DOCUMENT RESUME

ED 322 160 TM 015 251

AUTHOR Houston, Walter M.

TITLE Using Collateral Information from Similar Institutions To Predict College Freshman Grade Average. ACT Research Report Series 87-8.

INSTITUTION American Coll. Testing Program, Iowa City, Iowa.

PUB DATE Sep 88

NOTE 31p.; For related documents, see TM 015 252 and TM 015 254.

AVAILABLE FROM ACT Research Report Series, 2201 North Dodge Street, P.O. Box 168, Iowa City, IA 52243.

PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.

DESCRIPTORS Bayesian Statistics; *College Freshmen; Colleges; Comparative Analysis; Estimation (Mathematics); Evaluation Methods; Grade Point Average; *Grade Prediction; Higher Education; Least Squares Statistics; *Mathematical Models; Predictive Measurement; *Predictor Variables; Regression (Statistics); *School Statistics

IDENTIFIERS Bayesian M Group Regression; *Collateral Information; *Cross Validation; Pooled Least Squares Method

ABSTRACT

Two methods of using collateral information from similar institutions to predict college freshman grade average were investigated. One central prediction model, referred to as pooled least squares with adjusted intercepts, assumes that slopes and residual variances are homogeneous across selected colleges. The second model, referred to as Bayesian m-group regression, allows estimates of slopes and variances to vary across colleges without ignoring the available collateral information. These models were compared with the more usual procedure of deriving regression equations within each college considered in isolation from other colleges. Data were obtained from colleges that participated in the American College Testing predictive research services program during the 1983 and 1984 years, and that had fewer than 100 records in 1983. Two groups of colleges were used: (1) 9 four-year colleges with "liberal", or "open," enrollment; and (2) 10 two-year colleges with more than 20 freshmen over the age of 25 years. It was found that both models using collateral information resulted in more accurate predictions, on cross validation, than did the within-college model, and that the Bayesian approach slightly outperformed the pooled least squares approach. It is noted that the Bayesian simultaneous regression model is highly adaptive to different regression structures and therefore can be expected to perform as well as the other two models across most situations. Seven tables present study data. (Author/SLD)

ACT Research Report Series 87-8

Using Collateral Information From Similar Institutions to Predict College Freshman Grade Average

Walter M. Houston

September 1987

ACT

For additional copies write:
ACT Research Report Series
P.O. Box 168
Iowa City, Iowa 52243

USING COLLATERAL INFORMATION FROM SIMILAR INSTITUTIONS TO PREDICT COLLEGE FRESHMAN GRADE AVERAGE

Walter M. Houston

ABSTRACT

Two methods of using collateral information from similar institutions to predict college freshman grade average were investigated. One central prediction model, referred to as pooled least squares with adjusted intercepts, assumes that slopes and residual variances are homogeneous across selected colleges. The second model, referred to as Bayesian m-group regression, allows estimates of slopes and variances to vary across colleges without ignoring the available collateral information. These models were compared with the more usual procedure of deriving regression equations within each college considered in isolation from other colleges. It was found that both models employing collateral information resulted in more accurate predictions, on cross validation, than did the within-college model, and that the Bayesian approach slightly outperformed the pooled least squares approach. It is noted that the Bayesian simultaneous regression model is highly adaptive to different regression structures and therefore can be expected to perform as well as the other two models across most situations.

ACKNOWLEDGMENTS

The author wishes to thank Ming-Mei Wang for the computer program to perform the empirical Bayesian m-group regression analyses and Richard Sawyer for many helpful comments and suggestions.

USING COLLATERAL INFORMATION FROM SIMILAR INSTITUTIONS TO PREDICT FRESHMAN GRADE AVERAGE

The American College Testing Program offers predictive research services to postsecondary institutions that use ACT Assessment data in their admissions procedures. An important component of the predictive services is the capability of predicting freshman college grade point average (GPA) from a linear combination of the four subtests included in the ACT Assessment: the English Usage Test (E), the Mathematics Usage Test (M), the Social Studies Reading Test (SS), and the Natural Sciences Reading Test (NS), and from students' self-reported high school grades in these areas (ACT, 1987). Grade predictions can be provided to students, high school counselors, and colleges for each participating college selected by the students when they register for the ACT Assessment.

Currently, regression equations are calculated within each college separately using standard least squares methods. In calculating within-college equations (by whatever statistical procedure), one can encounter several practical problems. Among these potential problems are the necessity for "adequate" sample sizes within each college, the presence of negative regression weights, a lack of stability over time of estimated regression parameters, and the loss of predictive accuracy on cross validation. Under some circumstances, the need for adequate sample sizes would preclude the possibility of deriving separate regression equations for relevant subpopulations within a college. In addition to the use of within-college regression equations, other factors that could lead to these problems are the low reliability of available criterion measures, differing degrees of range restriction both within and across colleges due to disparate applicant populations and the criteria imposed for admittance, and different grading standards across colleges and across curricula within colleges.

