ACT Research Report Series 88-4

# Central Prediction Systems for Predicting Specific Course Grades

Walter Houston
Richard Sawyer

December 1988

ACT

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

Central Prediction Systems for
Predicting Specific Course Grades

Walter Houston and Richard Sawyer
The American College Testing Program

Acknowledgment

The authors thank Ming-Mei Wang for the computer program to perform the empirical Bayesian m-group regression analyses.

## ABSTRACT

Methods for predicting specific college course grades, based on small numbers of observations, were investigated. These methods use collateral information across potentially diverse institutions to obtain refined within-group parameter estimates. One method, referred to as pooled least squares with adjusted intercepts, assumes that slopes and residual variances are homogeneous across selected colleges. The second method, referred to as Bayesian m-group regression, allows estimates of slopes and residual variances to vary across colleges, without ignoring available collateral information. These central prediction 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 for both models employing collateral information, a sample size of 20 resulted in a level of crossvalidated prediction accuracy comparable to that obtained using the within-college least squares procedure at colleges with 50 or more observations. The Bayesian approach outperformed the pooled least squares approach. It is noted that the Bayesian approach is highly adaptive to different structures and can thus be expected to outperform the other two procedures across most situations.

## Central Prediction Systems for

## Predicting Specific Course Grades

Through the ACT Assessment Program, postsecondary institutions can predict their freshmen students' grades in specific courses. Typically, institutions use specific course grade predictions for placement in courses requiring varying levels of academic development. For example, students with low predicted chances of success in a standard freshman English course might be advised or required to enroll in a remedial English course. On the other hand, students with a high predicted probability of success in an accelerated course might be encouraged to enroll in it.

Institutions usually make placement decisions on the basis of explicit selection on test scores; for example, students with ACT English test scores less than 16 might be placed in a remedial English course. Through the ACT Standard Research Service (SRS), institutions can make placement decisions using all four ACT test scores and student's self-reported high school grades. Such placement decisions, based on more information, are potentially more accurate than decisions based on single test scores.

The placement rules derived through participation in SRS are based on prediction equations for specific course grades using student's ACT test scores in English, mathematics, social studies, and natural sciences, and the student's self-reported high school grades in these same areas. Associated with each prediction equation are an estimated intercept, estimated regression slopes, and an estimated residual variance. The multiple correlation coefficient and residual variance are the standard measures used to assess the accuracy of the course grade predictions.

For deciding on whether to place a student in a standard level or remedial course, a predicted grade in the standard level course is computed.

The predicted grade is then converted into an estimated probability of C or better (or of B or better). The student is placed in the standard level course depending on whether his or her estimated probability of success exceeds a predetermined threshold.

A more sophisticated approach to placement would recognize the costs of incorrectly placing students in the remedial course (false negative error), as well as those of incorrectly placing them in the standard course (false positive error). If false positive and false negative errors are associated with equal loss, and if correct decisions carry no loss, then the optimal decision rule would be to admit students to the standard level course when their estimated probability of success is greater than .5. Other loss functions, of course, would lead to different decision rules. In many applications, the cost of false positive and false negative errors can not be expected to be the same, and a more systematic, decision theoretic model is required to establish decision rules for placement.

In another refinement of the procedure, course grades could be dichotomized, based on the definition of success, and the probability of success modeled directly. Such procedures require nonlinear (e.g. logistic) regression models, and are the focus of other research being done at ACT.

In evaluating the validity of decision rules for course placement, nontraditional validation strategies are required. The standard measures of association used in establishing criterion-related validity, the multiple correlation coefficient and the residual variance, measure the strength of the relationship between predictor and criterion variables, averaged across the range of the predictor score scales. With the assumption of multivariate normality, these statistics are useful in establishing the validity of decision rules for specific course placement (Sawyer, 1988). However, without an assumption of multivariate normality, these indices are largely uninformative. In validating placement decision rules, it is perhaps necessary, but not sufficient to describe the relationship between the predictor variables and the criterion variable by these traditional indices.

SRS grade predictions are based on within-college least squares (WCLS) procedures, which use data from each institution separately, in isolation from the data of other institutions. Potential problems encountered in using within-group regression equations include the presence of negative regression weights and a lack of adequate prediction accuracy on crossvalidation. As sample sizes become smaller, these problems increase in severity. Other factors that could lead to these problems are the low reliability of specific course grades and extreme collinearity among the predictor variables.

The present research investigated alternative methods for predicting the freshman course grades of students from their ACT scores and high school grades. In contrast to within-college least squares procedures, the alternative methods, called central prediction systems, use information from several institutions collaterally to derive a prediction equation for each individual institution.

There has been research over the past 20 years (e.g., Novick, Jackson, Thayer, & Cole, 1972; Rubin, 1980; Braun, Jones, Rubin, & Thayer, 1983; Houston, 1987) on both the mathematical and empirically observed properties of central prediction systems. A typical finding in these studies has been that using collateral information from several institutions can increase both the prediction accuracy and the stability of estimated regression weights over time. Thus, central prediction systems have the potential to be useful in situations where small sample sizes would preclude calculating within-group least squares prediction equations.

Situations where ACT Assessment users frequently express a need for prediction equations, but where the sample sizes needed for within-group least squares predictions are not available, include predicting specific course grades, predicting overall GPAs of students enrolled in different academic colleges of a university, and predicting overall GPAs of students with special background characteristics. The ability to predict specific course grades from small data sets could greatly increase the number of institutions that could make full use of AAP data for placement. For this reason, and because of the need to establish priorities, the current research was limited to the prediction of specific course grades. We hope to extend the research in the future to address the other two applications.

The two most common types of central prediction systems are Bayesian m-group regression and pooled least squares. Bayesian m-group regression uses the observed variability in least squares regression coefficients and residual variances across institutions to obtain refined estimates of the regression parameters for each individual institution. The refined parameter estimates are, roughly speaking, weighted averages of the within-group estimates and the estimates obtained from averaging the within-group estimates across all institutions. In the pooled least squares approach, the regression surfaces within each institution are assumed parallel, but not coincident. Under this assumption, estimates of the common slopes are pooled across institutions; intercepts are not assumed to be constant and are estimated separately for each institution. The assumptions of the pooled least squares approach are identical to the assumptions in traditional analysis of covariance models. Thus, the pooled least squares approach is hereafter referred to as the ANCOVA approach.

Bayesian m-group regression models allow estimates of regression slopes and variances to vary across institutions. They use collateral information only to the extent appropriate; when colleges differ greatly in their regression structures, or as sample sizes become large, Bayesian parameter estimates converge to the within-group least squares estimates. 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 similarities of the within-group regression structures across colleges, it is more flexible than the ANCOVA approach.

The ANCOVA model, in contrast, assumes that regression slopes and residual variances are homogeneous across institutions. To the degree that institutions differ in their regression slopes, the ANCOVA approach introduces prediction bias. To the degree that institutions differ in their residual variances, this approach introduces bias into the estimated probabilities associated with grade expectancies. On the other hand, the ANCOVA model is simpler to implement and operate than the Bayesian m-group model.

