0	1.081	1.380
β_2	0.5	0.330	0.720
β_3	—0.2	—0.810	—0.510
β_4	—0.3	0.630	0.210

$\hat{\sigma}^2 = 0.0022$, $S^2 = 0.0625$

TABLE 4.19

Price and Income Elasticities from Ordinary Least Squares and Bayes Estimates (Including Prior Information)

Model		*I*		*II*		*III*	
		SR	*LR*	*SR*	*LR*	*SR*	*LR*
Prior:	Income Elasticity	1.50	1.50	1.00	1.50	1.00	1.50
	Price Elasticity	—0.50	—0.50	—0.50	—0.50	—0.20	—0.50
OLS:	Incomo Elasticity	1.39	1.39	1.07	1.45	1.08	1.41
	Price Elasticity	—0.32	—0.32	—0.39	—0.39	—0.81	—0.18
Bayes:	Income Elasticity	1.66	1.66	1.0041	1.47	1.38	2.10
	Price Elasticity	—0.59	—0.59	—0.3900	—0.39	—0.51	—0.30

Note: The symbols SR and LR mean short run and long run, respectively.

CHAPTER V

SUMMARY, CONCLUSIONS, AND SUGGESTIONS

The main purpose of this study was to demonstrate the theory and logic of the biased method of statistical estimation, which was introduced by the chemical engineer Hoerl and it is called ridge regression.

The first chapter reviewed the assumptions under which the method of least squares has some very attractive statistical properties. It was shown that one crucial assumption is that the number P of columns in X matrix is less than the number T of observations. In other words, there is no exact linear relationship among the X variables. In addition, it was discussed that when this assumption is only just satisfied, a common situation in practice, a multicollinearity problem exists.

Chapter two developed the theoretical side of ridge regression. The effect of multicollinarity on ordinary least squares estimation was explored in the first section. It was shown that, according to the Gauss-Markov theorem, the least squares estimator is linear, unbiased and has minimum variance in the class of unbiased linear estimators. But there is no guarantee that the variance of the least squares estimator will be small. In the particular case of multicollinearity, the variances of the estimated coefficients tend to be large. It was also shown that multicollinearity tends to produce least squares estimators that are too large in absolute value.

The ridge estimator was defined to be that estimator which minimizes the sum of the squared distances of the points from the estimated line subject to a constraint on the length of the estimating vector. It was shown that as the ridge parameter k increases, the ridge estimators get smaller and smaller in absolute size and the ridge regression produces an estimator with a smaller variance than ordinary least squares. Because the ridge estimator is biased its technique was compared to ordinary least squares in terms of mean square errors. It was demonstrated that if the ridge parameter is chosen between zero and σ^2/Θ^2_{max}, the ridge mean square error will be less than the mean square error of the ordinary least squares estimator. The result arises because as k increases the reduction in variance exceeds the increase in bias.

Chapter two concluded with a discussion of two methods of choosing the ridge parameter. It was noted that these procedures make k a function of the sample data, and, therefore, k becomes stochastic.

Chapter three surveyed the critical analysis of ridge regression that have been developed by statisticians outside the classical least squares framework. The decision theory of biased estimators was reviewed and it was shown that for choosing k without reference to the data the ridge regression does not yield a minimax estimator. In addition, the relationship between Bayesian statistical inference and ridge regression were discussed. It was proven that for choosing the prior mean equal to zero, and a common variance for all regression coefficients, there is a particular value of k for which the ridge estimator is a Bayes estimator. This close relationship between the ridge estimator and the Bayesian estimator shows that the ridge estimator is an attempt to incorporate prior information into the estimation process.

Chapter four presented the ridge regression method in practice. Its technique was compared to ordinary least squares in the context of estimating price and income elasticities for Greek imports. A number of specifications were examined. These specifications differed by the amount of multicollinearity.

The ordinary least squares estimates show that for Models I and IV the system is not seriously different from orthogonality, while for the rest of the models a substantial multicollinearity problem is present. The employment of the ridge trace method confirmed these results. The coefficients of logRGDP and logRGDP_{-1} in Models II and III changed as k was increased. This method gave for the import data a small value of k. In the interval (0.001, 0.002). Two alternative methods of computating k also gave similar results. However, the ridge regression did not yield a minimax estimator, while the Bayesian approach gave an estimator approximately equal to ordinary least squares.

In conclusion, it appears that the conventional model of Greek imports is subject to a substantial multicollinearity problem if lags are included in the model. Although ridge estimation provided more stable estimators for small values of the ridge parameter, the estimator was not minimax. In the case of a lagged price specification, a wrong sign was encountered in both the ordinary least squares and ridge estimations.

BIBLIOGRAPHY

Brown, William G., and Beattie, Bruce R. Improving Estimates of Economic Parameters by Use of Ridge Regression with Production Function Applications. *American Journal of Agricultural Economics*, 1975, 57, pp. 21-34.

Chatterjee, Samprit, and Price, Bertram. *Regression Analysis by Example*. New York: John Wiley and Sons, 1977.

Chow, Gregory C. *Econometrics*. New York: McGraw-Hill, 1983.

Christou, George C. *Introductory Econometrics*. 2nd ed. Thessaloniki: Gutenberg, 1982.

Dhrymes, Phoebus J. *Introductory Econometrics*. New York: Springer-Verlag, 1978.

