# Prior Information and Heuristic Ridge Regression for Production Function Estimation

## Oscar R. Burt, Michael D. Frank, and Bruce R. Beattie

A heuristic criterion for choosing an acceptable level of bias in ridge regression is presented. The criterion is based on a noncentral $F$-test of the stochastic restrictions implicit in the ridge estimator. An appropriate significance level for the test is based on conjunctive use of strong and weak mean square error criteria. The procedure is illustrated in estimating a Cobb-Douglas production function for the Central Valley of California using factor shares as priors rather than the null vector. Preliminary results suggest that a conjunctive SMSE/WMSE criterion with more "reasonable" priors selects an estimator with smaller bias than ridge trace.

*Key words:* factor shares, mean-squared error, prior information, ridge regression.

It has been shown that ridge regression may be a viable estimation procedure for mitigating the deleterious effects of multicollinearity (Hoerl and Kennard 1970a, b; Swindel). Brown and Beattie demonstrated the applicability of ordinary ridge regression (ORR) to Cobb-Douglas production function estimation. Yet, two problems seriously hamper the efficacy of ORR in production function applications: (*a*) the lack of a good sample-based criterion for determining the bias of the estimator, and (*b*) the implicit shrinkage of the OLS estimator toward an untenable prior, namely, the null vector. The purpose of this paper is to outline and demonstrate a procedure for ameliorating these major shortcomings of ORR.

Fomby and Johnson have suggested an alternative form of the ridge estimator based on stochastic prior information. Their stochastic prior ridge (SPR) estimator allows for estimator selection based on a noncentral $F$-test of the implicit restrictions. However, the procedure has an objective statistical basis which does not depend on the individual analyst only when taking the biasing variable as given and making a comparison with ordinary least squares (OLS): the important practical problem of choosing the level of bias is not addressed directly.

This paper develops a heuristic criterion for choosing the level of bias in the SPR model. An alternative formulation of the SPR procedure is utilized in conjunction with a sample-based criterion to determine the level of bias in estimating Cobb-Douglas production function parameters for the central California valley. Because factor shares provide a robust point estimate of factor elasticities, they are used as the point toward which the ridge estimator is forced instead of the null vector. We begin by reviewing mean square error (MSE) criteria and the ORR and SPR estimators. A strategy for comparing alternative ridge estimators is then proposed. The OLS, ORR, and SPR parameter estimates for the application are presented and discussed and conclusions are drawn.

## A Review of MSE Criteria and OLS, ORR, and SPR Estimators

Consider the classical linear model,

(1)                $y = X\beta + \epsilon$

where $y$ is an $(n \times 1)$ vector of observations on the dependent variable, $X$ is an $(n \times p)$ matrix of fixed observations on the explanatory variables, $\beta$ is a $(p \times 1)$ vector of unknown parameters, and $\epsilon$ is an $(n \times 1)$ vector of random disturbances distributed $N(0, \sigma^2 I_n)$. Also, assume that $y$ and each column of $X$ have been standardized to have zero sample means and unit variances.

### MSE Criteria

Two mean square error criteria are commonly used in comparisons of estimators of the unknown parameter vector $\beta$. For any estimator of $\beta$ (denoted $b$), let $MSE(b)$ be the mean square error matrix of $b$, i.e., $MSE(b) = E(b - \beta)(b - \beta)'$. The strong mean square error (SMSE) criterion ranks estimator $b^1$ superior to $b^2$ when $MSE(b^2) - MSE(b^1)$ is a positive semidefinite matrix. This implies that any arbitrary linear combination of the components of $\beta$ is estimated with at least as small MSE using $b^1$ as using $b^2$. The weak mean square error (WMSE) criterion ranks estimator $b^1$ superior to $b^2$ when the trace of $MSE(b^2) - MSE(b^1)$ is non-negative.[1]

The latter criterion can be interpreted geometrically as a preference for the estimator with a smaller expected squared distance from the unknown parameter vector. It is also the criterion traditionally used in the evaluation of ridge regression estimators, such as in the seminal paper of Hoerl and Kennard (1970b).