The WCLS model, the Bayesian m-group regression model, and the ANCOVA model may be compared along a continuum. If all of the colleges were entirely different in their regression structures, then WCLS would likely be more appropriate than ANCOVA. If all the colleges were very similar in their regression structures except for intercepts, then the ANCOVA model would be more appropriate. The Bayesian m-group regression model strikes a compromise between these two positions, and may be heuristically thought of as encompassing the other two models. Bayesian m-group regression has the effect of regressing the within-group parameter estimates toward common values. Unlike the ANCOVA approach, however, the extent of the regression effect is determined by the data rather than only by assumption.

In many previous empirical studies of Bayesian m-group regression, the colleges investigated were specially selected to be very similar in the demographic characteristics of their students and in their curricula. The feasibility and cost of identifying highly similar colleges would appear to diminish the usefulness of this methodology in routine or large-scale operations. The current research was designed to investigate applications involving small numbers of students enrolled in colleges with potentially diverse characteristics.

The following research questions are addressed in this study:

1. Do central prediction systems permit calculating specific course prediction equations from samples smaller than those currently required? If so, how much smaller?
2. What is the preferred method of central prediction with respect to prediction accuracy, practical feasibility, and defensibility of the assumptions required?

Both questions were investigated empirically, using real data from diverse groups of colleges.

## Method

### Data Source

Data for this investigation were obtained from colleges that participated in ACT's Standard Research Service (SRS) during the 1983-84 academic year and during at least one of the academic years 1984-85, 1985-86, or 1986-87. Alternative methods were investigated for predicting grades in the following three specific college courses: writing/grammar, algebra, and biology. Courses were identified from information collected by Noble and Sawyer (1987).

Colleges were selected so that the number of observations at a given college for a particular course was less than 100 in the 1983-84 base year and greater than or equal to 20 in at least one of the crossvalidation years (1984-85, 1985-86, or 1986-87). Including colleges with both the required and less than required base year sample size of 50, relative to currently published SRS guidelines, facilitated making a more precise evaluation of the potential benefits that may be realized from using collateral information. In order to obtain a sufficient number of colleges with less than the required sample size, all analyses were conducted for males and females separately. Colleges were selected only on the basis of available sample sizes and on our ability to identify the specific course.

The number of colleges available within each course group and level of sex, the total number of observations and the ranges of observations within colleges for both the base year and crossvalidation years, are presented in Table 1. The term "analysis group" is hereafter used to indicate a particular course group and level of sex. The colleges within each analysis group were considered to be exchangeable for the Bayesian portion of the analyses. Briefly, the assumption of exchangeability implies that one's subjective

Table 1. Distribution of Institutional Sample Sizes, by Course and Sex

| | | | Base year number of students | | | | Crossvalidation year number of students | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Course | Sex | Number of institutions | Min. | Med. | Max. | Total, all institutions | Min. | Med. | Max. | Total, all institutions |
| Writing/grammar | Male | 17 | 14 | 52 | 88 | 938 | 32 | 75 | 226 | 1278 |
| Writing/grammar | Female | 16 | 15 | 60 | 99 | 915 | 30 | 98 | 346 | 1582 |
| Algebra | Male | 13 | 20 | 55 | 90 | 682 | 20 | 73 | 184 | 948 |
| Algebra | Female | 11 | 33 | 60 | 98 | 696 | 38 | 85 | 193 | 940 |
| Biology | Male | 15 | 13 | 50 | 97 | 764 | 23 | 68 | 184 | 1016 |
| Biology | Female | 16 | 10 | 41 | 99 | 812 | 33 | 78 | 268 | 1250 |

judgements about the within-college regression parameters are the same for all colleges. See Lindley (1971) for a discussion of this concept.

**Procedure**

Two different prediction equations were studied: an eight-variable equation based on the four ACT subtests and the four high school grades, and a two-variable equation based on the ACT Composite and high school grade point average. For each prediction equation, regression coefficients and residual variances were estimated using three different models: within-college least squares (WCLS), pooled least squares with adjusted intercepts (ANCOVA), and empirical Bayesian m-group regression.

The specific Bayesian m-group regression model used in this study (Wang, 1988), is an extension of an empirical Bayesian model developed by Rubin (1980) and Braun, Jones, Rubin, and Thayer (1983). In all of these models, data are used to estimate hyperparameters of a common prior distribution on the within-college regression coefficients (intercept and slopes). The model in this study (denoted BAYES) differs from the previous empirical m-group regression models in its treatment of the within-group error variances. In the BAYES model, data-based estimates are obtained for the degrees of freedom and scale parameter of the inverse chi-square prior distribution for the exchangeable within-group error variances. Point estimates of the within-group regression parameters are taken to be the modes of the posterior distributions. The model developed by Rubin (1980) and Braun, et al. (1983) uses joint maximum liklihood estimates for the within-group error variances. The BAYES model used in this study has an empirically determined informative prior distribution on the within-group error variances and, therefore, regresses estimates of within-group error variances toward common values. No comparable regression of within-group error variances occurs in the model used by Rubin (1980) and Braun, et al. (1983). More accurate estimation of residual variances is important because it permits more accurate estimation of the probability of course success. The WCLS, ANCOVA, and BAYES models are described in greater detail by Houston (1987).

Both theoretical considerations and the empirically derived results to date suggest that using collateral information across groups effectively increases within-group sample sizes. However, shifts over time in the population and/or changes in grading standards tend to decrease prediction accuracy, regardless of the method and sample sizes used to derive base year prediction equations. Therefore, all prediction equations were cross-validated.

Data from the 1983-84 base year were used to calculate prediction weights for each of the combinations of number of predictor variables and three estimation models. The prediction equations for each institution were then used to predict the specific course grades of students at the same institution in the 1984-85, 1985-86, or 1986-87 crossvalidation year. Where adequate sample sizes were available in more than one crossvalidation year, data from the latest year were used. Indices of predictive accuracy utilized in the crossvalidation analyses include mean squared error (MSE), mean absolute error (MAE), zero-order correlation between predicted and obtained course grades (R), and prediction bias (BIAS). Colleges were grouped according to base year sample sizes ($<50$ and $\geq 50$), and the distributions of crossvalidation indices were summarized across institutions.

## Results

For the eight-variable prediction equation, the WCLS model produced a large number of negative weights. The percentages of negative regression slopes across colleges and variables and within analysis groups ranged from 29% (biology-female) to 38% (writing/grammar-male; algebra-male).

For three of the six analysis groups, the eight-variable ANCOVA model eliminated all of the negative regression slopes. For the remaining three analysis groups, pooled estimates of the regression slopes associated with one or more of the predictor variables were negative. A negative weight was associated with the ACT Natural Sciences subtest in all three of these groups.

