DOCUMENT RESUME

ED 274 265		HE 019 705
AUTHOR	Brinkman, Paul T.; Niwa, Shelley	
TITLE	Statistical Cost Estimation in Higher Education: Some Alternatives.	
PUB DATE	31 Aug 83	
NOTE	58p.	
PUB TYPE	Information Analyses (070) -- Reports - Evaluative/Feasibility (142)	
EDRS PRICE	MF01/PC03 Plus Postage.	
DESCRIPTORS	Algorithms; *College Students; *Cost Estimates; Full Time Students; Higher Education; *Least Squares Statistics; Part Time Students; *Statistical Analysis; *Two Year Colleges	
IDENTIFIERS	*Econometrics; *Ridge Regression Analysis	

ABSTRACT

Recent developments in econometrics that are relevant to the task of estimating costs in higher education are reviewed. The relative effectiveness of alternative statistical procedures for estimating costs are also tested. Statistical cost estimation involves three basic parts: a model, a data set, and an estimation procedure. Actual data are used to assess whether the ridge techniques provide a viable alternative to the more familiar ordinary least squares approach within the collinear environment characteristics of translog models. The translog model that is used for the study generates marginal cost estimates for full-time and part-time students at two-year colleges. In every comparison conducted for the study, the ridge procedure was superior to the ordinary least squares approach. Of importance were ridge improvements in the precision and stability of estimated coefficients, since marginal cost estimates were a function of a set of coefficients. In addition, the ridge regression provided a means for data and model exploration. Comparing ordinary least squares and ridge estimates, and especially by examining ridge traces and variance inflation factors, can also promote understanding of the effects of multicollinearity (i.e., highly correlated explanatory variables) in a given situation. Algorithms and graphs are included. A five-page list of references concludes the document. (SW)

Statistical Cost Estimation in Higher Education: Some Alternatives

Paul T. Brinkman

and

Shelley Niwa

August 31, 1983

Statistical Cost Estimation in Higher Education:

Some Alternatives

The need to understand cost behavior is a perennial one in higher education. The driving force underlying the need may differ from one place to another and it certainly changes over time. For instance, it could be the task of evaluating the cost structure of programs at institutions competing for the same support dollars, or of ascertaining the conditions under which economies of scale might be expected to accompany enrollment growth. At the present time, there is much concern about what may happen to unit costs when enrollments decline. Whatever the motivation, the cost analyst is faced with the continuing challenge of finding new and, presumably, better ways of understanding costs.

As the extensive survey done by Adams, Hankins, and Schroeder (1978) makes quite clear, most of what passes for cost analysis in higher education is essentially a cost calculation of one sort or another. The ubiquitous average cost per student credit hour is a case in point. This cost figure can be calculated directly, given data on total costs and total credit hours. But other costs, such as the cost of an additional credit hour, are usually not directly calculable. Instead they must be estimated, using either statistical or accounting procedures. The statistical approach is far more common, and constitutes the focal point for this study. Specifically, the intent of this study is 1) to review some recent developments in econometrics that are relevant to the task of estimating costs in higher education, and 2) to test the relative effectiveness of alternative statistical procedures for estimating costs. The material included should be useful to researchers and analysts who have need to estimate higher-education costs, and to those who are interested in statistical cost estimation more generally.

Statistical cost estimation involves three basic parts: a model, a data set, and an estimation procedure. For present purposes, data-related issues will be dealt with summarily because these issues (for instance, the quality of financial data, the problem of finding acceptable output measures, and so on) are generally familiar ones for the higher-education analyst. By contrast, the thinking among econometricians regarding the structure of cost estimation models has undergone considerable development over the years, and may not be as familiar. Similarly, there have been developments in estimation procedures that have not received much attention in higher-education circles, but are worth considering here. Of course, this paper cannot offer the broad coverage appropriate to textbooks. The selection problem is made easier, though, by the fact that econometric thinking on models has converged somewhat, and because some of the preferred models can lead to statistical problems which in turn make certain estimation procedures more attractive. To put names to these matters, the models in question are translog cost functions, the statistical problem is multicollinearity, and the estimation procedure is ridge regression. Actual data will be used in assessing whether the ridge techniques provide a viable alternative to the more familiar ordinary least squares (OLS) approach within the collinear environment characteristic of translog models. The translog model that is used for the study generates marginal cost estimates for full- and part-time students at two-year colleges.

Cost Estimation

The behavior of costs in a particular industry, or more generally, the production structure of an industry, can be analyzed by estimating either a production function or a cost function. The procedures have been shown to be theoretically equivalent by Shepard (1953), for the single-product firm, and by McFadden (1978), for the multiproduct firm. When cost structure is the primary concern, estimating a cost function is the most direct approach. And, when the industry in question consists of multiproduct firms, as is certainly the case with respect to higher education, estimating a joint cost function offers the distinct advantage of making it relatively easy to model the structure of cost without imposing a priori restrictions on the structure of production--restrictions which are typically imposed when modeling the structure of multiproduct firms by estimating transformation (production) functions (Brown, Caves, and Christensen 1979).

The implicit form of a cost function can be written as

$$C = C(q; p; t) \tag{1}$$

where C is total cost, q is output, p is the input price, and t is a set of technological conditions that may have some effect on the relationship between C and q (McFadden 1978). Under theoretically ideal conditions (that is, intent to minimize costs coupled with full knowledge of how to do so), the cost function specifies the minimum cost for a given level of output. Whether such conditions ever hold entirely is doubtful. It is certainly unlikely that they hold for higher education; Bowen (1980) makes this point rather emphatically. (Pauly [1978] makes the same point for hospitals.) Most estimated cost functions, then, actually represent average rather than minimizing behavior. Cohn (1979) refers to such cost functions as "approximate."

Developing an explicit form for the cost function in a given situation is the essence of the modeling problem. Preference for particular types of explicit functional forms has changed over the years. Some of the earliest examples of cost functions date from Dean's studies in the 1930s of retail trade stores (reprinted in Dean 1976). These early efforts typically employed simple additive models at best. For instance, Yntema (1940) used the function

$$C = a_0 + a_1 q + a_2 q^2 \tag{2}$$

to estimate marginal costs for the steel industry.

With the advent of computers in the post WWII era, and the growing interest in econometrics, functional forms gradually became more complex, and, one might say, more thoughtful. That is, more attention was paid to the intervening or secondary variables that might influence the cost-output relationship, to the form of the function (particularly as it related to economic theory), and to the form of the variables (for example, raw versus logarithmically transformed) included in the estimating equation. Johnston's 1960 textbook on statistical cost estimation provides an excellent review of developments to that point (including criticisms of the various procedures). A review of the early period can also be found in Dean (1976) and Walters (1963).

Despite the variety of functional forms employed, virtually all cost functions up through the early 1970s had one important feature in common. They all imposed a priori restrictions on the cost and production structure. For example, in that simplest of forms shown above as equation 1, one restriction imposed (apart from consideration of price and technical conditions) is that marginal cost, the change in total cost (c) associated with an additional unit of output (q), must be constant; it can only be the estimated value of the parameter a^1, regardless of the value of q--or anything else for that matter. Other functional forms were less restrictive, but it was not until the 1970s that so-called flexible forms, which impose few if any restrictions, began to be used with some frequency. Diewert (1974) reviews several of the flexible forms that are designed for joint cost functions (where more than one type of output is involved). Griffin (1982) compares the approximation characteristics of three flexible forms: the generalized Leontief, the translog, and the generalized square-root quadratic. Of these forms, the translog function proposed by Christensen, Jorgenson, and Lau (1971; 1973) appears to be the most widely adopted (for example, see Brown, Caves, and Christensen [1979], Cowing and Holtmann [1983], and Spady [1979]). A variety of discussions and applications of the translog cost function can be found in Smith (1982).

The translog joint cost function for ℓ outputs, m inputs, and n technical conditions can be written

$$\ln C = a_0 + \sum_{i=1}^{\ell} a_i \ln q + \sum_{j=1}^{m} \beta_j \ln p + \sum_{k=1}^{n} \gamma_k \ln t + \tfrac{1}{2}\sum_{i=1}^{\ell}\sum_{j=1}^{\ell} \varepsilon_{ij} \ln q_i \ln q_j$$

$$+ \tfrac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\gamma_{ij}\ln p_i \ln p_j + \tfrac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\rho_{ij}\ln t_i \ln t_j + \tfrac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{m}\nu_{ij}\ln q_i \ln p_j$$
$$+ \tfrac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{n}\phi_{ij}\ln q_i \ln t_j + \tfrac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n}\omega_{ij}\ln p_i \ln t_j \qquad (3)$$

where $s_{ij}=s_{ji}$, etc. The expression (3) has one neutral scale parameter (a_0), $l + m + n$ first order parameters ($\alpha_i, \beta_i, \gamma_i$), and $(l+1)(l/2) + (m+1)(m/2) + (n+1)(n/2) + lm + ln + mn$ second order parameters. The only restriction imposed in using this form is the regularity condition that, if the function is to be understood as a cost function in the strict sense, then the function must exhibit homogeneity of degree one in factor prices (that is, total cost must change in the same proportion and direction as a change in factor prices). Otherwise there are none of the typical restrictions on the cost and production structure of the firms analyzed. Indeed, what were once a priori impositions on structure now become testable hypotheses. The second order logarithmic terms for the output variables allow for two inflection points in the estimated cost curves, thus allowing for economies or diseconomies of scale, while the complete set of interaction terms removes any separability assumptions from the model. As Brown, Caves, and Christensen (1979) have shown, imposed restrictions such as homogeneity and separability of output can make a significant difference in the results (parameter estimates) of the analysis.

Clearly, then, in the absence of a priori knowledge about the structure of production, there is good reason to adopt a flexible functional form, such as the translog model shown above, that lets the data speak for themselves. At the same time, models such as the translog are prone to the estimation problem known as multicollinearity, in which correlation among the explanatory variables can hide or distort their true relationship to the dependent variables (for example, see Cowing and Holtmann [1983]). Zero-order correlations tend to be quite high ($r>.95$) between a value in logarithms, its square, and related interaction terms. In estimating marginal costs in higher education, the situation may be exacerbated by what one might call "natural", as opposed to "model-induced" multicollinearity. That is, the key explanatory variables one might typically consider in estimating cost functions for colleges and universities tend to be collinear. For example, there would be utility in being able to compare the marginal costs of lower versus upper division students, but these two enrollment levels tend to vary together (Allen and Brinkman 1983). This naturally occurring phenomenon, when combined with the tendency of the translog model to generate highly collinear explanatory variables, creates a situation in which multicollinearity is likely to be a serious problem.

Thus, in considering how best to proceed in statistically estimating costs for higher-education institutions, the analyst is faced with a dilemma. What has emerged in econometrics as the preferred form for the joint cost function, a highly flexible translog model, brings with it the threat of severe multicollinearity, capable of distorting the very results whose integrity is protected by the flexibility of the model. Before considering the appropriateness of estimation techniques designed to get around this dilemma, we look more closely at the problem of multicollinearity itself.

Multicollinearity

When explanatory variables are highly correlated, regression coefficients estimated by applying an ordinary least squares criterion suffer from a number of problems. These include

1) The precision of estimation falls so that it becomes very difficult if not impossible to disentangle the relative influences of the various variables. The loss of precision has three aspects: specific estimates may have very large errors; these errors may be highly correlated with one another; and the sampling variances of the coefficients will be very large.

2) Investigators are sometimes led to drop a variable incorrectly from an analysis when its coefficient is not significantly different from zero due to collinearity, rather than to the absence of a relationship with the dependent variable.

3) Estimates of coefficients are very sensitive to particular sets of sample data; the addition or deletion of a few observations can sometimes produce dramatic shifts in the coefficients.