It has long been thought that some improvement on within-college least squares equations could be realized by using collateral information from similar institutions through some form of central prediction system. Two categories of model specifications found in the literature surrounding central prediction systems appear to be the most reasonable. One central prediction model, based on classical statistical methods, is pooled least squares with adjusted intercepts (denoted ADJUST in this report). A key assumption of this model is that the population regression coefficients be approximately equal across selected colleges, but that the intercepts may be quite different (reflecting differences in difficulty level) and must be estimated separately. Another model, motivated from a Bayesian perspective, is referred to as the m-group regression model. There are several variations of m-group regression, ranging from empirical Bayesian to Bayesian. The model used in this investigation (denoted BAYES) is an extension of an empirical Bayesian model developed by Rubin (1980) and Braun, Jones, Rubin, and Thayer (1983). Another centralized prediction model proposed by Dempster, Rubin, and Tsutakawa (1981) is closely related to these empirical Bayesian models.

The within-college least squares model (denoted WCLS), the BAYES model, and the ADJUST model may be compared along a continuum. If all of the colleges were entirely different in their regression structures, then the WCLS model would likely be more appropriate than ADJUST. If all of the colleges were identical except for intercept, the ADJUST model would be appropriate. The BAYES model strikes a compromise between these two positions, and may be heuristically thought of as encompassing the other two models. Bayesian m-group regression brings to bear the available collateral information for the estimation of the regression parameters, while allowing for potential differences to exist among groups. Because the m-group regression model does not commit one to rigid a priori assumptions about the regression structures of the colleges, it may prove to be more flexible than WCLS and ADJUST.

It is important to base evaluations of alternative prediction systems on criteria that reflect the manner in which the prediction equations are used. There are at least three statistical criteria on which to compare prediction models that take into account the intended uses of the prediction equations provided by ACT to postsecondary institutions. The first criterion is the predictive accuracy realized from the model predictions upon cross validation over time. The second criterion on which to compare models is the stability of the estimated regression parameters over time. The third criterion is the amount of prediction bias introduced by use of the model. Prediction bias, as used in this analysis, is defined as the expected difference between predicted and obtained criterion values, where the expectation is with respect to hypothetical base year and cross validation year populations.

Model Specifications

The observable quantities consist of the criterion scores Y_{ij} (first semester college GPA) and the predictor variables X_{ijk} (ACT subtest scores and high school grades) for $i = 1,\ldots,n_j$ students, $j = 1,\ldots,m$ colleges, on $k = 1,\ldots,p$ predictor variables. Let $n = \sum_{j=1}^{m} n_j$ denote the total number of observations across all m colleges.

<u>Within-College Least Squares (WCLS)</u>

The regression model within each college j is given by

$$Y_{ij} = \alpha_j + \sum_{k=1}^{p} \beta_{jk} X_{ijk} + e_{ij} \qquad j = 1,\ldots,m$$

where e_{ij} is normally distributed with mean 0 and variance σ_j^2

σ_j^2 is the residual variance at college j

α_j is the intercept for college j

β_{jk} is the regression slope for variable k at college j

Y_{ij} is the observed GPA for student i at college j.

This is the ordinary least squares regression model with independent, normally distributed homoscedastic error terms. Under this model regression slopes, intercepts, and residual variances are allowed to vary across colleges.

Pooled Least Squares with Adjusted Intercepts (ADJUST)

The regression model is given by

$$Y_{ij} = \alpha_j + \sum_{k=1}^{p} \beta_k X_{ijk} + e_{ij} \qquad j=1,\ldots,m$$

where e_{ij} is normally distributed with mean 0 and variance σ^2

σ^2 is the common residual variance across colleges

β_k is the common regression slope for variable k across the m colleges.

All other notation is as previously defined.

Under this model, the intercepts are allowed to vary while the slopes and residual variances are assumed constant across colleges. Thus, the regression surfaces within each college are assumed parallel but not coincident. Note that the model assumes homoscedasticity of residual variances both within and across colleges.

M-group Regression (BAYES)

The m-group regression model uses the observed variability in regression coefficients and residual variances across the m groups to estimate the within-group parameters. The m-group parameter estimates are a weighted average of the individual within-group estimates and the estimates obtained from a pooled analysis.

The m-group regression model is hierarchical and can be described in three stages. While we distinguish between empirical Bayesian and Bayesian models, the first two stages are identical in both approaches. At Stage 1, the standard normal linear regression model within each college j is assumed.

$$Y_{ij} = \sum_{k=0}^{p} \beta_{jk} X_{ijk} + e_{ij} \qquad j=1,\ldots,m \qquad (1)$$

This is the same as the WCLS model; the notation has been altered slightly by introducing a dummy suffix $k = 0$ and a dummy variable $X_{ijk} = 1$ in order to include the intercept as another regression weight. For subsequent development, rewrite (1) in matrix notation as

$$\underset{\sim}{Y}_j = \underset{\sim}{X}_j \underset{\sim}{\beta}_j + \underset{\sim}{e}_j \quad j=1,\ldots,m$$

It is well known from the theory of linear models that the conditional sampling distribution of the maximum likelihood estimates of the regression parameters $\underset{\sim}{\beta}_j$ (denoted $\hat{\underset{\sim}{\beta}}_j$, where $\hat{\underset{\sim}{\beta}}_j$ is a (p+1) x 1 random vector) has a multivariate normal distribution:

$$(\hat{\underset{\sim}{\beta}}_j | \underset{\sim}{\beta}_j, \sigma_j^2, \underset{\sim}{X}_j) \sim N_{p+1} [\underset{\sim}{\beta}_j, \sigma_j^2 (\underset{\sim}{X}_j{}' \underset{\sim}{X}_j)^{-1}] \tag{2}$$