### The OLS and ORR Estimators

The OLS estimator is

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

which is unbiased and efficient. As an alternative to (2), Hoerl and Kennard suggest augmenting the diagonal elements of the $X'X$ matrix with some constant, $k$. Their ORR estimator takes the form

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

This estimator, unlike the OLS estimator, is biased; but for any arbitrary vector $l$, $\text{var}(l'\hat{\beta}^*) \leq \text{var}(l'\hat{\beta})$. Hoerl and Kennard (1970b) further show that there always exists a ridge estimator which has a smaller WMSE than that of OLS. However, since there exists an entire family of ORR estimators (one for each $k$-value, $0 < k < \infty$), selecting the optimal estimator or even a relatively good estimator presents a problem because the optimum depends on unknown parameters. Hoerl and Kennard (1970a) suggest selection by means of the ridge trace, i.e., a graphic portrayal of the relationship between each $\hat{\beta}_i^*$ and $k$. It is suggested that $k$ be chosen at the smallest value where the components of $\hat{\beta}^*$ tend to "stabilize" in some subjective sense. As Fomby and Johnson point out, selecting $k$ based on the point estimates of the parameters is questionable.

### The SPR Estimator

Fomby and Johnson argue that the addition of $k$ to the estimation procedure is in effect the introduction of some prior information on $\beta$.[2] They showed that given prior information of the form

(4)                $\sqrt{k}B = \sqrt{k}\beta + \sqrt{k}\nu,$

with $B$ a random variable and $\nu$ an error vector representing the uncertainty of the prior $B$,[3] the resulting ridge estimator (referred to as the SPR estimator) is

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

Notice that the magnitude of $k$ affects the weight assigned to the prior information. For $k = 0$, the prior information is not considered and (5) becomes the OLS estimator. As $k \to \infty$, the prior information dominates the sample information and $\tilde{\beta}^*$ equals $B$.[4] Also, note that for $B = 0$ and $0 < k < \infty$, (5) would appear to be the ORR estimator; but $B$ identically zero, as in ORR, is inconsistent with (4), where $B$ is a random variable.

Clearly, the prior information represented by $B$ must be a random variable and not the mean of a prior distribution for $\beta$ as in a Bayes-

---

[1] This is what Wallace calls first WMSE.

[2] Brown and Beattie referred to this but without proof. Later, Smith showed that the ridge estimator could be written in a mixed estimator framework.

[3] The covariance matrix of $\nu$ is $(\sigma^2/k)I$.

[4] This result is illustrated by rewriting (5) as $(X'X/k + I_p)\tilde{\beta}^* = X'y/k + B$, and taking the limit of both sides as $k \to \infty$.

ian interpretation of ridge regression or Chipman's minimum mean-square-error estimator. In a production function application, $B$ could be factor shares from an independent set of data which was considered comparable to the sample at hand or factor shares from the sample itself. The latter estimate involves only the set of independent variables, so there is no violation of the assumed independence of $\epsilon$ and $\omega$ in (6) below.

The methodology of Fomby and Johnson for applying the SMSE criterion to ridge regression can be described by the mixed estimation model,

$$(6) \qquad \begin{bmatrix} y \\ \sqrt{k}B \end{bmatrix} = \begin{bmatrix} X & 0 \\ 0 & \sqrt{k}I_p \end{bmatrix} \begin{bmatrix} \beta \\ \gamma \end{bmatrix} + \begin{bmatrix} \epsilon \\ \omega \end{bmatrix},$$

where the distrubance vector is assumed to obey the classical assumptions of the regression model. Both $\beta$ and $\gamma$ are $p$-component vectors of unknown parameters, but $\gamma$ is more of an artifice than a genuine parameter vector.[5] In the unrestricted case where $\gamma$ is a free parameter vector, the observation vector $\sqrt{k}B$ in (6) is a set of "artificial" observations, and the least squares estimators of $\beta$ and $\gamma$ are unrelated. Hence, $\hat{\beta}$ is the OLS estimator for the classical model given in (2) while $\hat{\gamma} = B$.

The test statistic for the SMSE criterion is the ordinary $F$-statistic for testing the hypothesis $\gamma = \beta$ against the general case of $\gamma$ as a free parameter vector. Let $SSE_o$ be the sum of squares for error in the unrestricted case and $SSE_k$ be the same measure for the restricted case when using a particular value of $k$ in (6). Both sums of squares are associated with the model in (6). The test statistic, which follows the noncentral $F$ distribution, is

$$(7) \qquad F = \left(\frac{n - p}{p}\right)\left(\frac{SSE_k - SSE_o}{SSE_o}\right),$$

where $SSE_o$ is simply the sum of squared residuals from a least squares fit to (1) since necessarily $\sqrt{k}I_p\hat{\gamma} = \sqrt{k}B$.