The BAYES eight-variable prediction model failed to eliminate one or more of the negative regression slopes in all six analysis groups. In those analysis groups in which the pooled estimates obtained under the ANCOVA model were all positive, however, the BAYES model eliminated a substantial portion of the negative weights obtained with the WCLS model.

For the two-variable prediction equation, negative regression slopes obtained under the WCLS model were present in 4 of the 6 analysis groups. For these 4 analysis groups, the percentage of negative slopes across colleges and variables ranged from 3% (biology-male) to 15% (grammar/writing-male). For all 6 analysis groups, both the BAYES and ANCOVA two-variable models eliminated all of the negative regression weights.

Table 2 reports the means and standard deviations, across colleges within analysis groups, of the estimates of within-college error variances for two and eight predictor variables and for the three estimation models. For the eight-variable equation in the writing/grammar-males analysis group, the arithmetic mean across colleges of the maximum likelihood estimates of the

Table 2. Means (Standard Deviations) of Base Year Residual Variances Across Colleges[a]

| Course | Sex | Number of predictor variables | Prediction model: WCLS[b] | ANCOVA[c] | BAYES[d] |
|---|---|---|---|---|---|
| Writing/grammar | Male | 8 | .59(.26) | .73(.00) | .68(.14) |
| | | 2 | .74(.28) | .75(.00) | .74(.14) |
| | Female | 8 | .51(.17) | .69(.00) | .63(.10) |
| | | 2 | .66(.17) | .71(.00) | .67(.08) |
| Algebra | Male | 8 | .82(.30) | 1.12(.00) | 1.02(.08) |
| | | 2 | 1.12(.25) | 1.22(.00) | 1.21(.00) |
| | Female | 8 | .92(.31) | 1.11(.00) | .99(.15) |
| | | 2 | 1.10(.33) | 1.14(.00) | 1.11(.10) |
| Biology | Male | 8 | .73(.20) | .87(.00) | .84(.00) |
| | | 2 | .88(.21) | .88(.00) | .88(.00) |
| | Female | 8 | .52(.29) | .79(.00) | .73(.07) |
| | | 2 | .67(.20) | .80(.00) | .78(.04) |

[a]Quantities in parentheses are standard deviations across colleges.

[b]Within-college estimates are maximum likelihood estimates.

[c]Maximum likelihood estimates under the assumption that error variances are homogeneous across colleges within analysis groups.

[d]Within-college estimates are the modes of the marginal posterior distributions ($\chi^{-2}$).

within-college error variances obtained from the WCLS model is .59. The standard deviation, across colleges, of these estimates is .26. The ANCOVA model assumes that residual variances are homogeneous across colleges; under the assumption of homogeneity, the maximum likelihood estimate of the common within-college residual variance is .73. The BAYES model estimates the within-college error variances as the modes of the marginal posterior distributions on the error variances. The average of these Bayesian point estimates, across colleges, is .68. Their standard deviation, across colleges, is .14.

Note in Table 2 that for each analysis group, the BAYES model has substantially regressed the estimates of within-college error variances toward common values. The extent of the regression effect is reflected in the reduction in standard deviations between the WCLS and BAYES models. In three cases (algebra-male-2 variable; biology-male-8 variable; and biology-male-2 variable), the BAYES model regressed the estimates of within-college error variances virtually to a constant. For other analysis groups, the regression effect was more moderate.

The results of the crossvalidation analyses for each analysis group and prediction equation are reported in Appendix A. The tables there give the medians, across colleges within analysis groups, of the following cross-validation statistics: mean squared error (MSE), mean absolute error (MAE), zero-order correlation between predicted and observed course grades (R), and prediction bias (BIAS). Results are provided separately for the two-variable and eight-variable prediction equations, for the three different estimation models (WCLS, ANCOVA, and BAYES), and for colleges with different base year sample sizes (less than 50 and 50 or more). Results are also given for all colleges in each analysis group.

With the eight-variable WCLS prediction equations currently used in the ACT Standard Research Service, prediction accuracy is severely reduced when base year sample sizes are less than 50 (Sawyer, 1987). A striking example of this property of WCLS predictions is shown in Table A.1, where the median MSE across the 8 colleges with base year sample sizes less than 50 is 1.36, as compared to a median MSE of .69 for colleges with base year sample sizes greater than or equal to 50.

Under the WCLS model, the two-variable prediction equations were more accurate on crossvalidation than the eight-variable prediction equations for every analysis group. Pooled across base year sample sizes and averaged across analysis groups, the median MSE obtained for the two-variable prediction equations was 14% less than the median MSE obtained for the eight-variable equations. Under the ANCOVA and BAYES central prediction models, only very small differences between the two-variable and the eight-variable prediction equations in prediction accuracy on crossvalidation were found. For both central prediction models, the median MSE obtained for the two-variable prediction equations was about 1% less than that obtained for the eight-variable equations.

The comparisons of greatest interest in this study concern the prediction accuracy associated with the central prediction models at colleges with fewer than 50 observations in the base year. Table 3 reports differences between the median MSEs for prediction models at colleges with fewer than 50 base year observations and the corresponding median WCLS MSEs at colleges with 50 or more base year observations. These differences are given for each course and

Table 3. Difference Between Median MSE for Prediction Models at Institutions With Fewer Than 50 Base Year Observations and Median MSE for WCLS Model at Institutions With 50 or More Base Year Observations, Averaged Across Sexes

| Course | Number of predictor variables | Prediction model WCLS | ANCOVA | BAYES |
|---|---|---|---|---|
| Writing/grammar | 8 | .58 | .08 | .02 |
| | 2 | .25 | .11 | .06 |
| Algebra | 8 | .86 | .15 | .13 |
| | 2 | .37 | .24 | .14 |
| Biology | 8 | .36 | -.16 | -.18 |
| | 2 | -.01 | -.09 | -.09 |

number of predictor variables, averaged across sexes. Negative differences in Table 3 reflect smaller MSE (greater prediction accuracy) for the prediction models relative to the WCLS model at colleges with large base year sample sizes; positive quantities reflect larger MSE (lesser prediction accuracy). For example, the median MSE for males among the 8 colleges with base year sample sizes less than 50 was .68 under the BAYES model (Table A.1). The median MSE among the 9 colleges with base year sample sizes of 50 or more was .69 under the WCLS model. Thus, the relevant calculation is .68-.69 = -.01. From Table A.3, a similar calculation for females is .71 - .66 = .05. The average of these two differences is .02, as reported in Table 3.

The results in Table 3 indicate that for writing/grammar and algebra, the use of the central prediction models at colleges with fewer than 50 base year observations resulted in only modestly decreased prediction accuracy, as compared to the prediction accuracy associated with the WCLS model at colleges with 50 or more base year observations. For biology, the central prediction models at institutions with small base year sample sizes actually had greater prediction accuracy than did the WCLS model at institutions with large base year sample sizes.