At Stage 2 the Bayesian part of the model is introduced by the assumption that the unobservable vectors of regression parameters $\underset{\sim}{\beta}_j$ are independent realizations from a multivariate normal distribution with mean vector $\underset{\sim}{\mu}$ and a positive definite covariance matrix $\underset{\sim}{\Sigma}$:

$$(\underset{\sim}{\beta}_j | \underset{\sim}{\mu}, \underset{\sim}{\Sigma}) \sim N_{p+1} [\underset{\sim}{\mu}, \underset{\sim}{\Sigma}] \tag{3}$$

The quantities $\underset{\sim}{\mu}$ and $\underset{\sim}{\Sigma}$ are referred to as hyperparameters. In a fully Bayesian approach and in the approach utilized in this research, it is also assumed that the residual variances σ_j^2 are independent realizations from an inverse chi-square distribution with specified degrees of freedom used to incorporate the strength of prior information.

Given the prior belief that the m colleges (or given subpopulations within each college) have similar characteristics, the colleges are said to constitute exchangeable units. The reader is referred to Lindley (1971) for further discussion of the important concept of exchangeability. For present purposes, the assumption of exchangeability permits one to act as though the unobservable parameters were randomly sampled from the stated distributions, although no actual random sampling of colleges is implied.

If the residual variances σ_j^2 and the hyperparameters $\underset{\sim}{\mu}$ and $\underset{\sim}{\Sigma}$ were known, standard Bayesian results (DeGroot, 1970; Box & Tiao, 1973) show that the posterior distributions of the $\underset{\sim}{\beta}_j$ are independent, multivariate normal with parameters

$$E[\underset{\sim}{\beta}_j|\hat{\underset{\sim}{\beta}}_j,\underset{\sim}{\mu},\underset{\sim}{\Sigma},\sigma_j^2,\underset{\sim}{X}_j] = [\sigma_j^{-2}(\underset{\sim}{X}_j'\underset{\sim}{X}_j) + \underset{\sim}{\Sigma}^{-1}]^{-1}[\sigma_j^{-2}(\underset{\sim}{X}_j'\underset{\sim}{X}_j)\,\hat{\underset{\sim}{\beta}}_j + \underset{\sim}{\Sigma}^{-1}\underset{\sim}{\mu}] \quad (4)$$

$$\mathrm{Var}[\underset{\sim}{\beta}_j|\hat{\underset{\sim}{\beta}}_j,\underset{\sim}{\mu},\underset{\sim}{\Sigma},\sigma_j^2,\underset{\sim}{X}_j] = [\sigma_j^{-2}(\underset{\sim}{X}_j'\underset{\sim}{X}_j) + \underset{\sim}{\Sigma}^{-1}]^{-1} \quad (5)$$

The posterior mean centroid of $\underset{\sim}{\beta}_j$ is a weighted average of the sample estimates $\hat{\underset{\sim}{\beta}}_j$ and the mean of the prior distribution $\underset{\sim}{\mu}$, with the weights being the sampling precision of the $\hat{\underset{\sim}{\beta}}_j$ and the prior precision, respectively. The posterior precision is the sum of the sampling precision and the prior precision. (Precision is the inverse of variance.) Note that the estimate of $\underset{\sim}{\beta}_j$ is regressed toward the mean centroid of the prior distribution. The concept of regressed estimates of unobservable parameters is prominent both in Bayesian statistics and in classical psychometric theory.

The distinction between Bayesian and empirical Bayesian approaches to m-group regression arises because the hyperparameters $\underset{\sim}{\mu}$ and $\underset{\sim}{\Sigma}$ and the residual variances σ_j^2 are not known in most applications. Their determination constitutes the third stage of the hierarchical model. A Bayesian approach places prior distributions with fully specified parameters on these unknown quantities, whereas empirical Bayesian approaches estimate the prior distributions from the observed data.

In a Bayesian approach, the elements of the vector $\underset{\sim}{\mu}$ are typically assumed to follow a multivariate normal distribution, $\underset{\sim}{\Sigma}$ is taken to have a Wishart distribution, and the σ_j^2 are taken to be inverse chi-square (see Novick, Jackson, Thayer, & Cole; 1972). With these prior specifications, the joint distribution of the data, parameters, and hyperparameters can be found. In principle, the joint posterior density of the parameters is then obtained by integrating out the hyperparameters and conditioning on the data, though the estimation of the prior distributions and the numerical techniques involved in obtaining the joint posterior distributions are complex. The reader is referred to Lindley (1970), Jackson, Novick, and Thayer (1971), Novick, Jackson, Thayer, and Cole (1972) for details. Although not the approach used in this study, a simplified version of a Bayesian approach to m-group regression, developed by Molenaar and Lewis (1979) and employed by Dunbar, Mayekawa, and Novick (1986), appears to be promising. The Molenaar-Lewis model places greater restrictions on the specification of prior information in order to increase computational efficiency and avoid problems in estimation.

In the empirical Bayesian approaches developed previously, maximum likelihood estimates of $\underset{\sim}{\mu}, \underset{\sim}{\Sigma}$, and σ_j^2 ($j=1,\ldots,m$) are obtained from the data via implementation of the EM algorithm (Dempster, Laird, & Rubin, 1977). The joint likelihood function is integrated over the distribution of $\underset{\sim}{\beta}_j$ to produce a marginal likelihood; the EM algorithm is then used to obtain estimates of $\underset{\sim}{\mu}$, $\underset{\sim}{\Sigma}$, and σ_j^2 that maximize the marginal likelihood. The $\underset{\sim}{\beta}_j$ are then estimated from their conditional posterior distribution, conditioned on these maximum likelihood estimates and the data.