For the restricted case, (6) can be written as

$$(8) \qquad \begin{bmatrix} y \\ \sqrt{k}B \end{bmatrix} = \begin{bmatrix} X \\ \sqrt{k}I_p \end{bmatrix}\beta + \begin{bmatrix} \epsilon \\ \omega \end{bmatrix},$$

and the least squares estimator of $\beta$ is given by (5). $SSE_k$ is the sum of squared residuals

[5] This method of extending the parameter space to transform stochastic restrictions on the linear model into an exactly restricted least squares model is due to Judge, Yancey, and Bock.

associated with the observation vector $y$ and the ridge estimator $\tilde{\beta}^*$ in (5) plus $k$ times the sum of squared deviations of $B_i$ from $\tilde{\beta}_i^*$, $i = 1, 2, \ldots, p$.

The advantage of the SPR estimator is that since it is, in effect, a formulation of an exactly restricted least-squares model, the estimator for a specific $k$ and $B$ can be evaluated by the mean square error criterion. Toro-Vizcarrondo and Wallace have shown that a restricted estimator such as $\tilde{\beta}^*$ is better in SMSE than an unrestricted estimator, i.e., $\hat{\beta}$, if the noncentrality parameter, $\theta$, of the classical $F$-statistic of the restriction is less than or equal to $\frac{1}{2}$. Since $\theta$ is an unknown parameter, a test procedure was outlined. The null hypothesis $\theta \leq \frac{1}{2}$ is not rejected at an $\alpha$ level of significance if the $F$-value of a standard test of the restrictions is less than the critical value associated with a noncentral $F$-distribution with $\theta = \frac{1}{2}$ and probability $\alpha$ to the right of the critical value.

The SMSE criterion applies to any linear combination of the components of $\beta$ and $\gamma$, but only those involving $\beta$ alone are of interest because $\gamma$ is merely an artifice to implement a formal statistical test. The restricted estimator in the extended parameter space from (8) is $[\tilde{\beta}^{*\prime}\tilde{\beta}^{*\prime}]$, while the OLS estimator is $[\hat{\beta}'\hat{\gamma}']$. If the restricted estimator is superior to the OLS estimator in SMSE, then $\tilde{\beta}^*$ is superior to $\hat{\beta}$ as well because $\beta$ is a set of linear combinations of the extended parameter vector $[\beta'\gamma']$. Also, $\tilde{\beta}^*$ would be superior to $\hat{\beta}$ by the WMSE criterion because each component of $\tilde{\beta}^*$ has smaller MSE than the respective component of $\hat{\beta}$. But the methods given in Wallace to apply the WMSE criterion are not appropriate here because the results would have to apply to the extended parameter vector $[\beta'\gamma']$, not to $\beta$ alone.

## Strategies for Choosing an SPR Estimator

Let $H$ be an hypothesis to be tested by classical methods and the marginal significance level (often called the $p$-value) for a given outcome from the sample data is denoted $MSL$. De Groot has presented a logical basis for interpreting $MSL$ as an aproximation to the posterior probability that $H$ is true. By a different approach, Lindley also provides justification for the same type of Bayesian interpretation of tail areas for the $F$-statistic when the sample

**Table 1.   Factor Elasticity Estimates and Related Statistics, Central California Valley, 1974**

| Estimator | Intercept | Factor Elasticity Estimates[a] | | | |
|---|---|---|---|---|---|
| | | Land | Labor | Machinery | Water |
| Factor Shares (FS) | | .215 | .187 | .0344 | .325 |
| OLS | 4.999 | .191 (.105) | −.0770 (.178) | −.0257 (.171) | −.225 (.152) |
| ORR[d] | 11.917–17.223 | .212–.189 (.0275–.0186) | .162–.159 (.0260–.0182) | .145–.153 (.0230–.0142) | .0953–.110 (.0238–.0163) |
| $SPR(B = 0)$ | 5.695 | .276 (.0800) | .124 (.111) | .0684 (.136) | −.0260 (.0965) |
| $SPR(B = FS)$[e] | 7.973 | .292 (.074) | .204 (.0950) | .0330 (.121) | .0650 (.0840) |

[a] Values in parentheses are standard errors. Their limitation should be recognized in interpretation of biased estimators.
[b] The sum of the factor elasticity estimates may not equal the returns to scale reported due to rounding error.
[c] Degrees of freedom are 6 and 8 for the two versions of the SPR estimator.
[d] The ranges represent the resulting estimates which might be subjectively chosen from the ridge trace (see fig. 1).
[e] Each prior factor elasticity set equal to the estimated factor share (FS) of each input (see footnote 6).

size is large. This interpretation of $MSL$ as a posterior probability that $H$ is true is applied to the noncentral $F$-distribution in the evaluation of the SPR estimator using an MSE criterion.