It is also useful to state these results as proportionate changes in MSE, relative to the median MSE associated with the WCLS model at colleges with 50 or more base year observations. For writing/grammar, the increases in median MSE associated with the BAYES model at institutions with small base year sample sizes were 3% (eight-variable prediction equation) and 9% (two-variable prediction equation). For the ANCOVA model, the corresponding increases in MSE were 11% (eight-variable prediction equation) and 17% (two-variable prediction equation). For algebra, the increases in MSE associated with the

BAYES model were 10% (eight-variable) and 11% (two-variable) and the increases associated with the ANCOVA model were 12% (eight-variable) and 19% (two-variable). Using the central prediction models at colleges with fewer than 50 base year observations actually decreased median MSE for biology; for the BAYES model, the decreases were 16% (eight-variable) and 8% (two-variable) and for the ANCOVA model, they were 15% (eight-variable) and 8% (two-variable). These same trends were found for mean absolute error (MAE) as well.

Averaged across analysis groups and prediction equations, using the BAYES model at institutions with fewer than 50 base year observations decreased the correlation between predicted and obtained course grades by 8% relative to the median correlation using the WCLS model at institutions with 50 or more base year observations. The corresponding decrease in crossvalidated correlation using the ANCOVA model was 12%.

We sought to determine minimum within-course sample sizes that would be feasible using the BAYES model in terms of comparable prediction accuracy relative to that obtained using the WCLS model with current base year sample size requirements. To do this, we plotted MSE for the eight-variable prediction equation against base year sample size for the WCLS and BAYES models. (These plots, for each analysis group, are presented in Appendix B.) We then fit a curve to the scatterplot for each prediction model using least squares criteria.

For all 6 analysis groups, logarithmic regression curves adequately summarized the relationship between crossvalidated MSE and base year sample size for the WCLS model. For 3 analysis groups, logarithmic regression curves were also adequate in summarizing the relationship between MSE and base year sample size for the BAYES model. However, for the analysis groups grammar/

writing-male, biology-male, and biology-female, the relationship between MSE and base year sample size among the data was described more effectively by a straight line.

The plots in Appendix B indicate that the BAYES model permits calculating specific course prediction equations from smaller sample sizes than does the WCLS model. For example, in Figure B.1 the ordinate of the WCLS curve corresponding to a sample size of 50 is approximately 0.9. The Bayes line lies below this level for all sample sizes. We are not claiming that the relationship between BAYES MSE and base year sample size is linear everywhere; clearly, the BAYES line will at some point start to increase as base year sample size decreases. The curves in Figure 1 do suggest, however, that BAYES MSE will be less than WCLS MSE at small sample sizes.

Notwithstanding considerations of prediction accuracy, the WCLS model requires that the number of within-group observations be greater than the number of predictor variables; otherwise, the parameter estimates are not uniquely determined. (This requirement corresponds to positive degrees of freedom for error.) As noted by Braun, Jones, Rubin, and Thayer (1983), empirical Bayesian models can, in principle, be used even when the within-group data are of deficient rank. Although the plots for 2 courses (writing/grammar and biology) suggest the possibility that adequate prediction accuracy might be obtained for colleges with data of deficient rank, we do not recommend permitting sample sizes to be this low. Figures B.3 and B.4, corresponding to the analysis groups algebra-male and algebra-female, suggest that a base year sample size requirement of 20 would result in a reasonable level of accuracy, relative to that obtained using the WCLS model with the current sample size requirement. A similar sample size requirement was found for the ANCOVA model, as well.

## Conclusion

The effectiveness of central prediction models in eliminating the negative slopes frequently obtained under the WCLS model depends upon the number of predictor variables. The results for the eight-variable prediction equations suggest that neither central prediction model investigated will eliminate all of the negative regression slopes when there are large numbers of correlated predictor variables, though the ANCOVA model seems to be somewhat more effective than the BAYES model in this regard. For the two-variable prediction equations, however, both central prediction models eliminated all of the negative regression slopes in every analysis group.

Under the WCLS model, the two-variable prediction equations were found to be more accurate on crossvalidation than the eight-variable equations. For the central prediction models, on the other hand, the prediction accuracy obtained for the two-variable and eight-variable equations were essentially the same.

The results of this research are consistent with previous findings that central prediction models permit calculating prediction equations from fewer observations than are required with standard least squares methods. Only a slight reduction in prediction accuracy was found, relative to that obtained using the WCLS model with the current sample size requirement of 50. Averaged across analysis groups and number of predictor variables, using the BAYES model at institutions with less than 50 base year observations resulted in a 1.5% increase in MSE relative to the MSE obtained using the WCLS model at colleges with 50 or more base year observations. The corresponding increase in MSE for the ANCOVA model was 6%. Our analysis indicates that under either the BAYES or ANCOVA models, a base year sample of 20 observations would result in a level of accuracy comparable to that obtained using current procedures.

The BAYES and ANCOVA central prediction models achieved comparable levels of prediction accuracy, with the BAYES model slightly outperforming the ANCOVA model. This result is consistent with results for predicting freshman GPA, as reported by Houston (1987).

We have demonstrated that both central prediction models are practically feasible. All of the prediction models investigated (WCLS, BAYES, and ANCOVA) require calculating a sum of squares and cross products matrix of the predictor and criterion variables within each group. Given that the elements of these matrices have been calculated, the additional cost of using the central prediction models is small.

Moreover, the ANCOVA model makes the assumptions that the within-group regression surfaces are parallel and that the residual variances about each regression surface are homogeneous across groups. To the extent that colleges are carefully selected for inclusion into the central prediction system, the assumptions required by the ANCOVA model may be defensible. However, in large scale operations, a careful matching of colleges based on similarities in curricula and demographic characteristics of their students is not feasible.

Bayesian m-group regression models assume that the colleges are exchangeable, i.e., subjective a priori judgments about the within-college regression parameters are the same for all colleges in the system. The defining characteristics that colleges must possess in order to be considered exchangeable should, of course, be modified and extended, as warranted by experience.

Because the Bayesian approach is highly adaptive to different regression structures, practically feasible and theoretically defensible, and slightly more accurate on crossvalidation than the pooled least squares approach, we believe the m-group regression model is preferable.

## References

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.

Houston, W. M. (1987). Using collateral information from similar institutions to predict college freshman grade average. (ACT Research Report No. 87-8). Iowa City, Iowa: The American College Testing Program.

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

Noble, J. P., & Sawyer, R. L. (1987). The validity of ACT test scores and self-reported high school grades for predicting grades in specific college freshman courses. (ACT Research Report No. 87-20). Iowa City, Iowa: The American College Testing Program.

Novick, M. R., Jackson, P. H., Thayer, D. T., & Cole, N. S. (1972). Estimating multiple regressions in m groups: A crossvalidation 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. (1988). Validating the use of ACT scores and high school grades for remedial course placement in college. Unpublished manuscript, The American College Testing Program, Iowa City, Iowa.