4) Estimates of coefficients are very sensitive to the addition or deletion of a variable in the model.

The multicollinearity problem is discussed extensively in the literature of econometrics and statistics. These discussions may be roughly divided into two broad categories: (1) those dealing with its nature and potential consequences (e.g., Blalock, 1963, 1964; Darlington, 1968; Goldberger, 1964; Gordon, 1968; Johnston, 1972; Kumar, 1975; Leamer, 1973; Wichers 1975); and (2) those discussing strategies for dealing with the problem such as variable selection (Gorman and Toman, 1966; Gunst and Mason, 1977; Hocking, 1976), reduction to canonical form (Baranchik, 1970; Chatterjee and Price, 1977), and biased estimation procedures. Techniques discussed under biased estimation procedures include Stein estimators (Mallows, 1973; Mayer and Wilkie, 1973; Sclove, 1968), Bayesian estimators (Leamer, 1973; Lindley and Smith, 1972; Theil, 1963), ridge estimators (Bulcock, Lee, and Luck, 1977; Darlington, 1979; Dempster, Schatzoff and Wermuth, 1977; Hoerl and Kennard, 1970; Marquadt, 1970; Vinod, 1978), and generalized inverse or fractional rank estimation (Hemmerle, 1975; Marquadt, 1970).

In strict mathematical terms, collinearity is said to exist if there are one or more linear dependencies between predictor variables (Silvey, 1969). Less restrictive definitions (e.g., Willan and Watts, 1973) suggest that collinearity exists when linear relationships hold approximately. Farrar and Glauber (1967) define collinearity as a statistical rather than a mathematical condition). Viewed from this latter perspective, the task becomes one of identifying the degree of collinearity and its effects.

Several indices are available for describing both the degree of ill-conditioning in the data and its effects on estimated coefficients. These include variance inflation factors (Marquadt, 1970), the mean squared error of the estimated coefficient vector (Hoerl and Kennard, 1970), the squared length of the estimated coefficient vector (Hoerl and Kennard, 1970), the forecasting error variance (Johnston, 1972), and the ridge trace of the standardized regression coefficients (Hoerl and Kennard, 1970). Statistical tests for the degree of ill-conditioning in the data have also been suggested by Bartlett (1950), Farrar and Glauber (1967), Haitovsky (1969), and Wichers (1975). Chatterjee and Price (1977) demonstrate how the method of principle components analysis can be used to locate collinear relationships.

We have found two of the above indices to be particularly useful: the variance inflation factors (VIFs) suggested by Marquadt (1970), and the ridge trace developed by Hoerl and Kennard (1970). The VIFs for a particular model are readily obtained from the diagonal of the inverse of the correlation matrix of the predictor variables, $(X'X)^{-1}$. More precisely, we can see from the equation

$$V(\tilde{\beta}) = \sigma^{2}(X'X)^{-1} \tag{4}$$

that the precision of an estimated regression coefficient is measured by its variance which is proportional to $\tilde{\sigma}^2$, the error variance of the regression model. The constant of proportionality for a given $\tilde{\beta}_i$ is taken from the i-th term of the principal diagonal of $(X'X)^{-1}$. The constant of proportionality is referred to as the "variance inflation factor" for $\tilde{\beta}_i$ (Marquadt 1970).

It is easily demonstrated that the VIF for a given B_i is equal to $1/(1-R^2)$, where R^2 is the square of the multiple correlation coefficient from the regression of the i-th explanatory variable on all other explanatory variables in the equation. Hence as R^2 tends toward 1.0, indicating the presence of a linear relationship between the explanatory variables, the VIF for $\tilde{\beta}_i$ tends to infinity as does the associated variance estimate. The estimated variance for any specific coefficient may then be written as

$$V(\tilde{\beta}_i) = \tilde{\sigma}^2 (1 - R_i^2)^{-1} \tag{5}$$

It has been suggested that values for VIFs greater than 10.0 are indications that multicollinearity may be causing estimation problems (Chatterjee and Price, 1977; Marquadt and Snee, 1975). A VIF of 10.0 for a particular explanatory variable X_i, implies a multiple correlation of .95, when X_i is regressed on the other explanatory variables in the model.

The second method for detecting multicollinearity, the ridge trace method, flows directly out of the ridge analysis which will be employed in this study as an alternative to the OLS procedure. The ridge trace method will be discussed and demonstrated in the sections that follow.

Ridge Estimators

As previously noted, the particular class of biased estimators employed in the present study are the ridge estimators first proposed by Hoerl and Kennàrd (1970). Ridge estimators were chosen for three reasons. First, they are designed to be more reliable than the least squares estimator in the presence of an ill-conditioned data matrix. Second, the "ridge trace" conveys both the degree of ill-conditioning, and the imprecision inherent in interpreting collinear data. Third, ridge-type solutions provide estimates under varying sample and collinearity conditions which appear to be at least as good if not better than available alternatives (cf., Dempster, Schatzoff and Wermuth, 1977; Hoerl, Kennard and Baldwin, 1975; McDonald and Galarneau, 1975; Vinod, 1978). With this in mind, we will proceed with the development of the general class of ridge-type estimators.

The least squares estimate of B may be written in a more generalized formula as

$$\hat{\beta}(k) = (X'X + kI)^{-1}X'Y \tag{6}$$

where "k" is a scalar. In the least squares estimator, k=0 so the above equation reduces to the familiar

$$\hat{\beta} = (X'X)^{-1}X'Y \tag{7}$$

When k>0, $\hat{\beta}(k)$ is a "biased" estimator of the true unknown coefficient vector. However, it can be shown that by allowing a little bias into the system, one obtains an estimator with a smaller total mean squared error value than by using OLS procedures. This may be stated analytically as

$$E[(\hat{\beta}(k) - \beta)'(\hat{\beta}(k) - \beta)] < E[(\hat{\beta} - \beta)'(\hat{\beta} - \beta)]$$

Using ridge estimation, then, entails adopting minimum mean square error (MSE) as a general criterion in place of the customary ordinary least squares criterion.

Procedures for Selecting k

Choosing a value for k is critical in using a ridge estimator. Hoerl and Kennard's 1970 article introducing ridge regression to the scientific community suggests that guidelines for selecting a particular value for k are straightforward:

1. At a certain value of k, the system will stabilize and have the general character of an orthogonal system.

2. Coefficients will not have unreasonable absolute values with respect to factors for which they represent rates of change.

3. Coefficients with improper signs at k=0 will have changed to have proper signs.

4. The residual sum of squares will not have been inflated to an unreasonable value. It will not be large relative to the minimum residual sum of squares or large relative to what would be a reasonable variance for the process generating the data. (p. 65)

Unfortunately, the guidelines are no more than general signposts. In reality, the optimal value of k cannot be determined with certainty (i.e., in terms of a closed-form solution) because it depends on the unknown parameter vector B and the unknown error variance σ^2. In practice, k must be determined subjectively or estimated from the data (Mayer and Wilkie, 1973; Judge et al 1980). This characteristic of k is at the root of the difference of opinion regarding the value of ridge estimation. Some would argue that the reduction in mean square error gained by the introduction of bias into the system has little value because of our inability to select the amount of bias in an optimal manner. To put it another way, we cannot evaluate any gain in accuracy for a particular problem without knowing the true values of the coefficients. Proponents of ridge techniques counter by claiming that one can use the <u>data</u> in a particular problem to help select a value of k that will produce an estimator superior to OLS. The rejoinder to that argument is that the resulting estimates are stochastic, while OLS estimators and ridge estimators based on a fixed k are nonstochastic. Thus it is argued that selecting a value for k based on sample data makes it improper to apply standard statistical tests (such as t-scores) in the ridge environment (Darlington 1978; Judge et al 1980). We will return to this important problem later.

A number of specific techniques have been developed for estimating the value of k. Each has its proponents. In the final analysis, the choice of technique depends on the assumptions the investigator is willing to make. Two techniques for estimating k were incorporated in this study: ridge trace (Hoerl and Kennard, 1970); and, the harmonic mean (Hoerl, Kennard, and Baldwin, 1975).

The ridge trace approach was used because it provides a description of (1) the severity of ill-conditioning in the data; (2) how collinearity conditions affect estimation; and (3) how increasing the degree of bias introduced into the regression model affects coefficient estimation. The approach entails introducing a specific amount of bias into the model and plotting the resultant biased coefficients against the bias value. The primary drawback of the ridge trace approach is that it does not provide a point estimate (i.e., a closed form solution) for k. The technique requires the analyst to visually examine the plot and make a subjective decision about where (i.e., at what value of k) the solution appears to stabilize. Because the technique makes no assumptions about the nature of the closed form solution but allows the analyst to plot the consequences of introducing all feasible bias values, ridge trace plots can simultaneously provide and depict the relationship between all feasible closed form solutions. The ridge trace procedure is formally developed in Appendix III.

The harmonic mean approach was used because (1) it provides a relatively simple procedure for calculating the bias parameter, i.e.,

$$k = p\tilde{\sigma}^2/\tilde{\beta}'\tilde{\beta} \tag{8}$$

(2) its assumptions are simple and relatively easy to understand making the procedure readily employable by the lay analyst; and (3) the procedure provides estimated values for k which appear to have optimal properties under varying conditions of collinearity. The rationale for the approach is presented in Appendix IV.

Ridge vs. OLS--Model for Testing

The results of numerous studies are available to the reader interested in theoretical comparisons of the merits of OLS versus ridge estimation procedures (Dempster, Schatzoff and Wermuth, 1977). Our concern here is not with attempts at proving that ridge estimators are always to be preferred when predictor variables are highly correlated. Arguments by Judge et al (1980) and many others clearly demonstrate that ridge estimators have many undesirable properties and, furthermore, lack many of the desirable properties claimed for them. Our intent here is to show how ridge estimators can be useful in a practical context--for providing both insights into the effects of multicollinearity and a viable means of mitigating some of those effects. Hence, rather than set up an artificial data set to use as a basis for comparing the results of OLS versus ridge estimates, actual data relating to a typical cost estimation problem in higher education are used in what follows. The former approach has the advantage of permitting knowledge of the true coefficients and true variances, around which comparisons could be made. It seemed more important, however, to show what working with ridge estimators is like under the normal condition of uncertainty.

The cost estimation problem to be used for testing purposes has been reported on earlier (Brinkman 1983). In that study, marginal costs for full-time and part-time students were estimated and compared for several standard expenditure categories at public two-year colleges. A translog joint cost function was developed and subsequently estimated by a ridge technique. For present purposes, the same model and data set will be used to compare ridge and OLS results. Initially, then, the testing procedure will in effect be looking behind the scenes to show what, if anything, was gained by using ridge regression rather than OLS. Additional comparisons will be made between OLS and ridge (at two different values of k) using progressively smaller samples (randomly chosen subsets of the original full sample), multiple samples of the same (small) size, sets of coefficients derived from one sample to estimate marginal costs for another sample, and the reestimation of a sample following the removal of outlier cases. Since the true values of the parameters are not known, the comparisons can only be suggestive, and not definitive. Nonetheless, by observing the results of working with real data, the potential user of the ridge procedures may gain some useful insights as to their practical utility.

The cost function used to estimate marginal costs at two-year colleges contains the following variables. The dependent variable is total instructional expenditures (in the original study, expenditures for student services and for educational and general purposes were also analyzed). The independent variables include: as outputs, the number of full-time students (FTS), the number of part-time students (PTS), and the number of non-credit students (NCS); as input price, the salaries paid to full-time faculty (SAL); and as technological conditions, the proportion of degree earners (DEG), the proportion of relatively high-cost programs (HCP), and the system-status of the campus (CSS). The last variable listed was in dummy form (1 or 0 depending on whether the institution had independent ·rus or was part of a system), and was not interacted. The single price variable was not interacted either, in the absence of any substitution possibilities, but was in logarithmic form. All other variables were logged, squared, and interacted in standard translog form.