The upper bound for the noncentrality parameter $\theta$ above which the OLS estimator is superior in MSE and below which the linearly restricted estimator is better is $\theta = \frac{1}{2}$. The null hypothesis, $H$, to be tested is $\theta \leq \frac{1}{2}$ against alternatives $\theta > \frac{1}{2}$. Because $\theta > \frac{1}{2}$ implies that the OLS estimator is better than the linearly restricted estimator in SMSE, a posterior probability of less than .50 that $H$ is true against the one-sided alternative that $\theta > \frac{1}{2}$ would make OLS the preferred estimator, i.e., the probability that the restricted estimator is "better" is less than the implicit probability that OLS has smaller MSE. An alternative interpretation of an $MSL < .50$ is that the posterior odds ratio favors OLS. However, in certain cases, e.g., ridge regression, the main purpose of the test is to identify those instances where the implicit posterior odds favor the biased estimator. Therefore, .50 is a lower bound on the appropriate significance level in a choice between OLS and a biased linearly restricted estimator when using SMSE.

For given $k$, the $F$-statistic in (7) can be used to compare OLS and SPR estimators at some significance level between .50 and 1.0, the exact level depending on the analyst's subjective choice. However, the more fundamental problem is to make a good choice of $k$, and the objective test permitted between OLS and any particular SPR estimator only provides a heu-

ristic basis for some strategies in this regard. As $k \to 0$, the $F$-statistic in (7) will approach zero, and the implied $MSL$ will go to 1.0 for the SPR estimator. As $k$ increases, the $MSL$ will decline for any given sample, and its expected value will decline in an *ex ante* sense. Taking the sample as given data, there will be a value of $k = k'$ at which $MSL = .50$. If $k > k'$, then $MSL < .50$ and posterior odds would favor the OLS estimator over the implied ridge estimator. This line of reasoning led the authors to consider strategies which take $k'$ as the upper bound on $k$ for a given sample. One possible exception to the existence of $k'$ is when the hypothesis $\beta = B$ cannot be rejected at the 50% level, in which case, $B$ would be an appropriate choice for the estimator of $\beta$.

The above reasoning provides a value of $k$ such that in a statistical sense the OLS and SPR estimators are equally good in SMSE, but the real task is to choose a value of $k$ which is near optimal, not just as good as OLS. On the basis of the previous arguments, the optimal choice of $k$ should lie between zero and $k'$. Because $k'$ is associated with the value $\theta_o = \frac{1}{2}$ in a test of the hypothesis $\theta < \theta_o$ and $MSL = .50$, the search for a near-optimal choice of $k$ can be focused on a value of $\theta_o < \frac{1}{2}$ and an $MSL = .50$ to obtain a value of $k < k'$. Such an estimator would have an MSE matrix for each value of $\theta_o$. If a scalar measure such as the determinant of this matrix, say $D(\theta_o)$, were graphed against $\theta_o$ on the interval $(0, \frac{1}{2})$, one would expect $D$ to decline as $\theta_o$ increased from zero, reach a minimum, and then continue increasing monotonically. If this relationship is

**Table 1. Extended**

| Factor Elasticity Estimates | | $R^2$ | Returns to Scale[b] | $k$ | $F^c$ | $\alpha$ |
|---|---|---|---|---|---|---|
| Energy | Misc. | | | | | |
| .0315 | .207 | | 1.000 | | | |
| .354 | .748 | .995 | .966 | | | |
| (.243) | (.283) | | | | | |
| .170–.170 | .173–.162 | .969–.952 | .957–.942 | .125–.225 | | |
| (.0194–.0121) | (.0138–.00859) | | | | | |
| .182 | .345 | .992 | .969 | .009 | 1.095 | .50 |
| (.163) | (.161) | | | | | |
| .0890 | .274 | .992 | .957 | .016 | 1.109 | .50 |
| (.139) | (.132) | | | | | |

fairly symmetric on the interval $(0, \frac{1}{2})$, $\theta_o = \frac{1}{4}$ should yield a relatively good estimator in SMSE.