The approach used in this study to estimate $\underset{\sim}{\mu}$, $\underset{\sim}{\Sigma}$, and σ_j^2 is a refinement of the empirical Bayesian approaches. Rather than estimate the residual variances σ_j^2 by the method of maximum likelihood, the current model allows for an informative prior distribution on the residual variances. In the current implementation, data-based estimates of the degrees of freedom and the scale parameter of the inverse chi-square distribution for the exchangeable within-college error variances are obtained. Residual variances are estimated by forming a weighted average of the mean of the prior distribution and the maximum likelihood estimates of the residual variances within each group. The model developed by Rubin (1980) and Braun et al. (1983) in effect places a locally uniform prior distribution (Box & Tiao, 1973) on the residual variances. As within-group sample sizes increase, the results obtained from the current approach and the approach used by Braun et al. (1983) converge. Informal comparisons made during this study indicate that even for small sample sizes the two approaches yield similar results.

For illustrative purposes, assume only one predictor variable and that the predictor and criterion variables have been standardized to zero mean and unit variance within each group (to avoid dealing with an intercept term). Analogous to equations (2) and (3) assume that

$$(\hat{\beta}_j | \beta_j, \sigma_j^2, \underset{\sim}{X}_j) \sim N[\beta_j, \sigma_j^2(\hat{\beta}_j)],$$

where $\sigma_j^2(\hat{\beta}_j) = \sigma_j^2 / \sum_{i=1}^{n_j} X_{ij}^2$ is the sampling variance of $\hat{\beta}_j$,

and $(\beta_j | \mu, \Phi) \sim N[\mu, \Phi]$,

where Φ represents the between group variance of the single regression weight β_j. The regression slope for college j can be estimated by the mean of its posterior distribution.

$$E[\beta_j | \hat{\beta}_j, \mu, \Phi, \sigma_j^2, \underset{\sim}{X}_j] = [\sigma_j^{-2}(\hat{\beta}_j) + \Phi^{-1}]^{-1} [\sigma_j^{-2}(\hat{\beta}_j)\hat{\beta}_j + \Phi^{-1}\mu] \qquad (6)$$

Equation (6) is the scalar equivalent of equation (4). Two observations are apparent from equation (6). As Φ, the "between-group" variance of the β_j, increases relative to the within-group sampling variance $\sigma_j^2(\hat{\beta}_j)$, greater emphasis is placed on the data from college j considered in isolation from the other m-1 colleges. For simultaneous regression procedures to prove more effective than within-college least squares, care must be taken to identify colleges that constitute exchangeable units. The weighted average approach also provides some protection against the inclusion of non-exchangable units.

Equation (6) also indicates that as the sampling precision of the $\hat{\beta}_j$ increases (through increased sample sizes or more properly selected design points), more emphasis is placed on estimates of β_j obtained from group j data considered in isolation. Conversely, if large sample sizes are not available, estimates of regression parameters may be substantially regressed toward the mean obtained from the exchangeable units. Simultaneous regression procedures are likely to prove more effective than within-group least squares when there are a sizable number of exchangeable units with small to moderate sample sizes available within each unit.

Method

Data Source

Data available for this investigation were obtained from colleges that participated in the ACT predictive research services during the 1983 and 1984 academic years, and that had fewer than 100 records in 1983. These data were a subset of data analyzed by Sawyer (1987). Of the 125 colleges in the data set, two groups were selected for subsequent analysis.

Group 1 colleges were selected from among four-year public institutions whose self-described freshman admission policies were "liberal" or "open." Hierarchical cluster analysis was used to select a subset of these colleges based on the percentages of students enrolled in various programs and majors. The nine colleges selected were characterized as having the vast majority of students enrolled in fine arts, humanities, and foreign language programs.

Group 2 colleges consisted of two-year public institutions with freshmen over the age of 25 years. Ten two-year public colleges were selected for which the number of freshmen over the age of 25 years was greater than 20 in both the

1983 and 1984 school years. The need for adequate sample sizes and for a moderate number of similar colleges within each group precluded using data from colleges with more selective admission policies.

Sample sizes for Group 1 and Group 2 colleges in both the 1983 and 1984 school years are presented in Table 1.

TABLE 1

Available Sample Sizes

Group	College	Year 1983-1984	Year 1984-1985
1	1	67	59
	2	56	73
	3	71	53
	4	56	98
	5	72	54
	6	51	49
	7	98	104
	8	51	51
	9	50	50
	(Total)	572	591
2	1	32	32
	2	34	75
	3	32	27
	4	28	50
	5	68	55
	6	53	58
	7	28	37
	8	22	27
	9	31	37
	10	47	47
	(Total)	375	445

The colleges within Group 1 and the colleges within Group 2 were considered to be exchangeable for the Bayesian portion of the analysis.