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

Wang, M. M. (1988). Personal Communication.

# Appendix A

## Crossvalidation Statistics for the WCLS, ANCOVA, and BAYES Prediction Models, by Course Group, Sex, and Number of Predictors

Table A.1. Medians, Across Institutions, of Crossvalidation Statistics
Course: Writing/Grammar
Sex: Male
*Number of predictors: 8*

| Prediction model | Crossvalidation statistic | Base year sample size: Less than 50[a] | Base year sample size: 50 or more[b] | All institutions |
|---|---|---|---|---|
| WCLS | MSE | 1.36 | .69 | 1.03 |
| | MAE | .94 | .66 | .81 |
| | R | .21 | .37 | .34 |
| | BIAS | - .11 | - .13 | - .12 |
| ANCOVA | MSE | .75 | .69 | .72 |
| | MAE | .68 | .65 | .67 |
| | R | .44 | .47 | .46 |
| | BIAS | - .08 | - .16 | - .10 |
| BAYES | MSE | .68 | .66 | .66 |
| | MAE | .65 | .63 | .64 |
| | R | .46 | .47 | .47 |
| | BIAS | - .08 | - .13 | - .09 |

[a] 8 institutions

[b] 9 institutions

Table A.2. Medians, Across Institutions, of Crossvalidation Statistics
Course: Writing/Grammar
Sex: Male
Number of predictors: 2

| Prediction model | Crossvalidation statistic | Base year sample size: Less than 50[a] | Base year sample size: 50 or more[b] | All institutions |
|---|---|---|---|---|
| WCLS | MSE | .90 | .70 | .86 |
| | MAE | .76 | .64 | .73 |
| | R | .27 | .43 | .40 |
| | BIAS | -.18 | -.20 | -.20 |
| ANCOVA | MSE | .78 | .70 | .77 |
| | MAE | .70 | .64 | .69 |
| | R | .44 | .45 | .45 |
| | BIAS | -.10 | -.18 | -.10 |
| BAYES | MSE | .72 | .68 | .68 |
| | MAE | .66 | .62 | .63 |
| | R | .44 | .45 | .45 |
| | BIAS | -.14 | -.15 | -.14 |

[a] 8 institutions

[b] 9 institutions

Table A.3. Medians, Across Institutions, of Crossvalidation Statistics
Course: Writing/Grammar
Sex: Female
Number of predictors: 8

| Prediction model | Crossvalidation statistic | Base year sample size: Less than 50[a] | Base year sample size: 50 or more[b] | All institutions |
|---|---|---|---|---|
| WCLS | MSE | 1.15 | .66 | .81 |
| | MAE | .87 | .65 | .70 |
| | R | .24 | .50 | .43 |
| | BIAS | .26 | -.00 | .01 |
| ANCOVA | MSE | .75 | .61 | .66 |
| | MAE | .66 | .62 | .64 |
| | R | .29 | .59 | .53 |
| | BIAS | .01 | .06 | .03 |
| BAYES | MSE | .71 | .61 | .66 |
| | MAE | .65 | .62 | .63 |
| | R | .31 | .58 | .54 |
| | BIAS | - .02 | .01 | .00 |

[a] 5 institutions

[b] 11 institutions

Table A.4. Medians, Across Institutions, of Crossvalidation Statistics
Course: Writing/Grammar
Sex: Female
Number of predictors: 2

| Prediction model | Crossvalidation statistic | Base year sample size: Less than 50[a] | Base year sample size: 50 or more[b] | All institutions |
|---|---|---|---|---|
| WCLS | MSE | .87 | .58 | .64 |
| | MAE | .73 | .62 | .64 |
| | R | .31 | .58 | .48 |
| | BIAS | -.06 | .06 | .02 |
| ANCOVA | MSE | .71 | .58 | .66 |
| | MAE | .67 | .62 | .65 |
| | R | .30 | .58 | .50 |
| | BIAS | .01 | .03 | .02 |
| BAYES | MSE | .67 | .57 | .64 |
| | MAE | .66 | .60 | .63 |
| | R | .33 | .59 | .53 |
| | BIAS | .01 | .04 | .03 |

[a] 5 institutions

[b] 11 institutions

Table A.5. Medians, Across Institutions, of Crossvalidation Statistics
Course: Algebra
Sex: Male
Number of predictors: 8

| Prediction model | Crossvalidation statistic | Base year sample size: Less than 50[a] | Base year sample size: 50 or more[b] | All institutions |
|---|---|---|---|---|
| WCLS | MSE | 2.34 | 1.37 | 1.65 |
| | MAE | 1.16 | .96 | 1.08 |
| | R | .21 | .47 | .33 |
| | BIAS | - .03 | .07 | .06 |
| ANCOVA | MSE | 1.42 | 1.24 | 1.37 |
| | MAE | .96 | .93 | .94 |
| | R | .44 | .50 | .48 |
| | BIAS | .11 | - .00 | - .00 |
| BAYES | MSE | 1.41 | 1.21 | 1.34 |
| | MAE | .94 | .90 | .91 |
| | R | .44 | .53 | .47 |
| | BIAS | .05 | - .05 | - .04 |

[a] 6 institutions

[b] 7 institutions

Table A.6. Medians, Across Institutions, of Crossvalidation Statistics
Course: Algebra
Sex: Male
Number of predictors: 2

| Prediction model | Crossvalidation statistic | Base year sample size: Less than 50[a] | Base year sample size: 50 or more[b] | All institutions |
|---|---|---|---|---|
| WCLS | MSE | 1.72 | 1.31 | 1.40 |
| | MAE | 1.06 | .92 | 1.00 |
| | R | .35 | .48 | .44 |
| | BIAS | - .05 | - .07 | - .07 |
| ANCOVA | MSE | 1.51 | 1.25 | 1.34 |
| | MAE | 1.02 | .93 | .97 |
| | R | .43 | .47 | .44 |
| | BIAS | .02 | - .05 | .01 |
| BAYES | MSE | 1.38 | 1.20 | 1.31 |
| | MAE | .98 | .90 | .95 |
| | R | .45 | .48 | .46 |
| | BIAS | .01 | - .12 | .00 |

[a] 6 institutions

[b] 7 institutions

Table A.7. Medians, Across Institutions, of Crossvalidation Statistics
Course: Algebra
Sex: Female
Number of predictors: 8

| Prediction model | Crossvalidation statistic | Base year sample size: Less than 50[a] | 50 or more[b] | All institutions |
|---|---|---|---|---|
| WCLS | MSE | 1.93 | 1.19 | 1.33 |
| | MAE | 1.09 | .87 | .91 |
| | R | .32 | .51 | .35 |
| | BIAS | - .45 | .03 | - .02 |
| ANCOVA | MSE | 1.44 | 1.17 | 1.18 |
| | MAE | 1.01 | .83 | .88 |
| | R | .37 | .53 | .50 |
| | BIAS | - .32 | .09 | .01 |
| BAYES | MSE | 1.40 | 1.13 | 1.16 |
| | MAE | .99 | .84 | .88 |
| | R | .39 | .56 | .52 |
| | BIAS | - .27 | .09 | - .02 |