The specific value $\theta_o = \theta^*$ is defined as that value (unknown) of $\theta_o$ such that in the test of the hypothesis $\theta \leq \theta_o$ and a search over $k$ yielding a marginal significance level of .50, $\theta_o = \theta^*$ gives the smallest MSE based on the SMSE criterion. Under the hypothesis $\theta \leq \theta^*$, choosing $k$ such that the $MSL = .50$ would imply a posterior probability of .50 that $k$ is too large, i.e., the associated value of $\theta$ is greater than $\theta^*$. If $\theta^*$ were known, $k$ determined in this way would play the same role for an upper bound on the optimal choice of $k$ as $k'$ did for choosing an upper bound on $k$ in the comparison of OLS with the biased estimator. The final choice of $k$ is an approximation where $\theta_o = \frac{1}{4}$ is used as an estimate of $\theta^*$ by appealing to symmetry in MSE on the interval $(0, \frac{1}{2})$.

However, let it be clear that the above argument has only heuristic appeal because it leans heavily on a logical basis that is appropriate for a fixed value of $k$ from sample to sample. But the actual ridge estimator is based on a value of $k$ which is determined by increasing $k$ incrementally for a given sample until an $MSL$ of .50 is reached under the null hypothesis. Only Monte-Carlo studies can ultimately determine how effective this strategy is within various structures for the parameters and the matrix $X$.

Critical values of the noncentral $F$-distribution for $\theta_o = \frac{1}{4}$ are not available in printed tables, but these critical values at the .50 level of significance are needed to implement the above method of choosing $k$. The computer

algorithm in the appendix of Goodnight and Wallace provides a satisfactory approximation to the upper tail of the noncentral $F$-distribution with degrees of freedom and $\theta$ as given data. This algorithm can be used in an efficient search procedure to find the value of $k$ associated with the SPR estimator which yields $MSL = .50$ for any given value of $\theta$, $\frac{1}{4}$ in particular.

Another strategy used in the application reported later implicitly utilizes the strong and weak MSE criteria jointly. The SMSE criterion may be more stringent than is feasible for a severely collinear data set in that the resulting biased estimates may suffer many of the same limitations as OLS. In other words, the low information content of the implicit design matrix, $X$, in (1) might prompt the analyst to resort to a more biased estimator than SMSE would justify, nevertheless defensible under the WMSE criterion. As mentioned earlier, the WMSE criterion cannot be applied logically in the SPR framework because the minimum expected-squared-distance measure would involve the extended parameter space. However, the fact that WMSE is unidimensional and a less demanding criterion than SMSE makes it clear that the logical upper bound of $\theta_o = \frac{1}{2}$ for SMSE might be quite conservative under the WMSE criterion.[6] This argument suggests what might be called a conjunctive SMSE/WMSE criterion for estimator selection.

Arguments given earlier suggested a strategy to delineate an upper bound on $k$ for an SPR

---

[6] As a practical matter, the critical values for the $F$-statistic are available for $\theta_o = \frac{1}{2}$ (Wallace and Toro-Vizcarrondo).

estimator under the SMSE criterion; it is the smallest value of $k$ at which $MSL = .50$ when testing the hypothesis that $\theta \leq \frac{1}{2}$. This value of $k$ is denoted $k_{max}$. Posterior odds would favor OLS over the SPR estimator if $k > k_{max}$ because $MSL$ would be less than .50. The conjunctive SMSE/WMSE criterion is to choose the SPR estimator associated with $k_{max}$ that gives the maximum defensible bias under the SMSE criterion, i.e., test the hypothesis that $\theta \leq \frac{1}{2}$ with $MSL = .50$. But under the WMSE criterion, the bias is expected to be rather modest.

Another consideration which justifies a greater amount of bias is the possibility of substantial specification error. This justification is particularly appropriate if the bias point is an a priori point estimate of the parameter vector instead of the null vector. The authors suspect that practitioners of ridge regression have a propensity to use larger $k$ values than might be justified by MSE criteria because of the pervasion of specification error in empirical work.