The data are taken from the 1979-80 Higher Education General Information Surveys, except for the data on non-credit enrollments which came from the American Association of Community and Junior Colleges' directory. The full sample consisted of all institutions that had complete data and were not a branch campus, except for a handful of outliers which were removed from the sample. The full sample consisted of 779 institutions, or about 75 percent of all public two-year colleges in 1979-80.

Results

Table 1 shows the estimated coefficients for the full sample using OLS and two ridge estimates. In looking at the coefficients (B) and their standard errors (SE) as estimated by OLS, one finds remarkably little outward evidence of collinearity. Roughly half of the coefficients are statistically significant (B/SE > 1.96, as measured using OLS estimates), and most of their signs are plausible. One unexpected result is the sign on the estimated coefficient for FTS. Since what we know about the costs of instruction suggests that the number of full-time students is usually the most important single determinant of total costs, it is surprising that the coefficient on full-time students (FTS) should be negative and statistically insignificant, instead of being positive and significant.

Despite what is suggested by the OLS coefficients, however, the variables are in fact highly collinear. The variance-inflation-factors (VIF) make this quite clear. Since a VIF of 1 is equivalent to orthogonality, it is clear that only a couple of variables, SAL and CSS, are relatively free of collinearity. The high VIFs on the remaining variables indicate a high degree of imprecision in the estimated coefficients.

The zero-order correlations among the explanatory variables may lend insight into sources of the collinearity problem. Table 2 shows the correlations for a subset of the variables in the model. The table clearly shows the "model-induced" collinearity discussed earlier. Some variables have more than a .99 correlation with their squares, and some interaction terms have well over a .90 correlation with one or both of the interacted terms. By contrast, the "natural" collinearity among the variables shown only runs as high as .688 (FTS with PTS), and is usually much less than that. Of course, zero-order correlations typically will understate the degree of collinearity in the system, as they reveal nothing of the collinearity which is due to combinations of variables. The VIFs do reflect the latter source of collinearity, however, and thus provide better insight into the extent and location of multicollinearity in a given model.

Table 1

Regression Results Using Alternative Estimators

	B	SE	VIF	B	SE	VIF
		FTS			(FTS)2	
OLS	-.319	.231	464.76	.097	.017	562.68
K=.003	.190	.069	44.33	.035	.005	39.69
K=.200	.225	.005	.24	.017	.000	.25
		PTS			(PTS)2	
OLS	.266	.131	388.62	.032	.006	160.91
K=.003	.030	.047	50.77	.019	.003	41.90
K=.200	.029	.003	.21	.004	.000	.31
		(FTS)(PTS)			SAL	
OLS	-.072	.016	680.40	.623	.063	2.04
K=.003	-.010	.004	52.03	.640	.062	1.94
K=.200	.008	.000	.11	.617	.043	.94
		NCS			(NCS)2	
OLS	-.028	.036	196.17	.008	.001	15.24
K=.003	-.038	.019	55.83	.007	.001	13.39
K=.200	-.007	.001	.19	.002	.000	.74
		DEG			(DEG)2	
OLS	-.985	.304	206.84	.136	.033	100.61
K=.003	-.550	.144	46.49	.108	.019	33.75
K=.200	-.089	.010	.20	.001	.002	.31
		HCP			(HCP)2	
OLS	-.229	.114	143.76	.023	.007	11.77
K=.003	-.218	.066	47.52	.022	.006	10.35
K=.200	-.013	.005	.25	.007	.002	.80
		(FTS)(NCS)			(FTS)(DEG)	
OLS	-.008	.005	194.75	.056	.038	243.48
K=.003	-.003	.003	55.57	.007	.015	37.16
K=.200	.001	.000	.21	.029	.001	.32
		(FTS)(HCP)			(PTS)(NCS)	
OLS	.004	.016	153.93	.002	.003	83.49
K=.003	.022	.009	46.53	-.001	.002	34.70
K=.200	.011	.001	.27	-.000	.000	.43
		(PTS)(DEG)			(PTS)(HCP)	
OLS	-.035	.023	180.15	.013	.011	88.49
K=.003	-.021	.011	40.99	-.004	.007	38.28
K=.200	-.001	.001	.30	-.000	.001	.26

	(NCS)(DEG)			(NCS)(HCP)		
OLS	-.002	.007	93.98	.007	.003	26.13
K=.003	-.001	.005	39.17	.005	.003	19.65
K=.200	-.001	.000	.33	-.000	.001	.63

	(DEG)(HCP)			CSS		
OLS	.012	.023	80.61	.005	.019	1.03
K=.003	.011	.015	34.68	.007	.019	1.02
K=.200	-.004	.002	.36	.018	.016	.70

Table 2

Zero-Order Correlations for a Subset of Variables

	TC	FTS	(FTS)2	PTS	(PTS)2	(FTS)(PTS)	HCP	(FTS)(HCP)
TC	1.000							
FTS	.924	1.000						
(FTS)2	.927	.997	1.000					
PTS	.744	.688	.694	1.000				
(PTS)2	.781	.723	.732	.990	1.000			
(FTS)(PTS)	.885	.880	.887	.945	.961	1.000		
HCP	.109	.070	.054	-.068	-.064	-.023	1.000	
(FTS)(HCP)	.458	.449	.434	.205	.221	.318	.914	1.000

Just how effective the VIFs are in pinpointing multicollinearity is a matter of controversy (see Judge et al 1980 for a discussion).

Table 1 also shows effects of introducing bias into the estimating system by means of the ridge procedure. Notice first the rapid reduction in VIFs for variables with extreme VIF values. Also note that the regression coefficients and their standard errors also change, depending on the amount of bias (the value assigned to k) introduced. We will discuss below the issues surrounding the choice of the amount of bias. For now, it is enough to note that K=.003 is a relatively small amount and K=.2 is a relatively large amount of bias for this particular situation, and that both values are plausible choices.

As shown in table 1, even at K=.003, the sign on FTS has switched from negative to positive, and the standard error has become small relative to the estimated coefficient. Additional bias does not change these desirable new features. Not all changes induced by the bias are as welcome. According to the OLS estimate, the sign on (FTS)(PTS) is negative. This result is certainly theoretically acceptable, as it indicates the existence of economies of scope: it is less expensive to instruct full-time and part-time students together, i.e., at the same institutions, than to do so separately. Unfortunately, one might say, the introduction of increasing amounts of bias into the estimating procedure eventually leads to a sign switch on (FTS)(PTS). Since this opposite result is also theoretically plausible, we are left with no substantive basis for arguing on behalf of either result. In other words, in the absence of a clear theoretical direction, it is difficult to feel comfortable with a sign change (especially when the standard errors are relatively small in both cases). Particular trouble with this variable might have been expected as it had the highest VIF of any variable in the model.

The ridge trace procedure allows us to "see" the effects of introducing bias into the system. Figure 1 shows the trace (the value of the standardized coefficient) as a function of the value of k for each of several variables. Figures 2 and 3 show the traces for additional variables in the model. The traces of the coefficients with high VIFs are much more sensitive to the amount of bias in the system. The reason why the ridge procedure is attractive, when collinearity is a problem, is the way in which it stabilizes or "tames" (Kennedy 1979) badly behaving coefficients. In other words, with enough bias, the coefficients of highly collinear variables can be made to behave as consistently as the coefficients of non-collinear variables. In part this is accomplished by reducing the absolute magnitude of the coefficients with respect to their OLS values. Interestingly, variables that are collinear and of little consequence in model have their coefficients reduced in magnitude to such an extent that they are, in effect, removed from the model.

While the behavior of a particular coefficient is of some interest, marginal cost estimates in the present context are the result of a combination of coefficients. Specifically, the marginal cost of an output q is equal to the first partial derivative of the estimated cost function with respect to q, multiplied by the estimated value of total cost for a particular value of q, divided by that value of q, or

$$MC_q = a_q \hat{C}/q \qquad (9)$$

where $a = \partial\hat{C}/\partial a$

In the present case, where the outputs of concern are full-time (FTS) and part-time (PTS) enrollments, the respective marginal cost calculations are as follows:

$$MC_f = (a_0 + 2a_1 FTS + a_2 PTS + a_3 DEG + a_4 HCP + a_5 NCS) \cdot \tilde{C}/FTS \qquad (10)$$

$$MC_p = (b_0 + 2b_1 PTS + b_2 FTS + b_3 DEG + b_4 HCP + b_5 NCS) \cdot \tilde{C}/PTS \qquad (11)$$