Procedure

Predictor variables of interest in this study are the four subtests comprising the ACT Assessment (E, M, SS, and NS) and high school grade point average (HSA). The criterion variable is first semester grade point average (GPA), reported on a scale from 0.0 to 4.0. Preliminary inspections of bivariate scatterplots were made for each college in order to identify any serious departures from the linearity and homoscedasticity assumptions of the within-college regression models. No serious violations of these assumptions were found.

Three separate regression models were applied to the nine Group 1 colleges for the 1983 base year. The three regression models were within-college least squares (WCLS), pooled least squares with adjusted intercepts (ADJUST), and Bayesian m-group regression across the nine colleges (BAYES). The prediction equations derived from each of these three models were then cross validated using 1984 data from the same schools. These procedures were repeated for the 10 Group 2 colleges using data only for students age 25 or over.

There are several criteria for comparing predicted versus obtained GPA. The cross validation analyses utilized three of the most common criteria: mean squared error (MSE), mean absolute error (MAE), and the squared correlation coefficient (R^2). MSE is defined as the squared deviation between predicted and observed GPA averaged across students at a given college. MAE is defined as the mean absolute deviation between predicted and observed GPA. R^2 is the squared zero-order correlation between predicted and observed GPA at a given college.

Cross validated prediction bias for non-traditional aged freshmen (over the age of 25 years) in the Group 2 colleges was also calculated. The following identity was used in the computation: $E(d^2) = Var(d) + BIAS^2$, where d is the prediction error, E denotes the expectation operator, and BIAS = E(d). The quantity Var(d) corresponds to error variance and the quantity BIAS represents prediction bias. Prediction bias for nontraditional aged freshmen was computed with respect to the three models WCLS, ADJUST, and BAYES, as well as with respect to the total group within-college least squares regression model that employed all freshman records from each college.

The definition of prediction bias used in this study provides an estimate of the average bias which occurs over the range of the predictor score scales and across all examinees. Houston and Novick (1987) have demonstrated that these indices of average bias may be misleading if there are selected cut-off points on the predictor variables. In such situations, regression equations derived from various models should be compared at these cut-off points. However, indices of average bias do provide one useful method for comparing how various models perform overall on cross validation.

Results

Group 1

The estimated regression parameters obtained from the within-group least squares (WCLS), the m-group regression (BAYES), and the pooled least squares with adjusted intercepts (ADJUST) models for Group 1 colleges during the 1983 school year are presented in Table 2.

TABLE 2

Estimated Regression Coefficients and Residual Variances
Group 1: 1983-1984 School Year

No.	Prediction method	Intercept	ACT English	ACT Math	ACT Social Studies	ACT Natural Sciences	HS Average	Residual Variance
1	WCLS	1.3214	.0653	.0280	.0036	-.0030	.1776	.6371
	BAYES	1.6718	.0466	.0188	.0070	.0071	.0826	.5775
	ADJUST	1.1270	.0392	.0209	.0031	.0125	.3132	.4393
2	WCLS	.8570	-.0268	.0547	.0074	.0080	.3198	.4048
	BAYES	.4529	.0102	.0361	.0052	.0116	.4143	.4248
	ADJUST	.2391	.0392	.0209	.0031	.0125	.3132	.4393
3	WCLS	.7329	.0228	.0273	.0057	-.0014	.3614	.3287
	BAYES	.6096	.0301	.0240	.0005	.0090	.3288	.3545
	ADJUST	.4155	.0392	.0209	.0031	.0125	.3132	.4393
4	WCLS	.4345	.0383	.0167	-.0100	.0131	.3858	.2673
	BAYES	.3811	.0303	.0228	.0003	.0116	.3729	.3200
	ADJUST	.3251	.0392	.0209	.0031	.0125	.3132	.4393
5	WCLS	.5479	.0475	.0146	-.0018	.0163	.2641	.4992
	BAYES	.5535	.0389	.0178	.0040	.0099	.3088	.4741
	ADJUST	.4456	.0392	.0209	.0031	.0125	.3132	.4393
6	WCLS	.0271	.0662	-.0129	.0360	.0161	.2450	.4686
	BAYES	.1960	.0553	.0047	.0116	.0136	.3171	.4631
	ADJUST	.3903	.0392	.0209	.0031	.0125	.3132	.4393
7	WCLS	.2493	.0555	.0072	.0065	.0156	.3032	.4022
	BAYES	.3487	.0461	.0120	.0069	.0119	.3236	.4059
	ADJUST	.4023	.0392	.0209	.0031	.0125	.3132	.4393
8	WCLS	.2421	.0130	.0264	.0098	.0223	.3600	.4969
	BAYES	.3171	.0276	.0238	.0000	.0127	.3858	.4696
	ADJUST	.2913	.0392	.0209	.0031	.0125	.3132	.4393
9	WCLS	-.7201	.0450	.0218	-.0238	.0053	.7638	.1848
	BAYES	-.3261	.0191	.0269	-.0057	.0190	.5439	.2809
	ADJUST	.0149	.0392	.0209	.0031	.0125	.3132	.4393

The within-college least squares estimates would seem to confirm that the institutions are somewhat similar. Notable features of the results include the presence of negative regression weights and the relatively small magnitude of the weights associated with the ACT Social Studies and Natural Sciences subtests across all nine colleges. The general effect of the m-group regression procedure has been to regress the within-group estimates toward the estimates obtained from the ADJUST analysis. Shrinkage of parameter estimates towards pooled values has the effect of eliminating the negative weights derived under the WCLS model. Note that the BAYES estimates of the regression parameters remain distinct across colleges. Although not reported in the table, squared correlations from the within-college analysis ranged from .16 to .67.