The highly collinear data set of Gorman and Toman provides an illustration of the results of applying the two methods discussed for a "good" choice of $k$, i.e., $\theta_o = \frac{1}{4}$ and $\theta_o = \frac{1}{2}$ with a $MSL = .50$. The shrinkage point is the null vector and the data involve ten independent variables and thirty-six observations. Fomby and Johnson present paired values of $k$ and the $F$-statistic in (7) which were used to make the necessary calculations. The results are not very accurate (for purposes here) because of the large interval on $k$ used by Fomby and Johnson. Calculations of $k$ were made by linear interpolation and reported accuracy is excessive, but the purpose is to provide comparative results.

The criterion based on an approximation to the optimal choice of $\theta_o$ under SMSE, which led to $\theta = \frac{1}{4}$, gave an $F$-statistic and value of $k$ equal to 1.007 and .0183, respectively; while the conjunctive SMSE/WMSE criterion ($\theta_o = \frac{1}{2}$) gave 1.056 and .0193, respectively. These values are far removed from those obtained by Hoerl and Kennard (1970a) based on the "stable region" using the ridge trace, where the $F$-statistic was around six and $k$ was in the interval .2 to .3. As a practical matter for this data set and shrinkage point, the choice of $\theta_o$ at $\frac{1}{4}$ or $\frac{1}{2}$ is not important. The authors suspect that this result emanates largely from the illogical shrinkage point. Strictly speaking, the SPR model is not appropriate here, but it was

applied for illustrative purposes under the artificial assumption that the null vector is a priori estimate of $\beta$.

## The Application

The study region chose for this application was the irrigated central California valley. County-level data were taken from the *1974 Census of Agriculture*. A Cobb-Douglas production function was hypothesized where output was value of crops harvested ($/county), and inputs were value of cropland harvested ($/county), labor expenditures ($/county), value of machinery ($/county), quantity of irrigation water applied (acre-feet/county), energy expenditures ($/county), and miscellaneous expenditures ($/county). OLS, ORR, and SPR parameter estimates are presented in table 1. Not surprisingly, the OLS estimates reveal the usual problems (unexpected signs, large standard errors) associated with a collinear data set.

*ORR results.* As originally suggested by Hoerl and Kennard (1970b), the ORR estimates were chosen using a ridge trace which is a subjective graphical method to evaluate the effects of the level of $k$ on the individual parameter estimates. The ORR estimates in table 1 are presented in interval form (corresponding to the bound of a "stable range," see fig. 1) to demonstrate the subjectivity of the method. Of course, nothing can be inferred about the MSE properties of the ORR estimates, even by Monte-Carlo studies, because of the very subjective criterion for choosing $k$.

*SPR results.* Fomby and Johnson's SPR estimator was applied in two versions. The first sets the factor elasticity priors at the zero vector ($B = 0$). Each prior factor elasticity estimate for the second version of the SPR estimator, $SPR(B = FS)$, was set equal to its respective factor share.[7] Factor shares are appealing as priors because there should be a tendency for factor shares to approximate factor elasticities