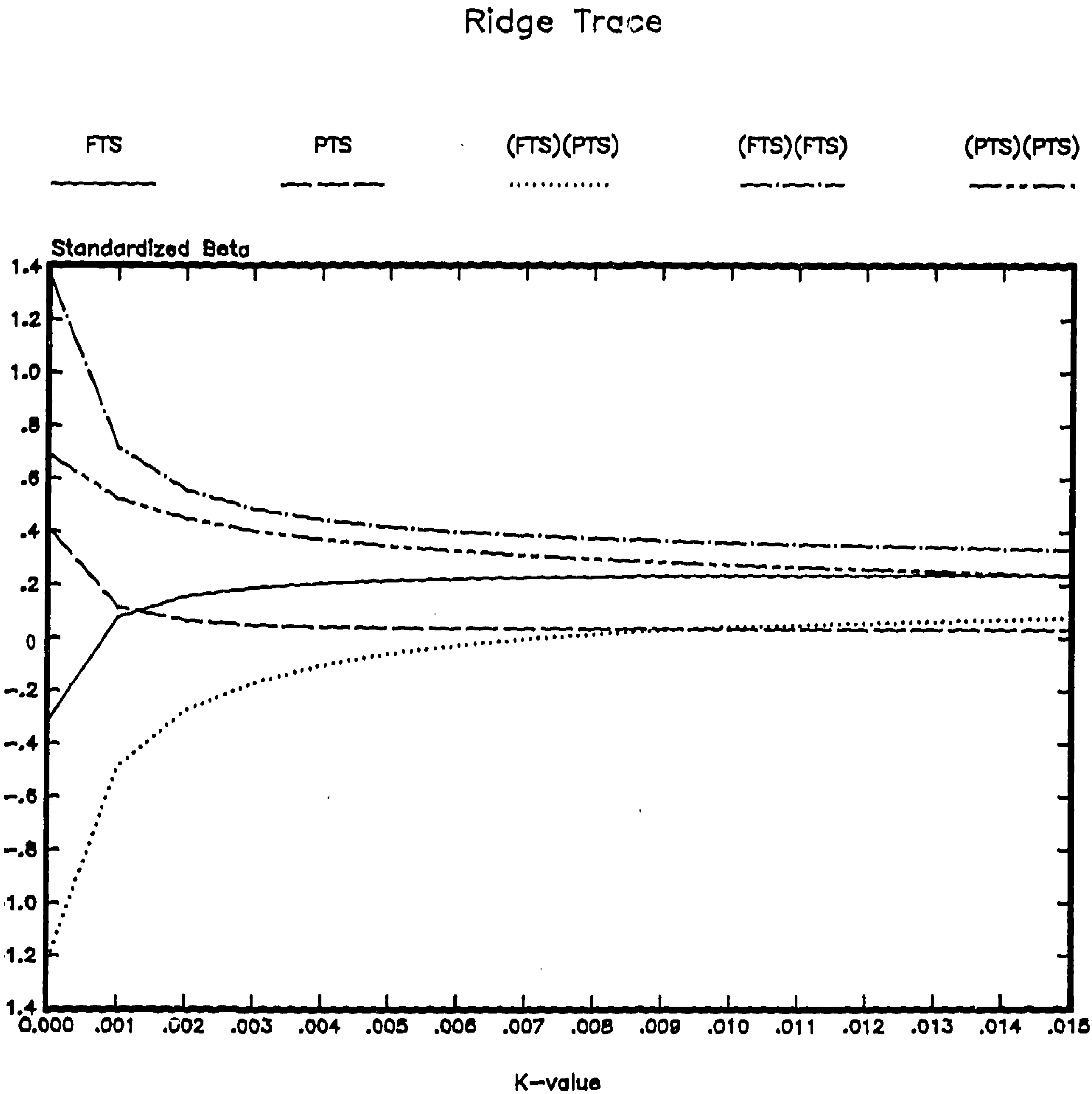

Fig.1. Ridge trace for key output variables, full sample

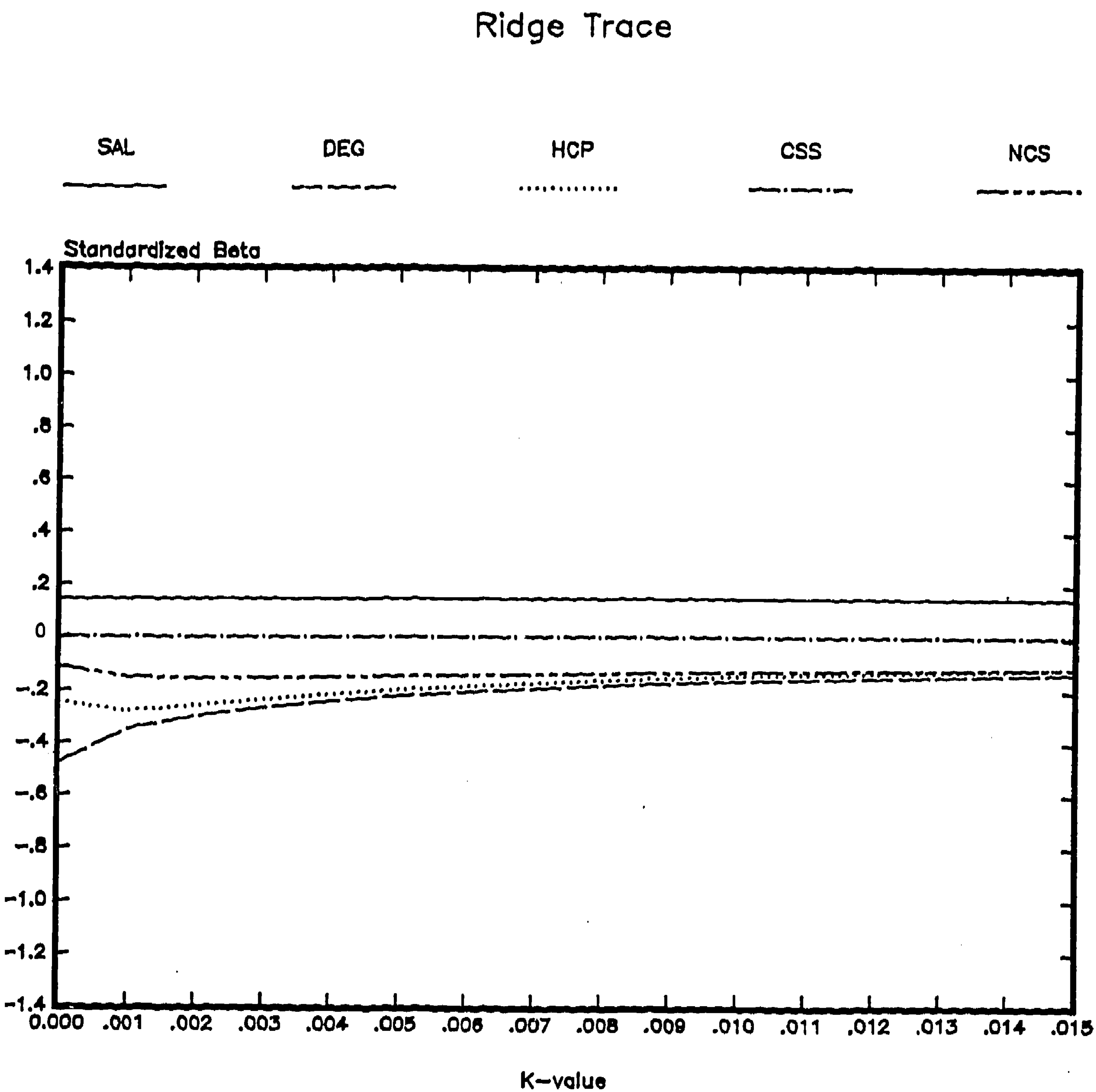

Fig.2. Ridge trace for subset of control variables, full sample

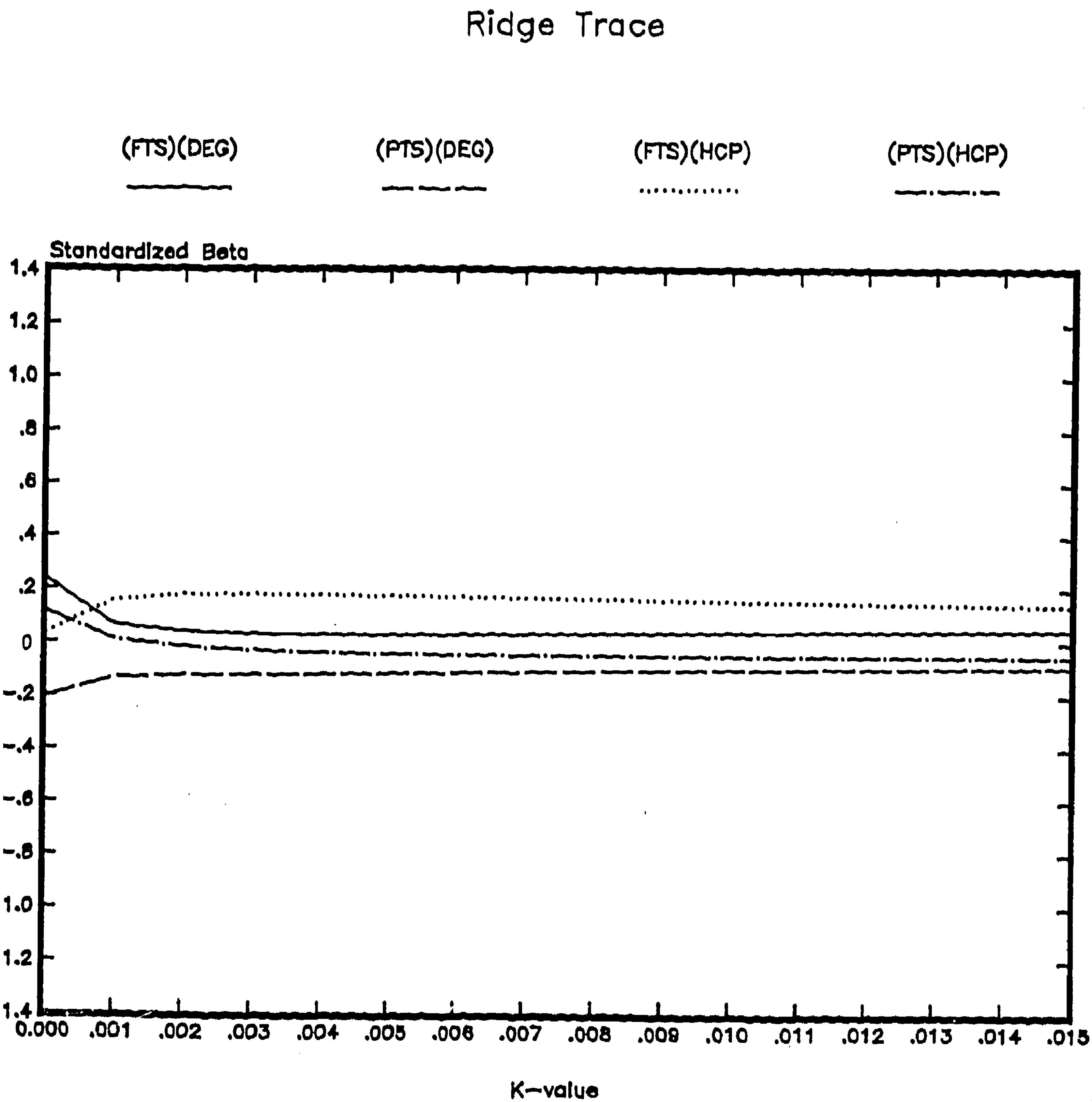

Fig.3. Ridge trace for subset of interaction variables, full sample

With respect to the concerns being addressed in this paper, it is perhaps worth noting the obvious: the accuracy of a marginal cost estimate for the model at hand will depend on the accuracy of six estimated coefficients. Table 3 shows the results of using these formulas with the three sets of coefficients shown in table 1. The calculations have been done for four levels of output in order to contrast the respective estimated costs across the observed range of enrollment. The category "small institutions" refers to institutions lying within the smallest five percent of those in the sample (as measured by enrollment). Data on 10 such institutions, randomly chosen, were averaged to create a data set for a "typical" small institution--284 full-time and 221 part-time students. In a similar fashion, data for a typical large institution were created--4,665 full-time and 12,885 part-time students. Between these extremes, two types of middle-range institutions are also represented in table 3. Section C shows the results of using raw enrollment means for the entire sample--1645 full-time and 2840 part-time students--to represent one such institution. Section B shows the results of using the logarithmic enrollment means for the entire sample--1150 full-time and 1366 part-time students--to represent the other. The raw data distributions are positively skewed, so the means of the logarithmic data are smaller. Fully two-thirds of all the institutions in the sample have enrollments equal to or less than the raw mean values.

In order to evaluate marginal costs at these various enrollment levels, values for the other independent variables in the model must also be selected. For the results shown in table 3, the following conditions were imposed: the raw mean values for percent of degree completion (29%) and percent of high cost programs (36.2%) were used in all sections; with respect to noncredit enrollment, the average of actual values was used for section A (165 students); the log mean value for section B (354 students), and the raw mean value for sections C and D (4335), and for faculty salaries the log mean value was used for section B ($18,215), the raw mean value for section C ($18,578), and the average of actual values for sections A ($13,625) and D ($23,090). Neither degree completion nor

Table 3
Marginal Costs of Instruction
Alternative Estimates
Full Sample

Institutional Size and Student Type	OLS	Ridge K=.003	Ridge K=.2
A. Small			
FT	$1057	$1335	$1436
PT	$ 349	$ 245	$ 349
FT/PT	3.03	5.45	4.11
B. Middle (Log Means)			
FT	$1494	$1500	$1455
PT	$ 290	$ 265	$ 258
FT/PT	5.15	5.65	5.64
C. Middle (Raw Means)			
FT	$1431	$1542	$1575
PT	$ 223	$ 208	$ 198
FT/PT	6.42	7.41	7.95
D. Large			
FT	$1871	$1941	$1809
PT	$ 179	$ 194	$ 151
FT/PT	10.45	9.94	11.98

program emphasis were correlated with full- or part-time enrollment levels, and thus both could be left at their respective mean values. Noncredit enrollment to some extent and salaries to a considerable extent were correlated with full- and part-time enrollment levels; thus values other than those at the mean were required to adequately represent typical combinations of institutional characteristics across the (credit) enrollment spectrum.