The results from the cross validation analysis for Group 1 colleges in the 1984 school year are given in Table 3. The table contains mean squared errors (MSE), mean absolute errors (MAE) and squared correlations (R^2) between predicted and observed criterion scores.

TABLE 3

Mean Squared Error, Mean Absolute Error, and Squared Multiple Correlation for Cross Validation Analysis of Group 1: 1984-1985 School Year

College	Prediction method	MSE	MAE	R^2
1	WCLS	.2712	.4139	.2187
	BAYES	.2611	.4042	.2572
	ADJUST	.2628	.4118	.2311
2	WCLS	.4529	.5660	.4331
	BAYES	.3718	.5061	.5444
	ADJUST	.3577	.4959	.5554
3	WCLS	.6029	.5432	.2318
	BAYES	.5986	.5330	.2357
	ADJUST	.6035	.5375	.2343
4	WCLS	.3484	.4678	.3113
	BAYES	.3451	.4601	.3162
	ADJUST	.3470	.4604	.2982
5	WCLS	.5780	.5135	.4400
	BAYES	.5612	.5100	.4612
	ADJUST	.5621	.5051	.4559
6	WCLS	.3768	.5051	.4181
	BAYES	.3292	.4820	.4872
	ADJUST	.3715	.5019	.4844
7	WCLS	.4825	.5440	.2756
	BAYES	.4716	.5350	.2863
	ADJUST	.4742	.5362	.2916
8	WCLS	.6328	.6291	.1414
	BAYES	.6118	.6283	.1490
	ADJUST	.6276	.6272	.1303
9	WCLS	.9989	.7810	.1772
	BAYES	.9756	.7518	.1875
	ADJUST	.9974	.7758	.1681
(AVERAGE)	WCLS	.5272	.5515	
	BAYES	.5029	.5345	
	ADJUST	.5115	.5391	

The results in Table 3 indicate a small yet consistent trend toward smaller errors of prediction on cross validation using an m-group regression model than those obtained from the classical models. These results are consistent with previous comparisons of m-group regression with conventional approaches (Novick et al., 1972). The average reduction in MSE, comparing the BAYES model to the WCLS model, was about 5%. Some improvement in MSE was found in each of the nine colleges. Somewhat smaller reductions were found for MAE, though the general trend was the same. Differences between the BAYES and ADJUST models in both MSE and MAE were very small.

Group 2

Table 4 presents the estimated regression parameters obtained from the three models for Group 2 colleges during the 1983 school year.

TABLE 4

**Estimated Regression Coefficients and Residual Variances
Group 2: 1983-1984 School Year**

No.	Prediction method	Intercept	ACT English	ACT Math	ACT Social Studies	ACT Natural Sciences	HS Average	Residual Variance
1	WCLS	2.3260	-.0097	-.0088	.0266	.0101	.1151	.2354
	BAYES	1.4728	.0303	-.0041	.0223	.0019	.2666	.2823
	ADJUST	1.2957	.0351	-.0071	.0239	-.0018	.3025	.3716
2	WCLS	.9640	.0204	.0086	-.0006	.0060	.4797	.2661
	BAYES	1.0017	.0287	.0028	.0127	.0026	.3767	.2834
	ADJUST	1.1366	.0351	-.0071	.0239	-.0018	.3025	.3716
3	WCLS	1.0038	.0626	.0017	-.0062	-.0273	.3710	.3591
	BAYES	.7757	.0410	-.0044	.0143	-.0054	.3793	.3748
	ADJUST	.8081	.0351	-.0071	.0239	-.0018	.3025	.3716
4	WCLS	2.4561	.0060	-.0308	.0598	-.0205	.1627	.3747
	BAYES	1.8813	.0346	-.0109	.0311	-.0066	.1687	.3811
	ADJUST	1.5632	.0351	-.0071	.0239	-.0018	.3025	.3716
5	WCLS	.3790	.0660	-.0335	.0187	.0186	.4467	.4679
	BAYES	1.1846	.0383	-.0060	.0203	.0036	.3078	.4836
	ADJUST	1.1741	.0351	-.0071	.0239	-.0018	.3025	.3716
6	WCLS	1.6193	.0497	-.0294	.0382	-.0161	.1014	.5791
	BAYES	1.1917	.0412	-.0091	.0222	-.0076	.2901	.5600
	ADJUST	1.0324	.0351	-.0071	.0239	-.0018	.3025	.3716
7	WCLS	1.7335	.0254	-.0107	.0267	.0010	.2652	.1256
	BAYES	1.7225	.0304	-.0056	.0260	-.0021	.2192	.1774
	ADJUST	1.5069	.0351	-.0071	.0239	-.0018	.3025	.3716
8	WCLS	.6735	.0016	-.0003	-.0296	.0600	.5346	.3660
	BAYES	1.0146	.0271	.0040	.0121	.0042	.3807	.3853
	ADJUST	1.1667	.0351	-.0071	.0239	-.0018	.3025	.3716
9	WCLS	.7593	.0026	.0266	.0173	.0026	.4680	.1828
	BAYES	.8738	.0257	.0064	.0093	.0057	.4141	.2278
	ADJUST	1.0918	.0351	-.0071	.0239	-.0018	.3025	.3716
10	WCLS	1.7222	.0492	-.0001	.0223	-.0149	.1849	.1374
	BAYES	1.6315	.0323	-.0045	.0240	-.0025	.2352	.1757
	ADJUST	1.4381	.0351	-.0071	.0239	-.0018	.3025	.3716