[a] 5 institutions

[b] 6 institutions

Table A.8. Medians, Across Institutions, of Crossvalidation Statistics
Course: Algebra
Sex: Female
Number of predictors: 2

| Prediction model | Crossvalidation statistic | Base year sample size: Less than 50[a] | Base year sample size: 50 or more[b] | All institutions |
|---|---|---|---|---|
| WCLS | MSE | 1.52 | 1.19 | 1.21 |
| | MAE | 1.04 | .88 | .89 |
| | R | .38 | .53 | .52 |
| | BIAS | - .34 | .08 | .01 |
| ANCOVA | MSE | 1.46 | 1.18 | 1.21 |
| | MAE | 1.03 | .86 | .91 |
| | R | .36 | .57 | .53 |
| | BIAS | - .33 | .09 | .04 |
| BAYES | MSE | 1.39 | 1.17 | 1.18 |
| | MAE | .98 | .84 | .87 |
| | R | .43 | .57 | .53 |
| | BIAS | - .21 | .09 | - .07 |

[a] 5 institutions

[b] 6 institutions

Table A.9. Medians, Across Institutions, of Crossvalidation Statistics
Course: Biology
Sex: Male
Number of predictors: 8

| Prediction model | Crossvalidation statistic | Base year sample size: Less than 50[a] | Base year sample size: 50 or more[b] | All institutions |
|---|---|---|---|---|
| WCLS | MSE | 1.62 | 1.17 | 1.18 |
| | MAE | 1.03 | .83 | .88 |
| | R | .38 | .50 | .45 |
| | BIAS | .17 | .20 | .20 |
| ANCOVA | MSE | .93 | 1.12 | 1.12 |
| | MAE | .76 | .82 | .81 |
| | R | .44 | .51 | .50 |
| | BIAS | .14 | .12 | .12 |
| BAYES | MSE | .90 | 1.03 | 1.02 |
| | MAE | .72 | .80 | .80 |
| | R | .44 | .54 | .54 |
| | BIAS | .02 | .20 | .11 |

[a] 6 institutions

[b] 9 institutions

Table A.10. Medians, Across Institutions, of Crossvalidation Statistics
Course: Biology
Sex: Male
Number of predictors: 2

| Prediction model | Crossvalidation statistic | Base year sample size: Less than 50[a] | Base year sample size: 50 or more[b] | All institutions |
|---|---|---|---|---|
| WCLS | MSE | 1.00 | 1.07 | 1.07 |
| | MAE | .81 | .83 | .83 |
| | R | .42 | .53 | .49 |
| | BIAS | .06 | .11 | .11 |
| ANCOVA | MSE | .92 | 1.02 | .99 |
| | MAE | .78 | .81 | .81 |
| | R | .44 | .53 | .52 |
| | BIAS | .14 | .10 | .11 |
| BAYES | MSE | .93 | .99 | .99 |
| | MAE | .75 | .80 | .80 |
| | R | .44 | .54 | .53 |
| | BIAS | .02 | .18 | .07 |

[a] 6 institutions

[b] 9 institutions

Table A.11. Medians, Across Institutions, of Crossvalidation Statistics
Course: Biology
Sex: Female
Number of predictors: 8

| Prediction model | Crossvalidation statistic | Base year sample size: Less than 50[a] | Base year sample size: 50 or more[b] | All institutions |
|---|---|---|---|---|
| WCLS | MSE | 1.26 | 1.00 | 1.07 |
| | MAE | .96 | .79 | .83 |
| | R | .42 | .54 | .53 |
| | BIAS | .17 | - .00 | .07 |
| ANCOVA | MSE | .92 | .90 | .91 |
| | MAE | .83 | .76 | .80 |
| | R | .65 | .59 | .59 |
| | BIAS | .15 | .01 | .08 |
| BAYES | MSE | .91 | .90 | .89 |
| | MAE | .78 | .76 | .79 |
| | R | .68 | .61 | .62 |
| | BIAS | .05 | - .00 | .03 |

[a] 9 institutions

[b] 7 institutions

Table A.12. Medians, Across Institutions, of Crossvalidation Statistics
Course: Biology
Sex: Female
Number of predictors: 2

| Prediction model | Crossvalidation statistic | Base year sample size: Less than 50[a] | Base year sample size: 50 or more[b] | All institutions |
|---|---|---|---|---|
| WCLS | MSE | .96 | .92 | .93 |
| | MAE | .80 | .77 | .78 |
| | R | .59 | .61 | .60 |
| | BIAS | .14 | -.01 | .08 |
| ANCOVA | MSE | .90 | .90 | .90 |
| | MAE | .77 | .76 | .77 |
| | R | .68 | .61 | .61 |
| | BIAS | .16 | .01 | .08 |
| BAYES | MSE | .89 | .88 | .88 |
| | MAE | .79 | .74 | .78 |
| | R | .68 | .62 | .62 |
| | BIAS | .00 | -.01 | -.00 |

[a] 9 institutions

[b] 7 institutions

# Appendix B

## Plots of Crossvalidated Mean Squared Error Against Base Year Sample Size for the WCLS and BAYES Models, by Course Group and Sex (Eight-Variable Prediction Equation)