In terms of the underlying management finance issues--especially tuition levels and appropriations or allocations per FTE--both the absolute value of the marginal costs for full- and part-time students, and the ratio between the two costs, are important. As table 3 shows, there are differences in the results by size of institution and by type of estimating procedure. The only result shown that appears somewhat implausible is the OLS estimate for full-time students at small institutions. On theoretical grounds we would expect that the cost curve, partially depicted by the four "points" shown in table 3, would be U-shaped, and there is evidence to that effect (Brinkman 1981). For very small institutions, estimated marginal costs would escalate rapidly according to the two ridge procedures, but would continue to decline according to OLS (not tabled). In part, the reason for the OLS result is the negative coefficient on FTS which was mentioned earlier. It could be argued, then, that the ridge technique "corrects" the sign on that coefficient and thereby produces a better estimate of marginal costs, particularly for the smaller institutions in the sample. (For those readers perplexed by the ability of small institutions to have lower marginal costs than the mid-sized institutions, as shown in table 3, we note that the primary reason is lower faculty salaries at the small institutions. If the small institutions paid their faculty at the [raw] mean rate for the sample [$18,215], instead of $13,625, they would in fact have higher marginal costs than the mid-sized institutions. See Brinkman [1983] for more details.)

It is also useful to look at some statistics for the system as a whole, in order to see what else happens beyond changes in estimated coefficients as bias is introduced. As shown in table 4, one result is a decrease in R^2, the conventional measure of "goodness of fit." Of course, R^2 must decrease since by definition OLS will always give the best fit measured in that way. Similarly, the increase in sums of squares (SSQ) and in the standard error (SE) are to be expected. Notice that the changes are quite small, which means that we can perturb this system quite considerably without losing much predictive power. In any case, the tradeoff is that the mean square error (MSE) is drastically reduced as bias is introduced. As Judge et al (1980) emphasizes, the reduction in MSE cannot be guaranteed, but obviously there is no question that the reduction is both real and substantial for the particular model estimated in this study.

Another way of comparing the OLS and ridge estimators is to look at their respective results for varying sample sizes. As noted earlier, an abundance of data (i.e., observations) is usually a good antidote to multicollinearity. The problem, of course, is that numerous cases are not always available, or their acquisition may be expensive, and so on. Thus the analyst may often have to make do with a relatively small sample, and to look elsewhere, such as to an alternative estimator, for help in handling multicollinearity.

Table 5 shows the marginal-cost results, at mean (log) values of the variables, of estimating the translog cost function for a series of decreasing sample sizes, starting with the full sample (N=779). The smallest of the randomly drawn subsamples, N=50, retains 27 degrees of freedom. Overall, the least amount of difference among the various samples, as measured by the range of values, is found in the estimates generated by the ridge programs when k=.2. This is particularly true for the cost estimates for full-time students. The ridge program at k=.2 also does the best job for sample size 50a, as the alternative procedures, OLS and ridge at k=.003, lead to implausible results in the form

Table 4

Overall Statistics for the Cost Function Model Using Alternative Estimators

	R^2	SSQ	SE	MSE	MAX VIF
OLS	.921	.081	.285	333.0	680.4
Ridge (k=.003)	.919	.083	.288	129.9	55.8
Ridge (k=.200)	.903	.100	.316	7.3	.9

Table 5
Marginal Cost Estimates Using OLS and Ridge Estimators
on Different Sample Sizes

Number of Cases	OLS			RIDGE (.003)			RIDGE .2)		
	FT	PT	FT/PT	FT	PT	FT/PT	FT	PT	FT/PT
779	$1494	$290	5.15	$1500	$265	5.65	$1455	$258	5.64
389	$1387	$246	5.64	$1438	$209	6.88	$1435	$235	6.11
176(a)	$1509	$276	5.47	$1535	$223	6.88	$1346	$261	5.16
176(b)	$1218	$323	3.77	$1261	$297	4.25	$1338	$259	5.17
100	$1390	$359	3.87	$1571	$300	5.24	$1433	$247	5.80
50(a)	$1716	$-64	--	$1605	$-56	--	$1385	$ 91	15.22
50(b)	$1431	$266	5.38	$1317	$236	5.58	$1393	$183	7.61

of negative marginal costs for part-time students. However, while the estimate of part-time marginal costs is positive at k=.2, the estimated value of $91 is only about a third of the value estimated on the basis of the full sample; nonetheless, the estimate is certainly less misleading than those derived from OLS or ridge at k=.003. The same two bias parameter values, .003 and .2, were used for each sample simply for ease of exposition. Normally, the selection of a value for k would be sample specific, as will be discussed below.

Given the amount of collinearity in the system, we might expect considerable variability in the coefficients, and thus the marginal cost estimates, from one randomly drawn small sample to another. Using table 5 again, we see that the marginal cost estimates derived from the first sample of 50 institutions (a) differ considerably from a second sample of the same size (b). The estimates based on OLS and ridge at k=.003 vary more than those based on ridge at k=.200; at the same time the two former estimates gave results for part-time students and for FT:PT that are closest to those derived from using the full sample. The picture is again somewhat mixed for the two samples (a and b) at N=176, although the cross-sample stability of the ridge estimates at k=.200 is remarkable. Also, even though OLS and ridge at k=.003 give good results for some of the small samples, they do so using negative coefficients on FTS (not tabled), which suggests that those estimators might yield implausible results for very small institutions--as opposed to institutions with mean or larger enrollments.

High collinearity tends to make regression coefficients highly sensitive to the inclusion in the sample of particular cases, especially outliers. Thus another way in which OLS and ridge estimators can be compared is their respective reaction to the removal of cases from the sample. The more stability, that is, the less the change in the coefficients, the better.

Table 6 shows the results of removing two cases from the sample N=100. The two cases removed were outliers in the sense that their predicted total costs were furthest (about 2 standard derivations) from their actual total costs among the institutions in the subsample. Table 6 shows two kinds of comparisons. Panel A contains marginal cost results, while Panel B contains the estimated coefficients for a subset of the variables in the model (those that are directly involved in the calculation of the marginal effects of FT and PT). As can be seen from the percentage change calculations, the ridge procedure at k=.200 provides considerably more stability than the OLS procedure, with respect to both the estimated marginal costs and the underlying regression coefficients. The ridge procedure at k=.003 generally yields more stable coefficients than OLS, but not in all instances.

Variability among subsamples can be examined in yet another way. The estimated coefficients from one sample can be used with the values of the variables from a second sample to yield predicted total costs for the second sample. These predictions can then be correlated with actual total costs across the second sample, with the degree of correlation expressed as the familiar R^2. The question for present purposes is whether coefficients estimated by OLS will do better or worse than those estimated by ridge--with respect to the amount that R^2 will shrink when the original coefficients are used with a new sample. Results of such a comparison for two randomly drawn subsamples (N=100) are shown in table 7. All three estimators yield high R^2 values for the original sample, with the bias-related decrease in R^2 again being evident (as in table 4). The cross-verification procedure (Daniel and Wood, 1980) shows much less shrinkage in R^2 for the ridge estimators. Ridge at k=.2 is especially resistent to shrinkage in this instance.

A sample size of 100 was chosen for this test because the ratio between the number of cases (N) and the number of variables in the model (P) was not extreme in either

Table 6
Effects of Removing Two Cases from a Small Sample (N=100)

	OLS			Ridge (K=.003)			Ridge (K=.200)		
	N=100	N=98	% Change	N=100	N=98	% Change	N=100	N=98	% Change
A. Marginal Cost Estimates									
FT	$1390	$1296	6.8%	$1571	$1542	1.8%	$1434	$1489	3.8%
PT	$ 359	$ 319	11.1%	$ 300	$ 274	8.7%	$ 247	$ 252	2.0%
Ratio	3.87	4.06	4.9%	5.24	5.63	7.4%	5.81	5.91	1.7%
B. Unstandardized Regression Coefficients									
FT	-1.188	-1.469	23.7%	-.003	-.029	866.7%	.241	.238	1.2%
(FT)2	.097	.095	2.1	.045	.044	2.2	.018	.018	0.0
PT	.368	.340	7.6	.136	.135	0.7	.024	.022	8.3
(PT)2	.008	.003	62.5	.010	.008	20.0	.003	.003	0.0
(FTxPT)	-.019	-.003	84.2	-.006	-.004	33.3	.007	.007	0.0
(FTxDEG)	-.128	-.202	57.8	-.046	-.059	28.3	.027	.026	3.7
(FTxHCP)	.304	.417	37.2	.088	.103	17.0	.019	.020	5.3
(FTxNCS)	-.007	-.007	0.0	-.003	-.003	0.0	.001	.001	0.0
(PTxDEG)	.004	.021	425.0	-.020	-.019	5.0	.000	.000	0.0
(PTxHCP)	-.058	-.077	32.8	-.013	-.013	0.0	.002	.002	0.0
(PTxNCS)	.003	.000	87.3	.003	.003	0.0	.000	.000	0.0

Table 7

Comparison of R^2 Shrinkage* for Three Estimators

	R^2 Values		
	OLS	Ridge (K=.003)	Ridge (K=.2)
Original Sample (N=100)	.944	.938	.920
Second Sample (N=100)	.842	.900	.915
Change in R^2	.102	.038	.005

* When regression coefficients estimated on the basis of an original sample are used to predict total costs for institutions in a second sample.

direction. The superiority of the ridge procedure with respect to R^2 shrinkage has been shown to be directly related to the magnitude of the number of predictors to the number of cases (P/N) (Faden 1978). While not extreme, the ratio in this instance, 22/100, is actually fairly low compared to some that are reported in recent literature. For example, in Cowing and Holtmann (1983), the ratio is 107/138, while in Brown, Caves, and Christensen (1979) it is 21/67. It appears, then, that there is some likelihood of encountering situations where the ridge procedure could be helpful, assuming, of course, that maintenance of predictive power in a cross-validation sense has value.

The Bias Parameter Revisited

In the previous section, it was shown that at least in the particular situation being analyzed in this study, the ridge estimators offered some advantages over the conventional least squares approach. The ridge estimators provided theoretically better estimates for marginal costs at small institutions based on a large sample size, more plausible estimates when relatively small samples were used, less shrinkage in R^2 when coefficients were used across samples, and more stable estimates when cases were removed from small samples. But the ridge procedure does not provide a single alternative to OLS. Rather, the procedure can generate a virtually unlimited number of alternatives (i.e., sets of estimated coefficients), with each alternative being a function of the value assigned to the bias parameter k. Unfortunately, as was pointed out earlier, the selection of a value for k is anything but straightforward. Which is not to say that there are not straightforward procedures, but rather that there are alternative procedures which lead to different values of k and no one procedure is acceptable to experts in the field.

In illustrating the capabilities of the ridge procedure in the previous section, results (coefficients and marginal cost estimates) were shown for assigned k values of .003 and .2. These values were chosen because they ranged from relatively small to relatively large amounts of bias, and because they had intuitive appeal to the authors.

That is, they led to results (a change in signs or in the magnitude of coefficients) which made sense. But what of the more rigorous methods suggested to assign a value to k? What k values do these methods suggest for the model and data used in the present study?

