ALTERNATIVE APPROACHES FOR EXAMINING THE TEMPORAL STABILITY OF PARAMETER ESTIMATES IN A MARKETING MODEL

James S. Moore, Indiana University-Purdue University at Fort Wayne
Zoher E. Shipchandler, Indiana University-Purdue University at Fort Wayne

ABSTRACT

The temporal stability of estimated parameters in multiple regression marketing models is important if the model is to prove useful in making economic inferences and in developing marketing strategies. OLS estimates are potentially distorted in the presence of collinear data sets that typify marketing models; consequently, any underlying temporal stability present may go undetected. This paper investigates the temporal stability of parameter estimates by comparing the results obtained from OLS, ridge, and latent root regression techniques in the presence of ill-conditioned data. Ridge regression provided improved individual coefficient stability and slightly greater predictive accuracy beyond the original estimation period.

Introduction

While multiple regression analysis (MRA) has been extensively discussed and utilized in marketing research, much of this history has focused on its aggregate predictive capabilities or one-time parameter estimates. Little has been done, as yet, to investigate the year to year parameter stability of MRA estimates for a common model specification. Such temporal stability of coefficient estimates is important if the model is to be accepted as theoretically sound and practically useful in developing a marketing plan for the future. A marketing manager would like to make operational plans, confident in the knowledge that the marginal impact of each major component of the model will remain reasonably constant even though the magnitude of the component itself will very likely change.

The intent of this study is to investigate the relative temporal stability of individual parameter estimates generated by three alternative MRA techniques, namely ordinary least squares (OLS), ridge regression (RR), and latent root regression (LRR). Each of these procedures is employed to estimate the co-efficients of the same linear model specification for each year from 1976 through 1982. These coefficients represent the separate marginal effects of each variable. The data set employed is one that is vulnerable to multicollinearity, as are data in many marketing situations. Consequently, our concern is with uncovering any year to year "staying power" of such marginal responses in the presence of ill-conditioned data.

Review of Literature

The OLS estimator, $\hat{\beta}$, is known to be the best linear unbiased estimator (B.L.U.E.) when the predictor variables are orthogonal. But when there is substantial intercorrelation among the predictor variables, $\hat{\beta}$ tends to deviate from the true value of the parameters being estimated in unpredictable ways. The problems associated with the presence of collinearity in statistical models have been well developed in the literature (Farrar and Glauber 1967). In sum, the OLS estimated coefficients, when generated from ill-conditioned data, may be compromised in terms of sign, magnitude and/or significance, and hence may not reflect only the influence of their associated predictor variables. This would invalidate the marginal response inferences an investigator might wish to make, and subsequently frustrate the search for any underlying temporal stability of these estimates.

Numerous approaches have been put forth in the econometric and marketing literature to address the problem of collinearity. The commonly employed approach of deleting collinear variables from the analysis, so as to reduce the degree of interdependence among the remaining variables, may often prove very inadequate for several reasons. First, the investigator may be responding to a symptom, such as inappropriate sign and/or low significance, and subsequently delete a variable that properly belongs in the equation specification. Second, the deletion of selected variables, whether they are identified via an arbitrary heuristic or by a stepwise regression procedure, can ultimately bias the remaining parameter estimates (Belsley et al, 1980).

Procedures for improving the conditioning of the data also have severe limitations, such as the unavailability of additional "clean" data. Another school calls for a form of artificial orthogonalization of the existing data. This often involves a transformation of variables with the resulting "new variables" no longer plagued by the collinearity. One such mechanism is factor analysis which aggregates several individual predictor variables into a smaller number of groups or factors. This results in the loss of some information and makes the interpretation of the new variables difficult.

Still another approach is the employment of a biased estimation procedure. Such techniques seek closer overall proximity to the true parameter by securing a much lower variance than a B.L.U.E. technique, in exchange for accepting their accompanying bias. If the amount of bias is kept small, then the estimate will be dominated by the reduced variance resulting in a lower overall mean square error. Two biased techniques that have gained exposure recently in the marketing literature are ridge regression and latent root regression.

Mahajan et al (1977) contrast ridge coefficients with those of OLS and show that while the ridge estimates are biased, they do possess smaller variance than the least squares estimates. Erickson (1981) found ridge regression preferable to OLS in handling the ill-conditioning of a highly autocorrelated data set as he sought to measure the cumulative impact of marketing efforts on sales beyond the period of the promotion's implementation.