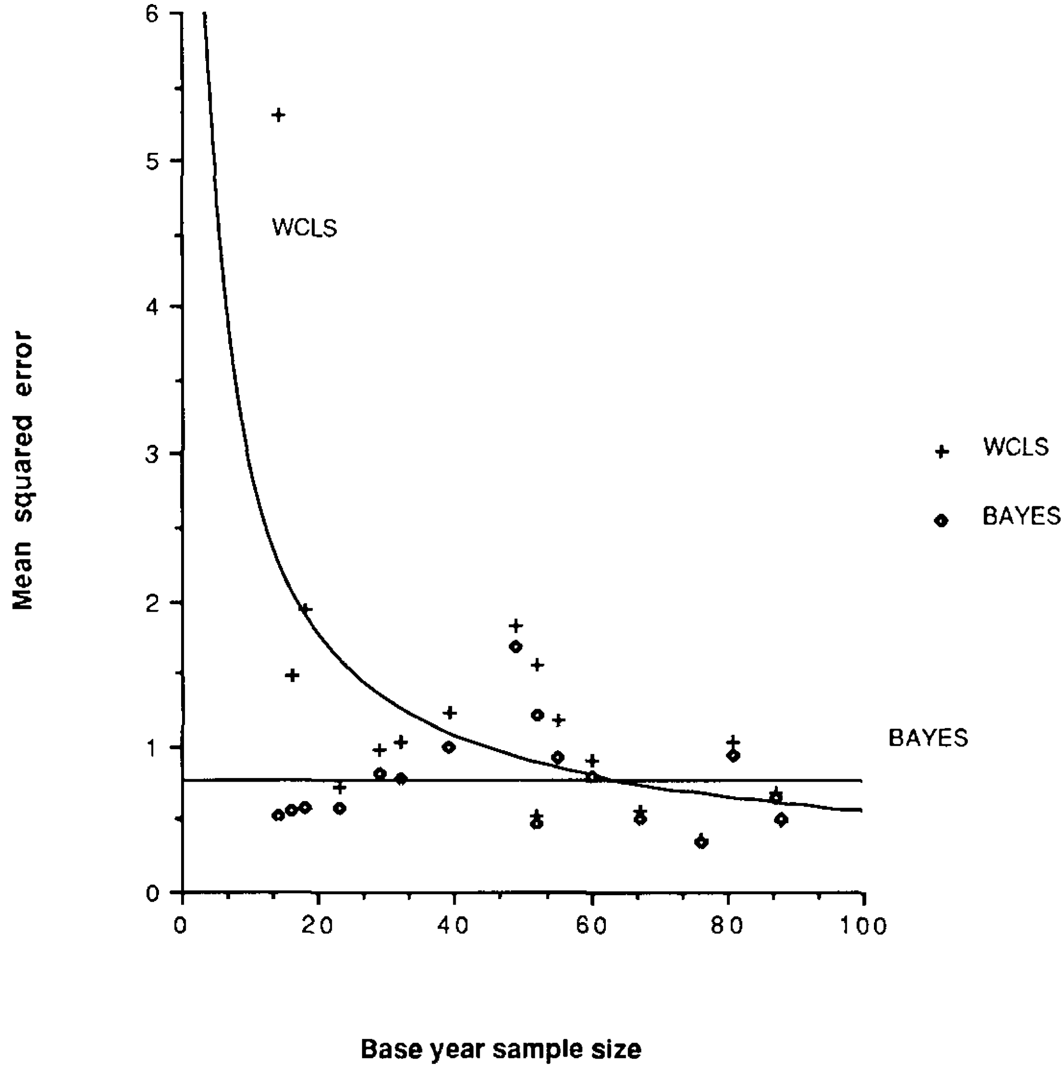


**Figure B.1 Crossvalidated MSE for WCLS and BAYES Models by Base Year Sample Size. (Analysis Group: Writing/Grammar-Male)**

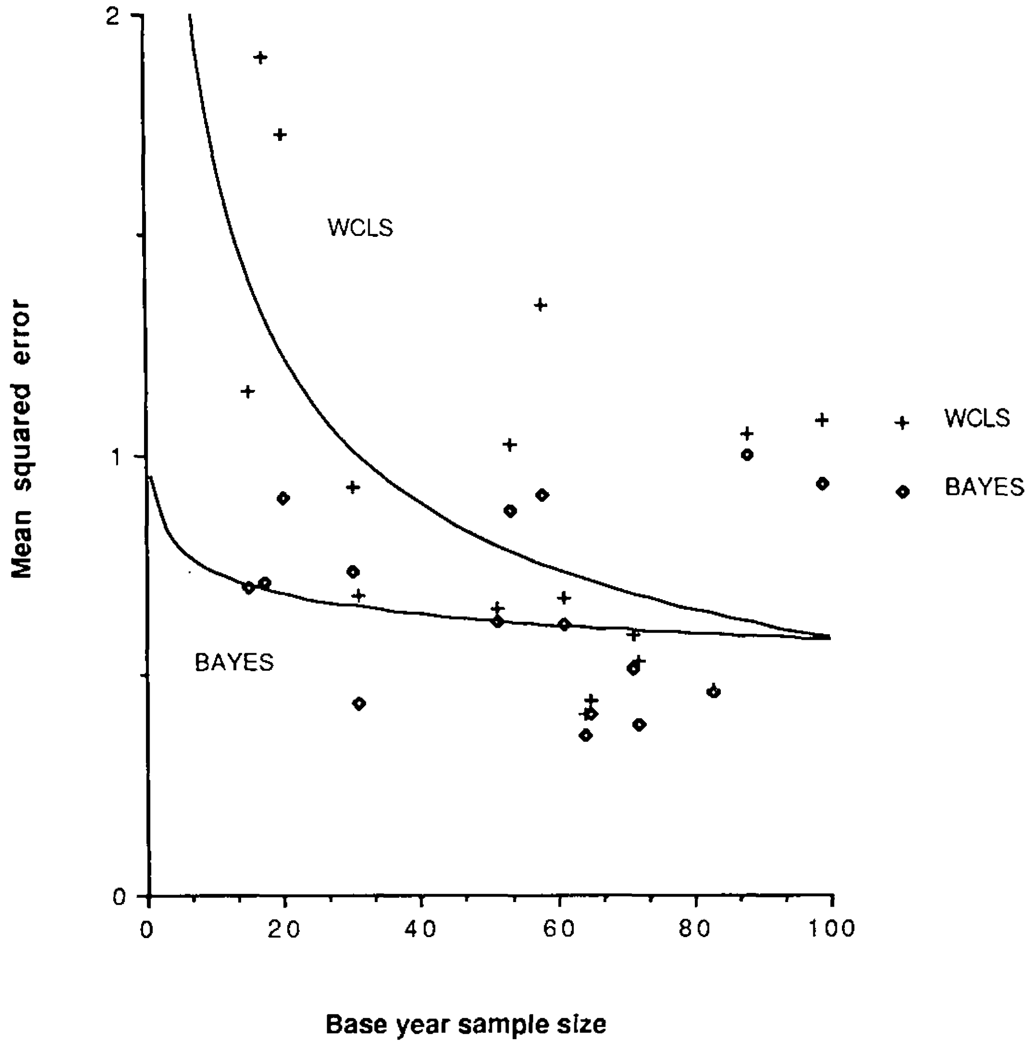


Figure B.2 Crossvalidated MSE for WCLS and BAYES Models by Base Year Sample Size.
(Analysis Group: Writing/Grammar-Female)