We start by returning to the ridge trace procedure, which was developed early on by the originators of the ridge estimator (Hoerl and Kennard 1970). One can see in figure 1 above why an analyst might elect to pick a value for k somewhere between .003 and .009. Relatively little bias seems to accomplish a great deal in terms of stabilizing the coefficients. As an alternative, using figure 4 which displays the trace over a greater range of k, consider selecting a value for k of .200, which was used for some of the estimates discussed above, or a value of .360, which is the value one obtains on the basis of the harmonic mean technique (as derived in Appendix III). Figure 4 shows that little is gained in terms of stable behavior by using a value greater than .2. For that matter very little additional stability is gained by using a value greater than .05. Are there advantages in using the least amount of bias that gains the minimum acceptable stability? Perhaps so, at least on an intuitive level. That is, for the analyst who considers the introduction of bias as at best a necessary evil to combat multicollinearity, there may be some utility in staying as close to the OLS solution as possible, although this position is not justified in the literature. One seeming advantage of using k=.003 for the full sample (N=779), for example, is that all the "t-scores" save that on FTS are of the same sign and order of magnitude as the t-scores for OLS. Strictly speaking, B/SE cannot be treated as a t-score in ridge (k>0), that is, a given value of the ratio cannot be assigned a level of significance, because the sampling distribution for the statistic is unknown when k is determined from the data (Obenchain, 1977). Practically speaking, however, the analyst might still be willing to use the statistic in interpreting how well the respective variables were performing if the amount of bias were very small. Furthermore, the smaller the bias the less the increase in the residual sums of squares for the model as a whole (see table 4).

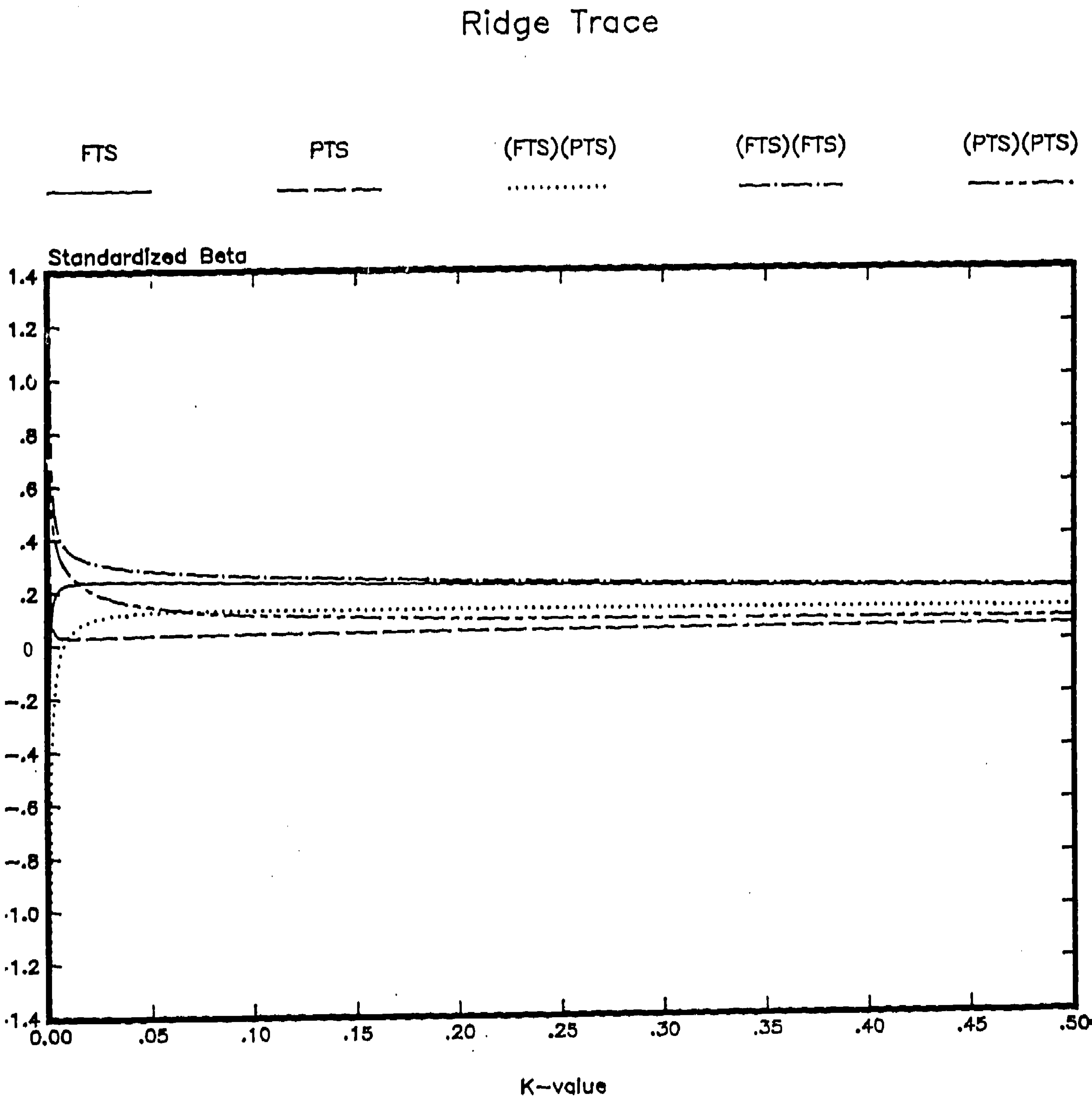

Fig.4. Ridge trace for key output variables, full sample, showing extended range for K−values

Figures 5 and 6 show the ridge trace for subsample 50(a), which was a "difficult" sample for all three estimators used in the analysis (see table 5). Note that in this instance coefficient stability is achieved with the introduction of somewhat more bias, roughly .015 or so. Note that the value of the coefficient on FT continues to increase all the way out to k=.50. Yet, the harmonic mean formulation suggests that k be set at .036, an order of magnitude less than indicated for the full sample (N=719). For the other very small subsample, 50(b), the harmonic mean formulation suggests that k be set at .046. The appropriate choice for k, then, as noted earlier, is entirely sample specific, and a function of a particular method of selection as well. Not the stuff, in other words, likely to impress the purist. On the other hand, it does seem in looking at the ridge traces that any amount of bias, within some range of k>0, would be a better choice than staying with OLS, assuming that the stability of particular coefficients was of greater concern than maximizing goodness of fit with respect to the predicted value of the dependent variable.

Discussion

The purpose of this study was to assess whether, in the face of extreme multicollinearity in estimating cost functions, the ridge procedure might be a useful alternative to the conventional least squares estimator. Utility will depend, of course, on perspective and need. As the problem was structured in the present study, the ridge procedure appeared to offer several modest advantages. The task was to estimate marginal costs for a multiproduct enterprise. Thus ridge improvements in the precision and stability of estimated coefficients were important--marginal cost estimates being a function of a set of coefficients. Similarly, related matters which are often of concern in estimating cost functions but not pursued in the present study, such as economies of scale and economies of scope, also depend on the value of estimated coefficients.

Ridge Trace

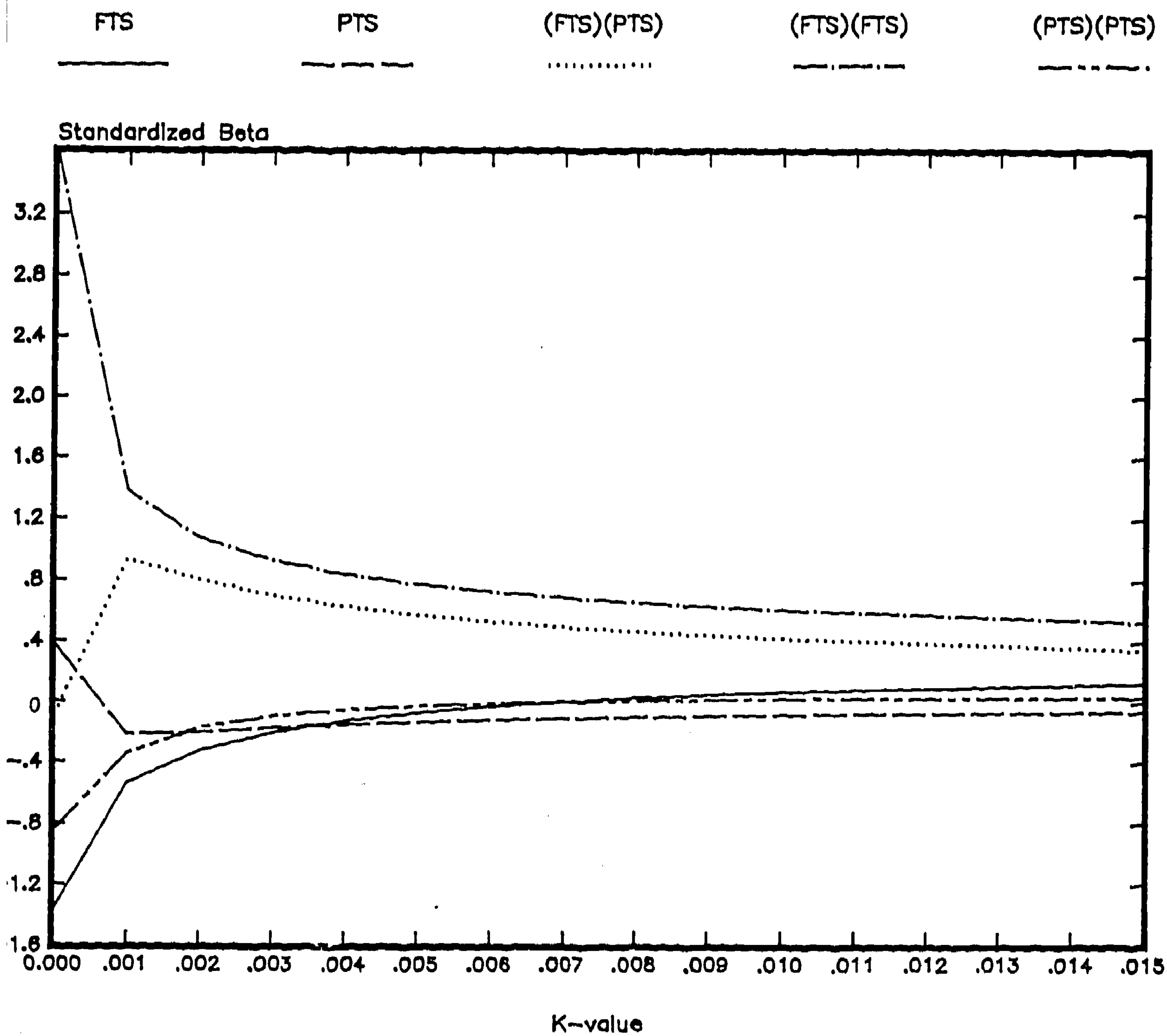

Fig.5. Ridge trace for key output variables, subsample (50a)