Latent root regression was introduced by Hawkins (1973) and Webster, Gunst and Mason (1974) as a modified least squares estimation procedure. LRR enables the user to detect the presence of near singularities and determine whether they possess any predictive value for the criterion variable. LRR estimates are obtained by deleting any latent vectors which are associated with the collinearity, and which do not appreciably influence the explained variation in the dependent variable. Thus, the LRR estimator is purged of the effect of any "non-predictive" near singularities while retaining the influence of any near singularities that do contain substantive information about the underlying model. Should all near singularities identified be found to be "predictive", then no latent vectors are deleted and the LRR estimator coincides with the OLS estimator.

While numerous econometric procedures for handling ill-conditioned data have been discussed in the literature, the temporal stability of the coefficients estimated by such techniques has remained largely unexamined. The intent of this study is to investigate and compare the year-to-year parameter stability of a common model specification across estimation procedures.

The Model

The vehicle for this investigation of temporal stability is the traditional additive multiple regression model $Y = X\beta + \varepsilon$ where ε is a n x 1 vector of random errors with mean zero and variance σ^2.

The data base consists of a systematic sample of 114 SMSA's selected from Sales and Marketing Management's Survey of Buying Power for each of the years 1976 to 1982. The model was estimated annually with cross-sectional data. The years 1976 through 1982 cover both a period of relative economic prosperity from 1976 to 1979 and a period of stagflation from 1980 to 1982. This affords the opportunity to compare estimation methods under conditions in which some of the predictor variables have undergone sizable changes.

The basic conceptual specification of the model can be stated as $S = f(C,M)$ where S = real annual dollar sales of furniture and household furnishings in an SMSA, C = yearly capacity to buy furniture and household furnishings in an SMSA, M = yearly market size of furniture and household furnishings in an SMSA. The actual predictor variables for each SMSA and time period include median household effective buying income (EBI) in constant dollars, the total number of households (HH), the percentage of the market area's total population in the age groups 18 through 24 (YOUNG), 25 through 49 (TARGET), and 50 and over (OLD), and the percentage of the market area's unemployed labor force (UNEMP).

Each regressor variable is expected to play a unique contributory role in influencing the aggregate volume of furniture sales in its respective SMSA. It is posited that, ceteris paribus, the dependent variable will be directly related to EBI, HH, and TARGET and inversely related to YOUNG, OLD and UNEMP. It is assumed that the prime age group for furniture is the TARGET group, whereas many persons in the YOUNG and OLD age groups are not active in the market for new home furnishings. To mitigate the influence of inflation over the seven year period, the dollar denominated variables were deflated to real terms.

Methodology

Given that marketing data are frequently plagued by multicollinearity and that our model was deliberately specified so as to be vulnerable to such ill-conditioning, we felt it appropriate to document the extent of the collinearity present. The traditional correlation matrix for our data set finds an average of three occurrences of .50 or more on each side of the primary diagonal, per year. However, such figures may seriously understate the true ill-conditioning present within the data as they reflect only the simple pairwise correlations and consequently fail to reveal any evidence of the collinearity if the dependencies are group interrelations instead. Moreover, it would be possible for three or more predictor variables to be collinear while no two such data series exhibit high correlation. In an effort to capture such "package" ill-conditioning, we employ the "multi-collinearity index" (mci) of Thisted and Morris (1980). Their index has a closed range $1 < \text{mci} < p$ where p is the number of regressors. Orthogonal data will yield an index near the upper end of the spectrum while values near unity are associated with a high degree of package collinearity. For our seven sets of data, mci ranged from 1.085 to 1.141 within a possible spectrum of $1 < \text{mci} < 6$.

Further, Belsley et al (1980) construct variance-decomposition proportions so as to identify the extent of the degrading impact of the multicollinearity. When the same singular value is associated with a large (>50%) proportion of the variance of two or more coefficient estimates, they cite that occurrence as "evidence that the corresponding near dependency is causing problems" with the "quality of the subsequent regression analysis..." In this study, the matrix of variance-decomposition proportions for each year was found to contain one or more instances of such degrading singular values.

Parameter estimates were then derived annually using OLS, RR, and LRR for the same model specified earlier. In an effort to gauge the temporal stability of each individual parameter in the model, the coefficient of variation

(σ / μ) over the seven annual estimates was employed as a unit free index of relative stability. The OLS results serve as a benchmark for comparison since it is the most commonly employed estimation technique. Yet OLS is known to be suspect with illconditioned data. Specifically, the OLS estimator, $\hat{\beta}$, diverges farther from the true population parameter, β, as the vector of regressors becomes less orthogonal. The primary manifestation of such ill-conditioning with OLS is on the individual parameter estimates and their respective variances.