---

[7] The factor share of the $i$th factor is defined as total expenditure in the time period divided by value of output. In this application, factor shares are easily determined for those inputs expressed in dollar terms but must be fabricated for the water input. In particular, assuming constant returns to scale, the factor share for water is determined as the difference between one and the sum of the factor shares of those inputs which were expressed in dollar terms. Such estimates of the factor shares do violate the assumed covariance structure in (6) because $B$ is correlated with $y$ since the factor shares estimates involve $y$. However, since all factor shares were estimated using the mean value of $y$, this correlation is only of order $1/\sqrt{n}$.
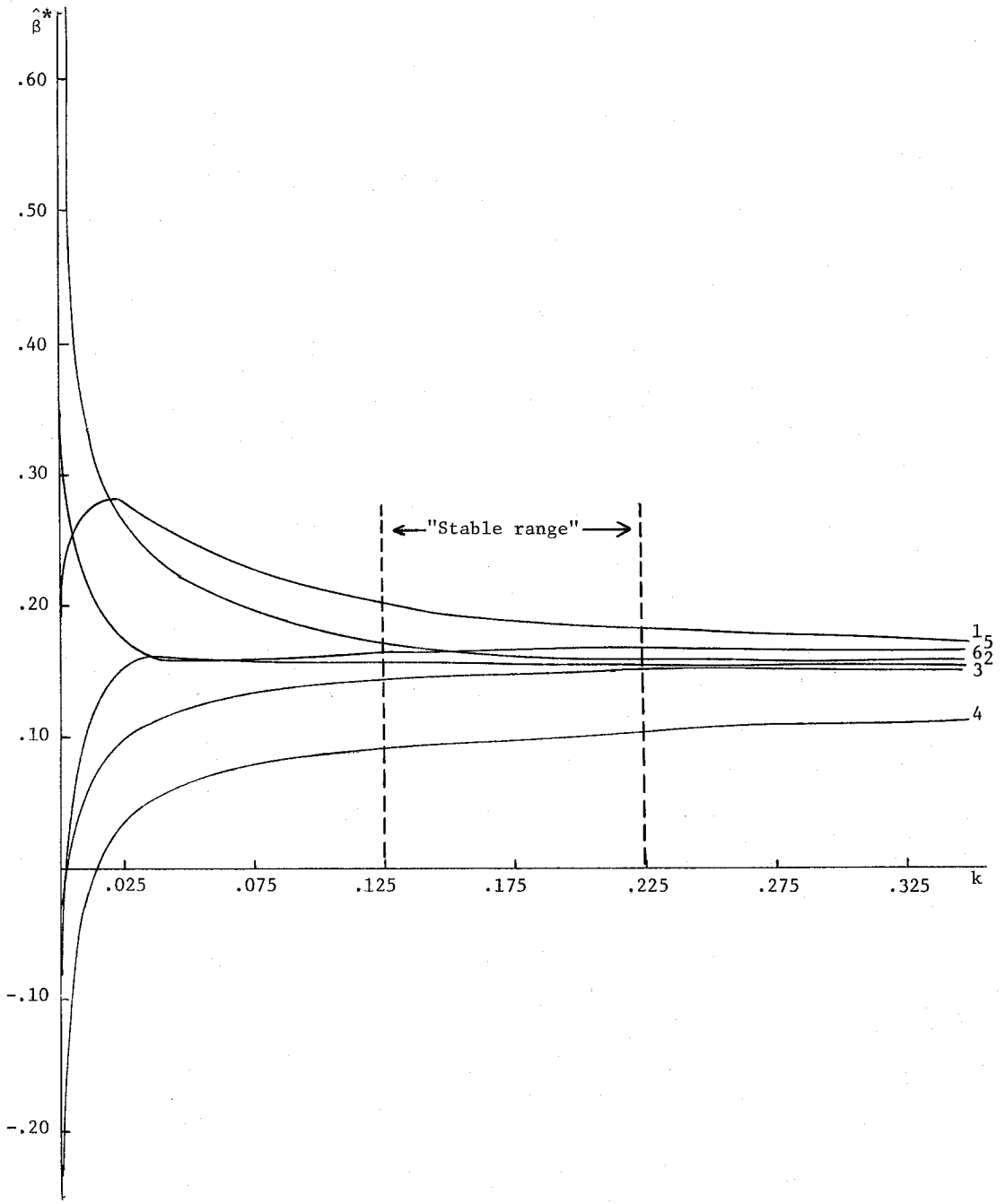
(Note: The numbers 1–6 conform to subscripts of $\hat{\beta}_i^*$ for corresponding variables defined in the text.)

**Figure 1.   Ridge trace of estimates for California Central Valley region**

in agriculture (Tyner and Tweeten). For *SPR* ($B = FS$) all coefficients were positive as expected, which was not the case for *SPR*($B = 0$). A marginal level of significance of .5 and $\theta_o = \frac{1}{2}$ under the SMSE criterion were used to determine the biasing parameter $k$ in both applications of the SPR estimator.[8]

It is informative to note that the $k$-values associated with the estimators for both *SPR* ($B = 0$) and *SPR*($B = FS$) are substantially lower than those corresponding to the "stable range" of the ridge trace (see fig. 1). The $k$-value for *SPR*($B = 0$) is .009; the comparable $k$-value for *SPR*($B = FS$) is .016. Yet, for ORR the lower-bound $k$-value appears to be in the vicinity of .125. It is conceivable that one might even choose a $k$-value as high as .225 when relying on visual interpretation of the ridge trace, a value about 14 times that for *SPR*($B = FS$). It would appear that "stability" of the ridge trace occurs at a point where too much bias is introduced according to SMSE. If one were to accord the prior information greater relative weight in the estimation, a lower significance level would be chosen which would allow a larger $k$-value and hence the introduction of greater bias. However, we find no justification for a significance level less than 50%.