Ridge Trace

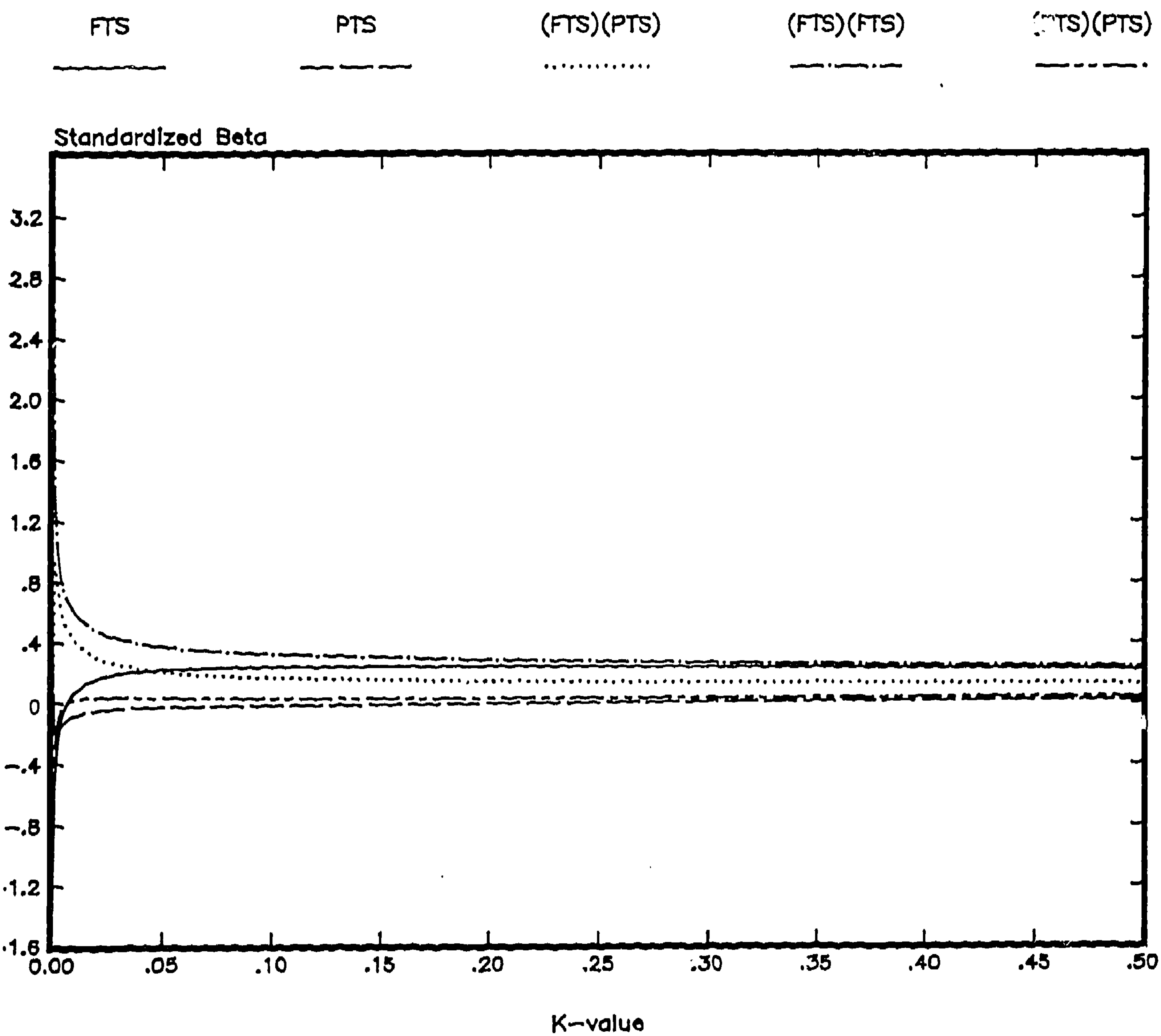

Fig.6. Ridge trace for key output variables, subsample(50a), showing extended range for K-values

If, on the other hand, the objective was to predict total costs with the least amount of error, then OLS has one immediate advantage. For any given sample, OLS will always provide the lowest possible residual sum of squares. Recall, though, that even when predicting total costs, the ridge procedure has a potential advantage when collinearity is high. If the coefficients estimated for one sample are to be used to predict total costs for the observations in another sample, the R^2 shrinkage incurred by a ridge estimator is likely to be less than that for the OLS estimator; the resultant, shrunken R^2 values for ridge estimators, then, may be higher than that for the corresponding OLS estimator.

There is another matter of perspective to consider, other than the specific aims of a cost estimation procedure. Roughly speaking, one might describe it as the difference between a theoretical versus a practical perspective. On the basis of reviewing the theoretically oriented literature, it appears as though there are serious, unresolved problems with the ridge procedure (the best summary of these problems is in Judge et al 1980). One might describe it simply as a situation in which the advantages offered by ridge are possible, but cannot be guaranteed theoretically. Furthermore, the failure to date to develop a theoretically unimpeachable way of assigning a value to the bias parameter has weakened the case for ridge.

Looked at practically, however, the ridge procedure does seem to offer hope in the battle against multicollinearity. In every comparison conducted for the present study, ridge was in some pertinent sense superior to OLS. All in all, it appears that the marginal cost estimates generated by ridge were less risky than those generated by OLS. From a practical perspective that may be enough to justify using the ridge procedure.

Finally, it should be apparent from this study that at the very least ridge regression provides a means for data and model exploration. By comparing OLS and ridge estimates, and especially by examining ridge traces and VIFs, the analyst can come to a better understanding of the effects of multicollinearity in a given situation. This is true whether one opts for a reduced form model, that is, elects to eliminate some of the collinear independent variables, or chooses to stay with a theory-driven model regardless of the attendant estimation problems, as was done in this study.

Appendix I: Properties of Ridge Estimators

The ridge estimator is a linear transformation of the least squares estimator, which is just

$$\tilde{\beta} = (X'X)^{-1} X'Y \tag{12}$$

Rearranging terms we obtain

$$(X'X)\tilde{\beta} = X'Y \tag{13}$$

Substituting in equation (6), we obtain (14)

$$\tilde{\beta}(k) = (X'X + kI)^{-1}(X'X)\tilde{\beta} \tag{15}$$

For $k > 0$, $\tilde{B}(k)$ is the ridge estimator.

The relationship of the ridge estimator of the OLS estimator is then given by

$$\begin{aligned} \tilde{\beta}(k) &= (X'X + kI)^{-1}(X'X)\tilde{\beta} \\ &= [I + k(X'X)^{-1}]^{-1}\tilde{\beta} \\ &= Z\tilde{\beta} \end{aligned} \tag{16}$$

so that $\tilde{B}(k)$ may be viewed as a linear transform of $\tilde{B}$.

If the squared length of the regression vector B is fixed at B^2, then $\tilde{B}(k)$ is the value of B that gives a minimum sums of squares of residuals. This is illustrated in Figure 7 for a two parameter problem by Marquadt and Snee (1975, p. 5) as follows:

> The point $\tilde{B}$ at the center of the ellipses is the least squares solution. $\tilde{B}$ the sum of squares of residuals, Φ, achieves its minimum value. The small ellipse is the locus of points in the B^1, B^2 plane where the sum of squares Φ is constant at a value larger than the minimum value. The circle about the origin is tangent to the small ellipse at $\tilde{B}(k)$. Note that the ridge estimate $\tilde{B}(k)$ is the shortest vector that will give a residual sum of squares as small as the Φ value anywhere on the small ellipse. Thus the ridge estimate gives the smallest regression coefficients consistent with a given degree of increase in the residual sum of squares.

Other key properties of B(k) include:

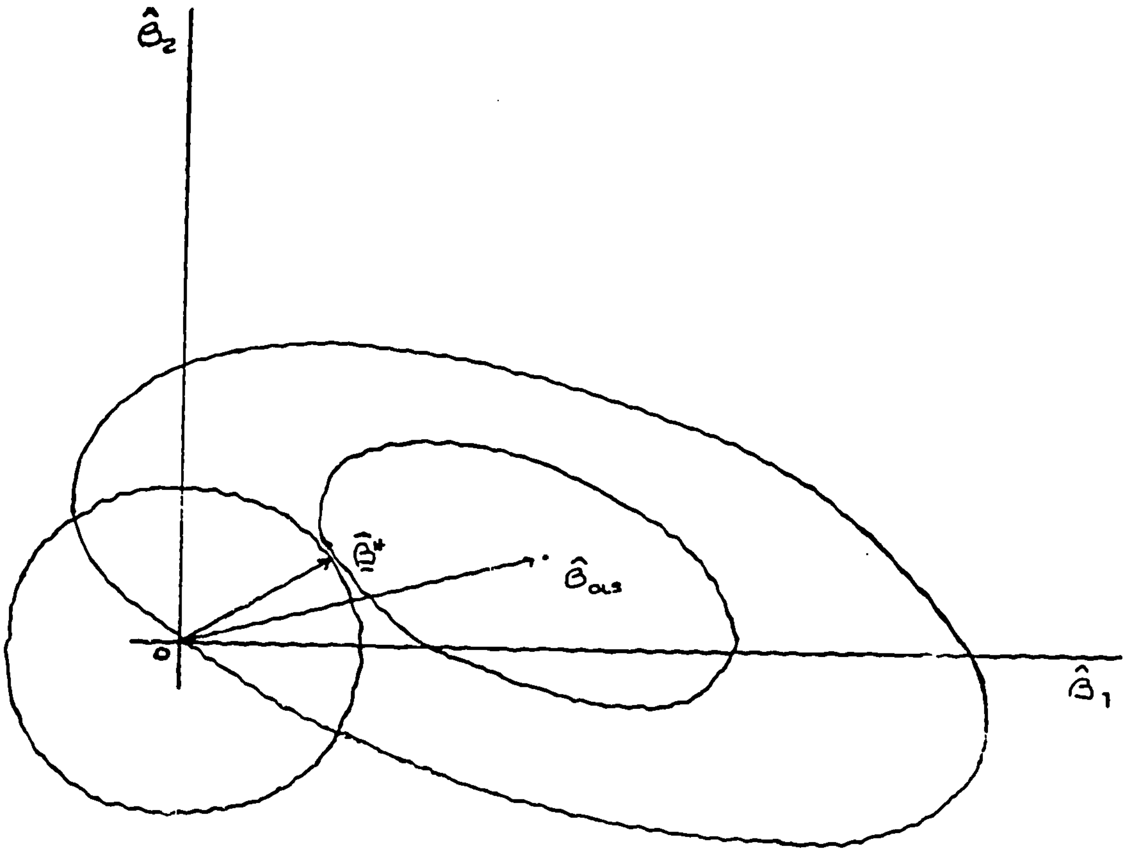

Fig. 7. The geometry of ridge regression

The length of $\tilde{B}(k)$ is a decreasing function of k.

The variance term is a decreasing function of k. That is,

$$V[\tilde{\beta}(k)] = \sigma^2 (X'X + kI)^{-1}(X'X)(X'X + kI)^{-1} = \sigma^2 Z (X'X)^{-1} Z \quad (17)$$

The bias term is an increasing function of k. That is,

$$ESD = tr\{V[\tilde{\beta}(k)]\} + \beta'(Z-I)'(Z-I)\beta = Variance + (Bias)^2 \quad (18)$$

where ESD denotes the expected squared distance to B.

This last property points out that the mean square error of $\tilde{B}(k)$ is composed of two components: (1) the sum of variances of all the estimated coefficients; and (2) the square of the bias introduced by substituting $\tilde{B}(k)$ for B.

Appendix II

The algorithms reported in Table 1 require that (X'X) has been transformed to the space of orthogonal predictor variables. In this form, the model expressed in equation (1) becomes

$$y = X^{*}\alpha + e \qquad (19)$$

where $X=X^{*}P$, $\alpha=PB$, $P'P=PP'=I$, $P'(X'X)P=\Lambda$, and Λ denotes the diagonal matrix of eigenvalues of (X'X).