The ridge estimator:

$$\beta(k) = (X'X + k\,I)^{-1} \; X'\,Y \text{ for } 0 \leq k \leq 1 \qquad (1)$$

employs the dimensionless parameter k and is a biased estimator. The ridge estimator is similar to the OLS estimator, $\hat{\beta}$, except that the main diagonal of the correlation matrix is augmented prior to inverting by a small positive quantity, k, where k is an index of bias. RR provides estimates that have lower variance and possible lower mean square error than $\hat{\beta}$. In fact, Hoerl and Kennard (1970) prove that there always is a $k > 0$ such that the variance plus the squared bias of $\beta(k)$ is less than the variance plus the squared bias of the OLS estimator. When k is sufficiently small, the variance decreases faster than the increase in the square of the bias. Thus as k increases from zero, the mean square error initially declines and then later increases. By accepting no more bias than necessary to stabilize the coefficients, the effect of the increased bias is more than offset by the reduced variance, generating a net reduction in the mean square error. Estimates of parameters that have a lower variance deserve consideration in model building where the major concern is to investigate the separate effect of each of the potentially collinear predictor variables in the specification.

The ridge estimator, founded on the hypothesis that the regression coefficients, other than the constant term, are zero, stochastically shrinks the estimates toward that target. Hence, the bias introduced is not an arbitrary, uncontrolled bias, but rather, a bias toward the hypothesis that the regression coefficients are zero. This directional influence is fundamentally consistent with the philosophy of predictor variable retention within a specified model. The burden of proof remains with each regressor to demonstrate that it does make a significant contribution by distinguishing its estimated coefficient sufficiently from zero.

The use of ridge regression necessitates the determination of an appropriate ridge constant, k. In reality the optimal k value is a function of the true parameter, β, and consequently cannot be established with certainty. Numerous mechanical techniques exist for establishing an acceptable k value and much controversy exists in the literature concerning k selection (Hoerl and Kennard 1970; Vinod 1978). This study employs the ridge trace estimate of the ridge constant as initially advocated by Hoerl and Kennard. The trace was applied independently to each of the seven annual equations with the resulting k values ranging from 0.04 to 0.18.

Latent root regression can be viewed as a modified least squares technique when both estimators are represented as a weighted linear combination of the associated latent vectors of the matrix of augmented correlation coefficients. Following the notation adopted by Webster et al (1974) and Sharma and James (1981)

each estimate $$\beta = -\left[\sum_{j=o}^{p} w_j \cdot \gamma_{oj} \cdot \lambda_j^{-1} \cdot \gamma_j^{o} \Big/ \sum_{j=o}^{p} w_j \cdot \gamma_{oj}^{2} \cdot \lambda_j^{-1}\right] \qquad (2)$$

where:

p = the number of predictor variables
λ_j = the latent roots of the augmented correlation matrix for $j = o,1,\ldots,p$ and ordered such that $\lambda_o \leq \lambda_1 \leq \ldots. \leq \lambda_p$
γ_j = the corresponding latent vectors
γ_{oj} = the first element of the jth vector
γ_j^{o} = the jth vector without the first element, and
w_j = 0,1 dummy deletion variables.

The OLS estimator has all $w_j = 1$ in equation (2) and hence contains all latent vectors regardless of their degree of collinearity or predictive ability. In contrast, the LRR estimator will set $w_j = 0$ in equation (2) for any identified vectors associated with "non-predictive singularities." Thus it retains only predictive singular and non-singular vectors in combination.

The latent root estimator has also been compared to the ridge estimator since both are biased estimation procedures (Hawkins 1975). Sharma and James (1981) note that the latent root estimator can be thought of as the ridge estimator with non-uniform k values i.e., k = 0 for predictive latent vectors (this would completely retain their influence) and $k = \infty$ for non-predictive latent vectors (this would completely eliminate their influence). However, since RR employs a uniform k value, it like OLS, does not distinguish between predictive and non-predictive singularities and thus becomes a weighted linear combination of all vectors.