Variation from group to group in the magnitude of the within-college least squares weights is evident, with a large number of estimates taking on negative values. In the absence of other data, a reasonable explanation of this finding is that the negative weights are due, in part, to the small within-group sample sizes employed and that the "true" coefficients are very small. However, a rather disturbing feature of the results presented in Table 4 is the negative weights associated with the mathematics and natural science subtests obtained from the ADJUST analysis in which a sample size of 375 was available. Once again, the general effect of the m-group regression procedure was to shrink parameter estimates toward common values. Note from Table 4 that m-group regression is not effective in eliminating negative regression weights when the weights derived from the pooled analysis are themselves negative. For those variables in which the ADJUST model yielded positive weights, the Bayesian procedure also proved effective in eliminating the negative weights obtained from the WCLS model. Although not reported, squared correlations from the within-college WCLS model ranged from .13 to .60.

Results from the cross validation analysis of Group 2 colleges are given in Table 5.

TABLE 5

Mean Squared Error, Mean Absolute Error, and Squared Multiple Correlation for Cross Validation Analysis of Group 2: 1984-1985 School Year

College	Prediction method	MSE	MAE	R^2
1	WCLS	.6352	.6626	.1056
	BAYES	.5262	.5769	.1589
	ADJUST	.6197	.6463	.1560
	WCLS (ALL)	.6190	.6476	.1505
2	WCLS	.3433	.4689	.5271
	BAYES	.3463	.4684	.5407
	ADJUST	.3796	.4871	.4775
	WCLS (ALL)	.3111	.4731	.4816
3	WCLS	.2593	.4158	.3856
	BAYES	.1880	.3433	.4543
	ADJUST	.2085	.3652	.4476
	WCLS (ALL)	.2272	.3868	.3469
4	WCLS	.4171	.4524	.1376
	BAYES	.4009	.4420	.1401
	ADJUST	.4301	.4517	.1183
	WCLS (ALL)	.4262	.4961	.1552
5	WCLS	.4205	.5021	.2809
	BAYES	.3610	.4808	.3557
	ADJUST	.3598	.4810	.3560
	WCLS (ALL)	.4327	.5100	.2735
6	WCLS	.4430	.5440	.1452
	BAYES	.3447	.4760	.2632
	ADJUST	.3851	.5094	.2533
	WCLS (ALL)	.4466	.5406	.1764
7	WCLS	.2667	.4129	.2218
	BAYES	.2512	.3990	.2305
	ADJUST	.2747	.4136	.2210
	WCLS (ALL)	.2874	.4396	.1849

TABLE 5 (continued)

College	Prediction method	MSE	MAE	R^2
8	WCLS	.6569	.6600	.2391
	BAYES	.6572	.6608	.2450
	ADJUST	.7139	.7137	.1945
	WCLS (ALL)	.8080	.7650	.2704
9	WCLS	.4479	.5278	.3612
	BAYES	.4311	.5226	.3516
	ADJUST	.4730	.5458	.3170
	WCLS (ALL)	.4941	.5443	.3493
10	WCLS	.3491	.4984	.0829
	BAYES	.3284	.4834	.0906
	ADJUST	.3434	.4990	.0773
	WCLS (ALL)	.3438	.5089	.0762
(AVERAGE)	WCLS	.4239	.5141	
	BAYES	.3835	.4853	
	ADJUST	.4189	.5113	
	WCLS (ALL)	.4396	.5309	

In order to compare the methods investigated in this study with the currently used model (total group within-college least squares), indices of predictive accuracy obtained from the cross validation of within-college least squares equations derived from all freshman records in the 1983 data set are presented under the model labeled WCLS (ALL). It is evident from Table 5 that using the Bayesian m-group regression model resulted in an increase in predictive accuracy compared to any of the other models investigated. Most notably, the m-group regression model achieved a 12.8% reduction in average MSE compared to the WCLS (ALL) model, and a 9.6% reduction in average MSE compared to the specific group WCLS model. The use of the Bayesian model resulted in a reduction in average MAE of 8.8% and 5.6% compared to WCLS (ALL) and WCLS, respectively. The use of the Bayesian model attained reductions of 8.5% and 5.1% in average MSE and MAE, respectively, compared to the ADJUST model.

Table 6 presents a comparison of the prediction bias that resulted from the use of the four models for Group 2 colleges.

TABLE 6

Bias Analysis for Group 2: 1984-1985 School Year

Average Absolute Bias Obtained from Regression Equations Evaluated Across Range of Predictor Score Scales Expressed in Raw Score Units

Prediction method	Prediction bias*
WCLS	.1851
BAYES	.1426
ADJUST	.175[illegible]
WCLS (ALL)	.2250

*Average absolute BIAS weighted by within college sample sizes

Note that use of the BAYES model resulted in less prediction bias than the other three models. Note also that all three models that utilized only those data from freshmen over the age of 25 attained substantially less prediction bias than the WCLS (ALL) model that derived regression equations based on all freshmen records in a college. The prediction bias reported in Table 6 was calculated by forming a weighted average of the absolute bias within each college.