Fienberg, Stephen E., and Zellner, Arnold. *Studies in Bayesian Econometrics and Statistics*. New York: North-Holland Publishing Company, 1977.

Goldberger, A. S. *Econometric Theory*. New York: John Wiley and Sons, 1964.

Greenberg, Edward, and Webster, Charles E., Jr. *Advanced Econometrics: A Bridge to the Literature*. New York: John Wiley and Sons, 1983.

Guilkey, David K., and Murphy, James L. Directed Ridge Regression Techniques in Cases of Multicollinearity. *Journal of the American Statistical Association*, 1975, 70, pp. 769-775.

Gujarati, Damodar. *Basic Econometrics*. New York: McGraw-Hill, 1978.

Henry, Neil W. A Note on Ridge Regression. *Sociological Methods and Research*, 1976, 4, pp. 495-500.

Hoerl, Arthur E., and Kennard, Robert W. Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, 1970, 12, pp. 55-67.

Hoerl, Arthur E., and Kennard, Robert W. Applications to Nonorthogonal Problems. *Technometrics*, 1970, 12, pp. 69-82.

Hoerl, Arthur E., Kennard, Robert W., and Baldwin, Kent F. Ridge Regression: Some Simulations. *Communications in Statistics*, 1975, 4, pp. 105-123.

Houthakker, H. S., and Magee, S. P. Income and Price Elasticities in World Trade. *The Review of Economics and Statistics*, 1969, 51, pp. 111-125.

Intriligator, Michael D. *Econometric Models, Techniques, and Applications*. New Jersey: Prentice-Hall, 1978.

Johnston, J. *Econometric Methods*. 2nd ed. New York: McGraw-Hill, 1972.

Judge, George G., Hill, Carter R., Griffiths, William E., Lütkepohl, Helmut, and Lee, Tsoung-Chao. *Introduction to the Theory and Practice of Econometrics*. New York: John Wiley and Sons, 1982.

Judge, George G., Griffiths, William E., Hill, Carter R., and Lee, Tsoung-Chao. *The Theory and Practice of Econometrics*. New York: John Wiley and Sons, 1980.

Kasarda, John D., and Shih, Wen-Fu P. "Optimal Bias in Ridge Regression Approaches to Multicollinearity". *Sociological Methods and Research,* 1977, 5, pp. 461-469.

Kidis, Adreas A. *Econometrics.* Athens: Gutenberg, 1982.

Lukacs, Eugene. *Probability and Mathematical Statistics.* New York: Academic Press, 1972.

Maddala, G. S. *Econometrics.* Tokyo: McGraw-Hill, 1977.

Malinvaud, E. *Statistical Methods of Econometrics.* 3rd ed. Amsterdam: North-Holland Publishing Company, 1980.

Mallows, C. L. Some Comments on Cp. *Technometrics,* 1973, 15, pp. 661-675.

Marquardt, Donald W. Generalized Inverses, Ridge Regression, Biased Linear Estimation, and Nonlinear Estimation. *Technometrics,* 1970, 12, pp. 591-612.

Marquardt, Donald W., and Snee, Ronald D. Ridge Regression in Practice. *American Statistician,* 1975, 29, pp. 3-19.

Mason, Robert, and Brown, William G. Multicollinearity Problems and Ridge Regression in Sociological Models. *Social Science Research,* 1975, 4, pp. 135-149.

Mayer, Lawrence S., and Willke, Thomas A. On Biased Estimation in Linear Models. *Technometrics,* 1973, 15, pp. 497-508.

Montgomery, Douglas C., and Peck, Elizabeth A. *Introduction to Linear Regression Analysis.* New York: John Wiley and Sons, 1982.

Obeuchain, R. L. Classical F-Tests and Confidence Regions for Ridge Regression. *Technometrics,* 1977, 19, pp. 429-439.

Obeuchain, R. L. Good and Optimal Ridge Estimators. *Annals of Statistics,* 1978, 6, pp. 1111-1121.

Raj, Balder, and Vllah, Aman. *Econometrics.* New York: St. Marins Press, 1981.

Rothenberg, Thomas J. *Efficient Estimation with A Priori Information.* New Haven, and London: Yale University Press, 1973.

Seber, G. A. F. *Linear Regression Analysis.* New York: John Wiley and Sons, 1977.

Smith, G., Campbell, F., Thisted, R. A., Marquardt, D. W., Van-Nostrand, R. C., Lindley, D. V., Obeuchain, R. L., Peele, L. C., Ryan, T. P., Vinod, H. D., Gunst, R. F. A Critique of Some Ridge Regression Methods. *American Statistical Association Journal,* 1980, 75, pp. 74-103.

Swindel, Benee F. Instability of Regression Coefficients Illustrated. *The American Statistician,* 1974, 28, pp. 63-65.

Theobald, C. M. Generalizations of Mean Square Error Applied to Ridge Regression. *Royal Statistical Society,* 1974, B-36, pp. 103-106.

Vinod, Hrishikesh D. Application of New Ridge Regression Methods to a Study of Bell System Scal Economies. *Journal of the American Statistical Association,* 1976, 71, pp. 835-841.

Wonnacott, Ronald J., and Wonnacott, Thomas H. *Econometrics.* New York: John Wiley and Sons, 1970.

Zellner, Arnold. *An Introduction to Bayesian Inference in Econometrics.* New York: John Wiley and Sons, 1971.