In LRR the identification of any near singularities present and their associated predictive natures involves the magnitudes of the latent roots, λ_j, and associated latent vectors, γ_j, of the augmented correlation matrix. The (p + 1) latent vectors define a set of mutually orthogonal axes Z_o, Z_1, ..., Z_p that are an alternative to the (p + 1) dimensional Euclidean space defined by the dependent variable, Y, and the p predictor variables, X_1, ..., X_p. The jth latent root, λ_j, measures the dispersion of the n data observations in the direction defined by the jth latent vector i.e., the Z_j axis. A "small" λ_j indicates there is little variability of data points in the Z_j direction and hence a high interdependence among the associated predictor variables. Should the "small" λ_j be affiliated with a latent vector that indicates Z_j is nearly orthogonal to the axis of the criterion variable, i.e., "small" γ_{0j}, then the near singularity can be labeled nonpredictive as the collinearity present is simply among the predictor variables with little or no spillover impact on the dependent variable. In such a case, that latent vector can be deleted so as to remove the unwanted effects of the near singularity from the parameter estimates involved in the singularity without major impact on the remaining estimates. The issue of just what constitutes "small" λ_j and γ_{0j} is not definitely resolved. Gunst et al (1976) defend a rule of thumb of λ_j of .3 or less coupled with a γ_{0j} of .1 or less, as indicators of nonpredictive singularity. The same rule of thumb is used in this analysis.

Results

Enforcing the identical specification of the model from year to year and across estimation techniques enables comparison of the temporal stability of the resulting coefficients with the OLS estimates serving as a benchmark. The comparison reveals a number of interesting differences.

1. In terms of incorrectly signed regression coefficients, both OLS and RR yield six estimates and LRR yields eight estimates that violate our a priori expectations. Interestingly, OLS and LRR failed to pick up the proper directional impact of the key variable, TARGET, that specifies the relative size of the true target market for sales of household furnishings. These same estimates under RR were all properly signed.

2. Relative to the number of statistically significant and appropriately signed coefficients, the ridge and latent root estimates provided very improved results. We recognize that the precise distribution of the ridge estimator is unknown. Several simulation studies suggest that, for modest values of k, any departure from the t-distribution will likely be minimal (Curcio et al 1984). Obenchain (1977) contends that the ridge estimator provides comparable F and t ratios as does OLS for hypothesis testing. As yet, an exact distributional theory and properties have not been derived for the LRR estimator since the latent roots and latent vectors are random variables with complex multivariate distributions. To employ the traditional F and t statistics, one must presume an appropriate underlying distribution. Webster, Gunst and Mason (1974) conducted simulation tests with generated data and observed an approximate F-distribution in the presence of vector deletion. Given the favorable results of Webster et al, F and t-ratios for LRR are interpreted cautiously. Only seventeen of forty-two OLS estimates were significant at the five percent level for a one-tailed test. In contrast, twenty four coefficients convincingly differentiated themselves from zero under ridge in spite of the bias which propels them toward zero, while twenty-five estimates were statistically significant under LRR. The increased significance for the ridge estimates and latent root is achieved through a substantial reduction in the individual coefficient variances.

3. Table 1 reports each technique's coefficient of variation to facilitate the comparison of temporal stability of the estimates. The ridge estimates appear, at first glance, to offer only mixed results regarding relative stability of the coefficients over time. However, for EBI and HH which are statistically significant under all three procedures, the coefficient of variation is lowest for ridge estimates. While the coefficient of variation for TARGET is lowest under LRR, the LRR estimates of TARGET were consistently inappropriately signed. Hence, even for TARGET, the ridge estimates offer greater stability over time and are properly signed. Also, as the population ages and consequently makes the transition from the classification of YOUNG to that of TARGET, the combined impact of a one percentage point reduction in YOUNG and a like increase in TARGET was found to be substantively more stable over time under RR.

TABLE 1

COMPARISON OF RELATIVE STABILITY OF ESTIMATES ACROSS TECHNIQUES σ/μ FOR 1976 THROUGH 1982

	OLS	RR	LRR	Technique With Least Relative Dispersion	Mean Coefficient for Selected Technique
UNEMP	.878	3.351	.863	LRR	-1.412
YOUNG	.268	.561	.369	OLS	-1.463
TARGET	1.846	.690	.590	RIDGE	1.946
OLD	.554	4.196	.592	OLS	-1.094
EBI	.209	.172	.207	RIDGE	.00840
HH	.0632	.0478	.0621	RIDGE	.3107
Combined Impact of Transition From Young to Target	.953	.226	1.213	RIDGE	3.338
UNEMPL 80-82	.211	.087	.238	RIDGE	-2.067

Furthermore, if one posits a structural change from the expansion years of 1976 through 1979 to the recession years of 1980 through 1982, one may expect a substantially different role for UNEMP in the model. While OLS, RR and LRR pick up on this with very different coefficient estimates for the recent period, the temporal stability of the recession years' coefficient is clearly improved with ridge.