Table 1. Closed Form Methods for Selecting k

1. Harmonic Mean (Hoerl, Kennard, and Baldwin, 1975)

$$k = p\sigma^2 / \tilde{\alpha}'\tilde{\alpha}$$

2. Empirical Bayes (Lawless and Wang, 1976).

$$k = p\sigma^2 / \sum_{i=1}^{p} \lambda_i \tilde{\alpha}_i^2$$

3. Iterative Estimation (Hoerl and Kennard, 1970).

$$k_i = p\tilde{\sigma}^2 / (\tilde{\alpha}_{k_{i-1}})' (\tilde{\alpha}_{k_{i-1}})$$

4. Variance Normalization (Bulcock, Lee, and Luck, 1977).

$$1/p \sum_{i=1}^{p} \lambda_i / (\lambda_i + k) = 1$$

5. Minimization of the Frequentist Expectation of the MSE (Dempster, Schatzoff, and Wermuth, 1977).

$$\sum_{i=1}^{p} (k_{q_i} - \tilde{\sigma}^2) / (\lambda_i + k) = 0$$

6. RIDGM Bayesian Approach (Dempster, Schatzoff, and Wermuth, 1977).

$$\sum_{i=1}^{p} \tilde{\alpha}_i^2 / (\tilde{\sigma}^2/k + \tilde{\sigma}^2/\lambda_i) = p$$

7. Generalized Ridge (Hocking, Speed, and Lynn, 1976).

$$k = \tilde{\sigma}^2 \sum_{i=1}^{p} \lambda_i^2 \tilde{\alpha}_i^2 / \sum_{i=1}^{p} \tilde{\alpha}_i^4$$

Appendix III: Ridge Trace

In linear estimation one postulates a model of the form

$$Y = X\tilde{\beta} + e \tag{20}$$

It follows from equation (20) that the residual sums of squares can be written as

$$\begin{aligned} \Phi &= (Y - X\beta)'(Y - X\beta) \\ &= (Y - X\tilde{\beta})'(Y - X\tilde{\beta}) + (\beta - \tilde{\beta})' X'X (\beta - \tilde{\beta}) \\ &= \Phi_{min} + \Phi(\beta) \end{aligned} \tag{21}$$

The Ridge Trace can be shown to be following a path through the sums of squares surface so that for a fixed Φ a single B is chosen which is of minimum length. This can be stated precisely as follows: Minimize B'B subject to,

$$(\beta - \tilde{\beta})'(X'X)(\beta - \tilde{\beta}) = 0 \tag{22}$$

This is graphically illustrated in Appendix I, Figure 7.

As a Lagrangian problem this is

$$F = \beta'\beta + (1/k)\left[(\beta - \tilde{\beta})'(X'X)(\beta - \tilde{\beta}) - \Phi \right] \tag{23}$$

where (1/k) is the multiplier. Then,

$$\frac{\partial F}{\partial \beta} = 2\beta(1/k)[2(X'X)\beta - 2(X'X)\tilde{\beta}] = 0 \tag{24}$$

Equation (23) reduces to

$$\beta = \tilde{\beta}(k) = [X'X + kI]^{-1}X'Y \tag{25}$$

The value of k is then chosen to satisfy the restraint imposed by equation (22). This is the ridge estimator. In practice it is easier to choose k>=0 and then compute Φ.

Appendix IV: Harmonic Mean Approach

The approach derives from two assumptions. First, if $X'X=I$, then a minimum mean square error term is obtained if (Hoerl and Kennard, 1970)

$$k = p\sigma^2/\beta'\beta \tag{26}$$

Secondly, the general form of equation (18) is rewritten as

$$[X'X + P'kP]^{-1}\tilde{\beta}(k) = X'Y \tag{27}$$

where $PkP=KI_p$. A minimum mean square error must be obtained when (Hoerl and Kennard, 1970, p. 63)

$$k_i = \sigma^2/\alpha_i^2 \tag{28}$$

Hoerl, Kennard and Baldwin (1975) argue that if the K_i are to be combined to obtain a single value of k, one would not want to use the arithmetic mean since very small α_i with no predictive power would yield very large values for k. They suggest that a more reasonable approach of averaging the k_i is to employ the harmonic mean. That is, calculate

$$\begin{aligned} 1/k &= (1/p)\sum_{i=1}^{p}(1/k_i) = (1/p)\sum_{i=1}^{p}(\alpha_i^2/\sigma^2) \\ &= (1/p\sigma^2)\sum_{i=1}^{p}\alpha_i^2 = \alpha'\alpha/p\sigma^2 = \beta'\beta/p\sigma^2 \end{aligned} \tag{29}$$

The value of k is then given by

$$k = p\sigma^2 / B'B \tag{30}$$

The results represented by equations (29) and (30) indicate that a reasonable choice for an automatic selection of k_i is an estimate of $(p\sigma^2/B'B)$. And that is what is used vis-a-vis

$$k = p\tilde{\sigma}^2 / \tilde{B}'\tilde{B}. \tag{31}$$

References

Adams, C. R., R. L. Hankins, and R. G. Schroeder. "The Literature of Cost and Cost Analysis in Higher Education." Volume 1 in A Study of Cost Analysis in Higher Education. Washington, D.C.: American Council on Education, 1978.

Allen, R. and P. Brinkman. Marginal Costing Techniques. Boulder, Colo.: National Center for Higher Education Management Systems [NCHEMS], 1983.

Baranchik, A. J. "A Family of Minimax Estimators of the Mean of a Multivariate Normal Distribution." Annals of Mathematical Statistics, 41 (1970), 642-645.

Bartlett, M. S. "Tests of Significance in Factor Analysis." British Journal of Psychology, Statistical Section, 3 (1950), 77-85.

Blalock, H. M. Causal Inferences in Experimental Research. Chapel Hill: University of North Carolina Press, 1964.

_____. "Correlated Independent Variables: the Problem of Multicollinearity." Social Forces, 3 (1963), 233-237.

Bowen, H. R. The Costs of Higher Education. San Francisco: Jossey-Bass, 1980.

Brinkman, Paul. "The Financial Impact of Part-Time Enrollments on Two-Year Colleges: A Marginal-Cost Perspective." Paper presented at the annual conference of the Association for Institutional Research, Toronto, May, 1983.

_____. Marginal Costs of Instruction in Public Higher Education. Ph.D. dissertation, University of Arizona, 1981. Ann Arbor, University Microfilms, 1982.

Brown, R. S., D. W. Caves, and L. R. Christensen. "Modelling the Structure of Cost and Production for Multiproduct Firms." Southern Economic Journal, 46 (July 1979), 256-270.

Bulcock, J. W., W. F. Lee, and W. S. Luck. "A Model of School Learning Using the Variance Normalization Method of Ridge Regression Analysis." I.E.A. Methodology Symposium, Michigan State University, 1977.

Chatterjee, S. and B. Price. Regression Analysis by Example. New York: John Wiley and Sons, 1977.

Christensen, D., W. Jorgenson, and L. J. Lau. "Transcendental Logarithmic Production Frontiers." Review of Economics and Statistics, 55 (February, 1973), 28-45.

_____. "Conjugate Duality and the Transcendental Logarithmic Production Function" (abstract). Econometrica, 3 (July 1971), 255-256.

Cohn, E. The Economics of Education. Cambridge: Ballinger Publishing Company, 1979.

Cowing, T. G. and A. G. Holtmann. "Multiproduct Short-Run Hospital Cost Functions: Empirical Evidence and Policy Implications from Cross-Section Data. Southern Economic Journal, 49 (January 1983), 637-653.

Daniel, Cuthbert and Fred S. Wood. Fitting Equations to Data. New York: John Wiley and Sons, 1980.

Darlington, R. B. "Reduced Variance Regression." Psychological Bulletin, 86 (1979), 1238-1254.

_____. "Reduced-Variance Regression," Psychological Bulletin, 85 (1978), 1238-1255.

Griffin, J. M. "The Approximation Characteristics of Generalized Functional Forms: Results from Pseudo-Data Experiments." Advances in Applied Micro-economics, 2 (1982), 3-18.

Gunst, R. F. and R. L. Mason. "Advantages of Examining Multicollinearity in Regression Analysis." Biometrics, 33 (1977), 249-260.

Haitovsky, Y. "Multicollinearity in Regression Analysis: Comment," The Review of Economics and Statistics, 59 (November 1969), 486-489.

Hemmerle, W. J. "An Explicit Solution for Generalized Ridge Regression." Technometrics, 12 (1975), 55-68.

Hocking, R. "The Analysis and Selection of Variables in Linear Regression." Biometrics, 32 (1976), 1-49.

Hocking, R. R., F. M. Speed, and M. J. Lynn. "A Class of Biased Estimators in Linear Regression." Technometrics, 18 (1976), 425-437.

Hoerl, A. E. and R. W. Kennard. "Ridge Regression: Applications to Nonorthogonal Problems." Technometrics, 12 (February 1970), 55-67.

Hoerl, A. E., R. W. Kennard, and K. F. Baldwin. "Ridge Regression: Some Simulations." Communications in Statistics, 4 (1975), 105-123.

Johnston, J. Econometric Theory, 2nd Edition. New York: McGraw-Hill, 1972.

______. Statistical Cost Analysis. New York: McGraw-Hill, 1960.

Judge, G. G., W. E. Griffiths, R. C. Hill and T-C. Lee. The Theory and Practice of Econometrics. New York: John Wiley & Sons, 1980.

Kennedy, Peter. A Guide to Econometrics. Cambridge, Mass.: MIT Press, 1979.

Kumar, T. Krishna. "Multicollinearity in Regression Analysis," The Review of Economics and Statistics, 57 (August 1975), 365-366.

Lawless, J. R. and P. Wang. "A Simulation Study of Ridge and Other Regression Estimators." Communications in Statistics, 5 (1976), 307-323.

Leamer, E. E. "Multicollinearity, a Bayesian Interpretation." Review of Economic Statistics, 55 (1973), 371-381.

Lindley, D. V. and A. Smith. "Bayes Estimates for the Linear Model." Journal of the Royal Statistical Society, Ser. B, 34 (1972), 1-41.

Mallows, C. L. "Some Comments on Cp." Technometrics, 15 (1973), 661-675.

Marquadt, D. "Generalized Inverses, Ridge Regression, Biased Estimation." Technometrics, 12 (1970), 591-612.

Marquadt, D. W. and R. D. Snee. "Ridge Regression in Practice." The American Statistician, 29 (1975), 3-20.

Mayer, L. S. and T. A. Wilkie. "On Biased Estimation in Linear Models." Technometrics, 15 (1973), 497-500.

McDonald, G. C. and D. I. Galarneau. "A Monte Carlo Evaluation of Some Ridge-Type Estimators." Journal of the American Statistical Association, 70 (1975), 407-416.

McFadden, D. "Cost, Revenue, and Profit Functions." Production Economics: a Dual Approach to Theory Application, Vol. 1, edited by M. Fuss and D. McFadden. Amsterdam: North-Holland Publishing Co., 1978.

Obenchain, R. L. "Classical F-Tests and Confidence Regions for Ridge Regression." Technometrics, 19 (1977), 429-439.

Pauly, M. V. "Medical Staff Characteristics and Hospital Costs." Journal of Human Resources, 13 (Supplement 1978), 77-111.

Sclove, S. "Improved Estimates for Linear Regression." Journal of the American Statistical Association, 63 (1968), 596-606.

Shepard, R. W. Cost and Production Functions. Princeton, New Jersey: Princeton University Press, 1953.

Silvey, S. D. "Multicollinearity and Imprecise Estimation," Journal of the Royal Statistical Society, Series B, 31 (1969), 539-552.

Smith, V. Kerry (ed.). Advances in Applied Micro-Economics, vol. 2. Greenwich, Conn.: JAI Press, 1982.

Spady, R. H. Econometric Estimation of Cost Functions for the Regulated Transportation Industries. New York: Garland Publishing, 1979.

Theil, H. "On the Use of Incomplete Prior Information in Regression Analysis." Journal of the American Statistical Association, 58 (1963), 401-414.

Vinod, H. D. "A Survey of Ridge Regression and Related Techniques for Improvements Over Ordinary Least Squares." Review of Economics and Statistics, 50 (1978), 121-131.

Walters, A. A. "Production and Cost Functions: An Econometric Survey." Econometrica, 3 (1963), 1-66.

Wichers, C. Robert. "The Detection of Multicollinearity: A Comment," The Review of Economics and Statistics, 57 (August 1975), 366-368.

Willan, A. R. and D. G. Watts. "Meaningful Multicollinearity Measures." Technometrics, 20 (1973), 407-412.

Yntema, T. D. "Steel Prices, Volume and Costs." United States Steel Corporation. Economic Committee Papers, 1 (1940), 223-323.