## Concluding Remarks

The results of this paper and empirical application suggest four conclusions:

(a) The SPR estimator in conjunction with an objective sample criterion for choosing the level of bias appears to work well in empirical estimation of Cobb-Douglas production functions, especially when good a priori estimates of the factor elasticities are used to determine the point in parameter space toward which the estimation vector is biased.

(b) The ridge trace estimator-selection procedure appears to introduce too much bias in the estimation, at least more bias than can be justified by MSE criteria. There is always the possibility that such estimators are closer to the true parameter vector because of specification error in the model and errors in the independent variables, but these considerations defy any objective analysis.

(c) Standard errors under biased estimation methods are of limited value because the bias component of MSE is unknown. However, these standard errors do give a lower bound on root MSEs and therefore are useful information in a one-sided context. That is, while low standard errors associated with biased estimators do not imply small MSE, large standard errors demonstrate weak precision and thus large MSE. The SPR standard errors in table 1 do provide information in the case of machinery, water, and energy—each standard error is close in magnitude to the point estimate of the respective elasticity which indicates that these estimates are not very reliable. This unpleasant result should not be surprising in that multicollinearity reflects serious limitations of the sample data for which no estimation method can really compensate.

(d) Of course, any conclusions based on the application such as this are only suggestive of the properties of various estimators. A Monte-Carlo study in the context of production function estimation would be most helpful.

[*Received July 1986; final revision received May 1987.*]

## References

Brown, W. G., and B. R. Beattie. "Improving Estimates of Economic Parameters by Use of Ridge Regression with Production Function Applications." *Amer. J. Agr. Econ.* 57(1975):21–32.

Chipman, J. S. "On Least Squares with Insufficient Observations." *J. Amer. Statist. Assoc.* 59(1964):1078–111.

De Groot, M. H. "Doing What Comes Naturally: Interpreting a Tail Area as a Posteriori Probability or as a Likelihood Ratio." *J. Amer. Statist. Assoc.* 68(1973):966–69.

Fomby, T. B., and S. R. Johnson. "MSE Evaluation of Ridge Estimators Based on Stochastic Prior Information." *Commum. Statist.* 6(1977):1245–58.

Goodnight, J., and T. D. Wallace. "Operational Techniques and Tables for Making Weak MSE Tests for Restrictions in Regressions." *Econometrica* 40(1972):699–709.

Gorman, J. W., and R. J. Toman. "Selection of Variables for Fitting Equations to Data." *Technometrics* 8(1966):27–51.

Hoerl, A. E., and R. W. Kennard. "Ridge Regression: Applications to Non-Orthogonal Problems." *Technometrics* 12(1970a):69–82.

———. "Ridge Regression: Biased Estimation for Non-

[8] The *SPR*($B = FS$) was also applied using $\theta_o = \frac{1}{4}$. However, similar to the results obtained for the Gorman and Toman data set, the choice of $\theta_o$ had little impact on the final parameter estimates.

Orthogonal Problems." *Technometrics* 12(1970b):55–67.

Judge, G. G., T. A. Yancey, and M. E. Bock. "Properties of Estimators after Preliminary Tests of Significance when Stochastic Restrictions Are Used in Regression." *J. Econometrics* 1(1973):29–47.

Lindley, D. V. *Introduction to Probability and Statistics from a Bayesian Viewpoint.* London: Cambridge University Press, 1965.

Smith, V. K. "Improving Estimates of Economic Parameters by Use of Ridge Regression with Production Function Applications: Comment." *Amer. J. Agr. Econ.* 57(1975):723–24.

Swindel, B. "Good Ridge Estimators Based on Prior Information." *Commum. Statist.* 5(1976):1065–75.

Toro-Vizcarrondo, C. E., and T. D. Wallace. "A Test of the Mean Square Error Criterion for Restrictions in Linear Regressions." *J. Amer. Statist. Assoc.* 63(1968): 558–72.

Tyner, F. H., and L. G. Tweeten. "A Methodology for Estimating Production Parameters." *J. Farm Econ.* 47(1965):1462–67.

U.S. Department of Commerce, Bureau of the Census. *U.S. Census of Agriculture, State and County Statistics.* Washington DC, 1974.

Wallace, T. D. "Weaker Criteria and Tests for Linear Restrictions in Regression." *Econometrica* 40(1972): 689–98.

Wallace, T. D., and C. E. Toro-Vizcarrondo. "Tables for the Mean Square Error Test for Exact Linear Restrictions in Regression." *J. Amer. Statist. Assoc.* 64(1969): 1649–63.