4. The relative ability of the OLS, RR and LRR models to forecast may be very distorted by a simple comparison of their respective R^2 values. While the average multiple correlation coefficient does decline from about .96 for OLS and LRR to approximately .78 for ridge, this comparison, based entirely upon the estimation period, may not be indicative of the techniques' relative predictive capability beyond that period.

   To test each technique's ability to forecast furniture sales a year beyond the estimation period, the regression coefficients estimated using a given year's data set were employed employed to project the dollar sales volume for the subsequent year. The mean forecast for each technique was then compared to the mean actual sales for that year. Six such annual forecast comparisons were possible over the seven year data period. In three of the six years (1977, 1978, and 1980) the mean predicted value of the criterion variable is closer to the mean actual value for RR than for OLS or LRR. For 1979, the forecasts were nearly identical for OLS, LRR and RR. In 1982, the closest average is obtained via LRR while RR and OLS are nearly indistinguishable. Only for 1981 did the least squares prediction prove more accurate. In all cases, the dispersion of the forecast distribution is less with ridge than with OLS or LRR. Thus, in spite of lower R^2, the true predictive ability of RR beyond the estimation period is, on the whole, better than LRR and OLS.

Conclusion and Summary

It has been demonstrated in this article that the temporal stability of parameter estimates through RR is better than or at least as good as OLS and LRR in the presence of ill-conditioned data. This is particularly true for coefficients of those variables with a substantial impact on variance explanation. RR estimates also result in a reduction in the incidence of improperly signed parameters and an increase in the number of statistically significant coefficients. Hence, in this model, the ridge technique was better able to separate the influences of the regressor variables in spite of their interdependence. This permitted improved economic inferences and better detection of any underlying coefficient stability over time. Additionally, despite lower R^2 values for RR, the package predictive ability was not in any way sacrificed.

REFERENCES

Belsley, D.A., Kuh, E., & Welsch, R. Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. New York: John Wiley and Sons, 1980.

Curcio, R.J., P. Marquette & J. Webb. Improved Prediction of Real Estate Values Through the Use of Ridge Regression. Property Tax Journal, June 1984, 107-121.

Erickson, G.M. Using Ridge Regression to Estimate Directly Lagged Effects in Marketing. Journal of the American Statistical Association, 1981, 76, 766-773.

Farrar, D.E., & Glauber, R.R. Multicollinearity in Regression Analysis: The Problem Revisited. The Review of Economics and Statistics, 1967, 49, 92-107.

Gunst, R.F., Webster, J.R., & Mason, R.L. A Comparison of Least Squares and Latent Root Regression Estimators. Technometrics, 1976, 18, 75-83.

Hawkins, D.M. On the Investigation of Alternative Regressions by Principal Component Analysis. Applied Statistics, 1973, 22, 275-286.

Hawkins, D.M. Relations Between Ridge Regression and Eigenanalysis of the Augmented Correlation Matrix. Technometrics, 1975, 17, 477-480.

Hoerl, A.E., & Kennard, R.W. Ridge Regression: Biased Estimation for Non-Orthogonal Problems and Applications to Non-Orthogonal Problems. Technometrics, 1970, 12, 55-82.

Mahajan, Vijay, Jain, A.K., & Bergier, M. Parameter Estimation in Marketing Models in the Presence of Multicollinearity: An Application of Ridge Regression. Journal of Marketing Research, 1977, 14, 586-591.

Obenchain, R.L. Classical F Tests and Confidence Regions for Ridge Regression. Technometrics, 1977, 19, 4, 429-439.

Sharma, Subhash & James, W.L. Latent Root Regression: An Alternative Procedure for Estimating Parameters in the Presence of Multicollinearity. Journal of Marketing Research, 1981, 18, 154-161.

Thisted, R.A., & Morris, C.N. Theoretical Results for Adaptive Ordinary Ridge Regression Estimators. Technical Report 94, University of Chicago, 1980.

Vinod, H.D. A Survey of Ridge Regression and Related Techniques for Improvements Over Ordinary Least Squares. Review of Economics and Statistics, 1978, 60, 121-131.

Webster, J.R., Gunst, R.F., & Mason, R.L. Latent Root Regression Analysis. Technometrics, 1974, 16, 513-522.