In order to compare the stability of the estimated regression parameters for the BAYES, ADJUST, and WCLS models over time, estimates of these parameters were obtained for Group 2 colleges in the 1984 data set in addition to estimates obtained from the 1983 data set already presented.

TABLE 7

Absolute Differences in Estimated Regression Parameters for 1983 and 1984 School Years Averaged Across Group 2 Colleges

Parameter	Prediction Method		
	WCLS	BAYES	ADJUST
ACT English	.025	.005	.006
ACT Mathematics	.031	.017	.026
ACT Social Studies	.021	.008	.015
ACT Natural Sciences	.023	.004	.004
H.S. Average	.140	.091	.027

Table 7 presents the absolute differences between the 1984 estimates and the 1983 estimates for each predictor variable for the three models averaged across colleges. The results in Table 7 indicate that the BAYES and ADJUST estimates are substantially less variable over time than the WCLS estimates. The greater stability of estimates obtained from the BAYES and ADJUST models suggests that using collateral information reduces the effects of year-to-year sampling fluctuations.

Discussion

The results of this study indicate that increases, both in predictive accuracy obtained on cross validation and in the stability of the estimated regression parameters over time, can be realized from the use of a Bayesian simultaneous prediction method. Increases in predictive accuracy were also attained by use of the ADJUST model, in which regression slopes are assumed approximately equal across selected institutions, while intercepts are allowed to vary. The results of this investigation provide evidence that the use of collateral information from similar institutions in the construction of prediction equations lead to increases in predictive accuracy and decreases in prediction bias obtained on cross validation. Although the empirical Bayesian method performed somewhat better on cross validation than the pooled least squares with adjusted intercepts method, the advantages may be offset by the increased cost, due to the added numerical complexity of the BAYES model.

An advantage of Bayesian simultaneous prediction methods is that the estimates of regression slopes and residual variances are allowed to vary across colleges, while traditional least squares methods either assume homogeneity of slopes and variances across colleges (ADJUST) or fail to utilize any collateral information (WCLS). It should be noted that the colleges in this study were selected to be very similar, and thus to make the Bayesian and pooled least squares procedures perform well. It has yet to be determined whether or not either procedure can perform meaningfully better than within-college least squares methods in more general situations. Because the Bayesian approach is highly adaptive to different regression structures, the BAYES model can be expected to perform as well as the other two models across the vast majority of situations. The justification for Bayesian simultaneous regression hinges on whether the flexibility inherent in the Bayesian system can achieve meaningful improvements over more easily implemented approaches.

The greatest potential for centralized prediction systems has to do with special prediction situations involving small numbers of students. Such situations include the prediction of specific course grades, the calculation of prediction equations for socially or educationally relevant subgroups, and the calculation of regression equations for small colleges with limited numbers of ACT tested students. From either a classical or Bayesian perspective, the use of collateral information from similar institutions may provide a viable alternative to within-college least squares regression equations in situations such as these.

REFERENCES

The American College Testing Program. (1987). ACT Research and Information Services. Iowa City, IA: Author.

Box, G. P., & Tiao, G. (1973). Bayesian inference in statistical analysis. Reading, MA: Addison-Wesley.

Braun, H. I., Jones, D. H., Rubin, D. B., & Thayer, D. T. (1983). Empirical Bayes estimation of coefficients in the general linear model from data of deficient rank. Psychometrika, 48, 171-181.

DeGroot, M. H. (1970). Optimal statistical decisions. New York: McGraw-Hill.

Dempster, A. P., Rubin, D. B., & Tsutakawa, R. K. (1981). Estimation in covariance components models. Journal of the American Statistical Association, 76, 341-353.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, 39, 1-38.

Dunbar, S. B., Mayekawa, S., & Novick, M. R. (1986). Simultaneous estimation of regression functions for Marine Corps technical training specialities. Journal of Educational Statistics, 11, 275-292.

Houston, W. M., & Novick, M. R. (1987). Race-based differential prediction in Air Force technical training programs. Journal of Educational Measurement, 24, 309-320.

Jackson, P. H., Novick, M. R., & Thayer, D. T. (1971). Estimating regressions in m-groups. British Journal of Mathematical and Statistical Psychology, 24, 129-153.

Lindley, D. V. (1970). A Bayesian solution for some educational problems III. Research Bulletin 70-33. Princeton, NJ: Educational Testing Service.

Lindley, D. V. (1971). The estimation of many parameters. In V. P. Godambe & D. A. Sprott (eds.), Foundations of statistical inference. Toronto: Holt, Rinehart, and Wilson.

Molenaar, I. W., & Lewis, C. (1979). An improved model and computer program for Bayesian m-group regression. (ONR Technical Report 79-5). Iowa City, IA: College of Education, the University of Iowa.

Novick, M. R., Jackson, P. H., Thayer, D. T., & Cole, N. S. (1972). Estimating multiple regressions in m-groups: A cross validation study. British Journal of Mathematical and Statistical Psychology, 25, 33-50.

Rubin, D. B. (1980). Using empirical Bayes techniques in the law school validity studies. Journal of the American Statistical Association, 75, 801-816.

Sawyer, R. L. (1987). Determining minimum sample sizes for estimating prediction equations for college freshman grade average. (ACT Research Report No. 87-4). Iowa City, Iowa: The American College Testing Program.