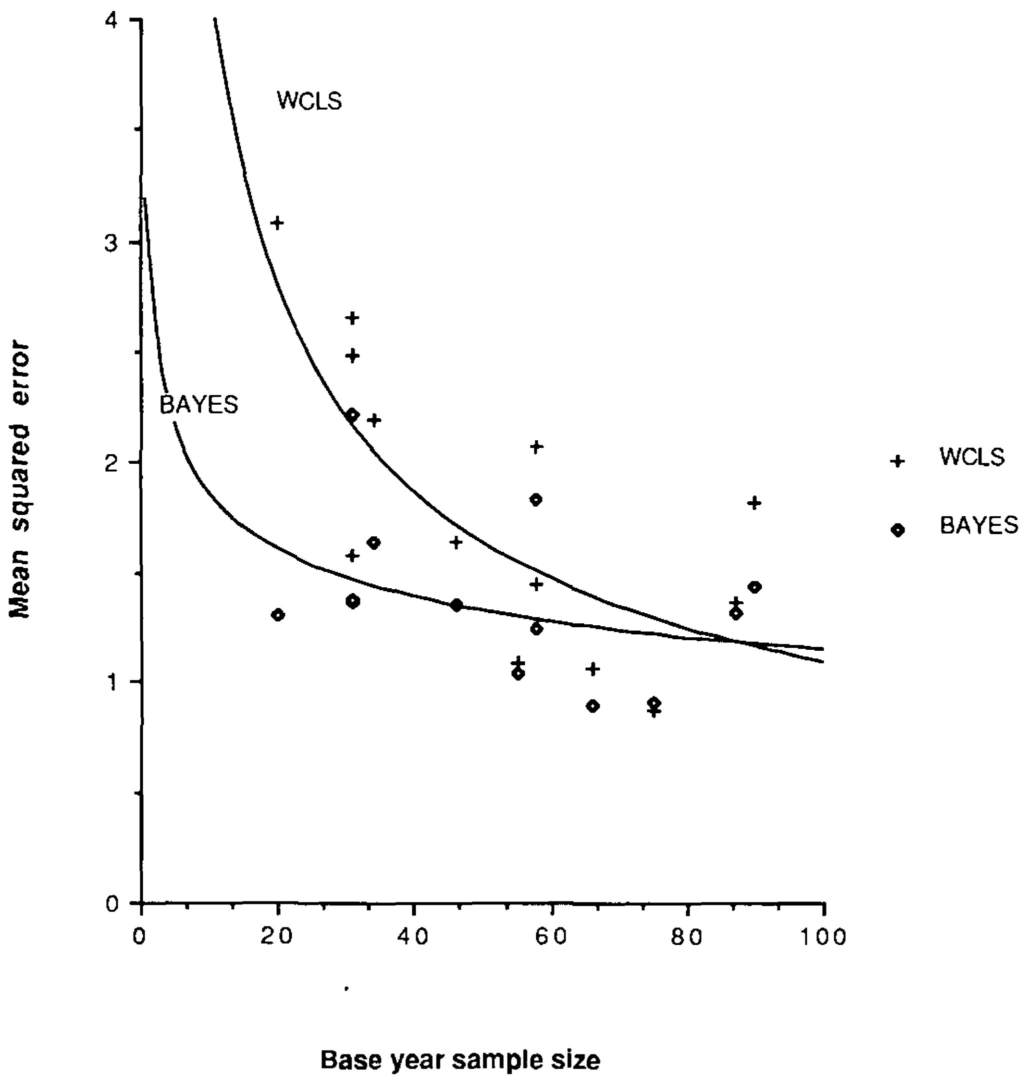


**Figure B.3 Crossvalidated MSE for WCLS and BAYES Models by Base Year Sample Size. (Analysis Group: Algebra-Male)**

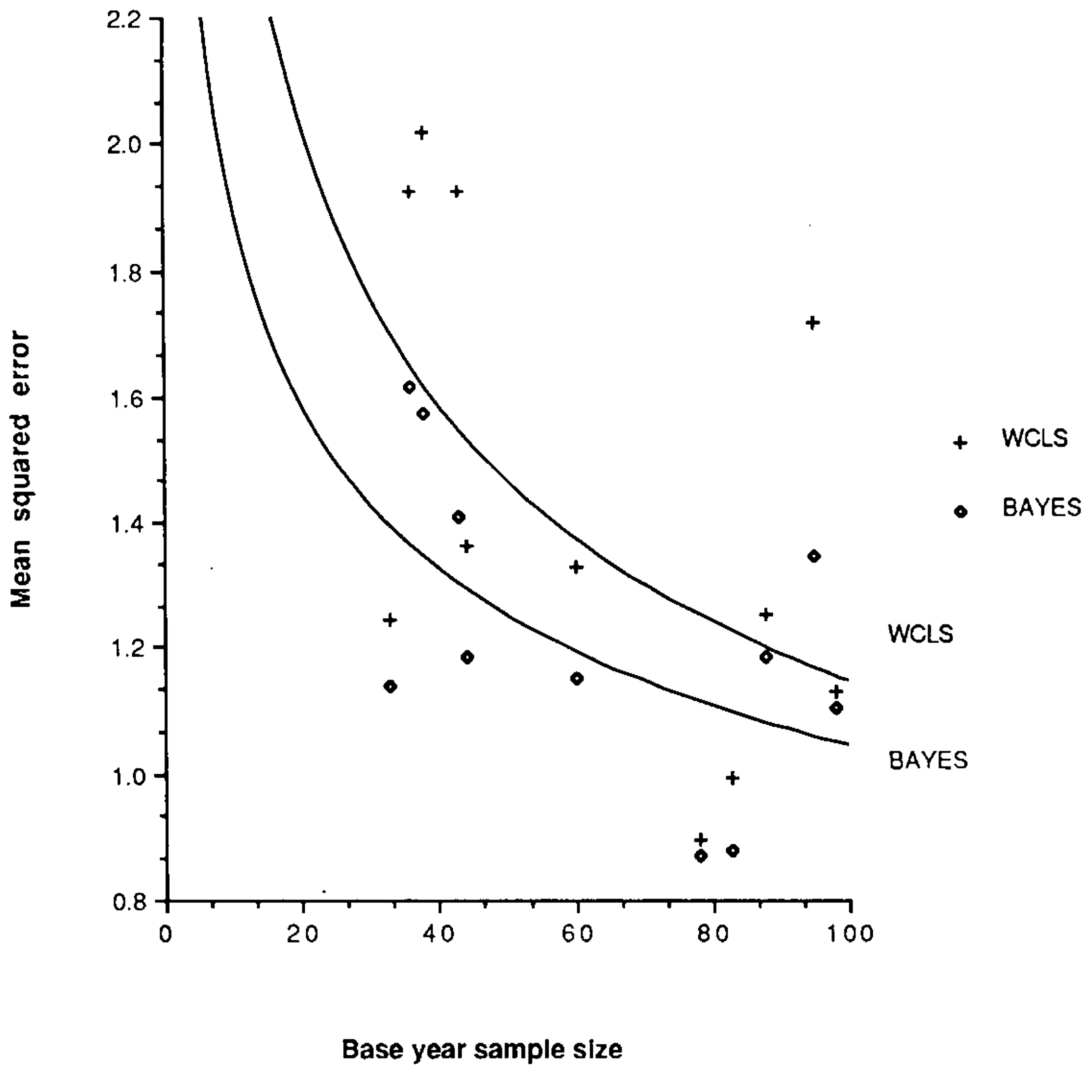


**Figure B.4 Crossvalidated MSE for WCLS and BAYES Models by Base Year Sample Size. (Analysis Group: Algebra-Female)**

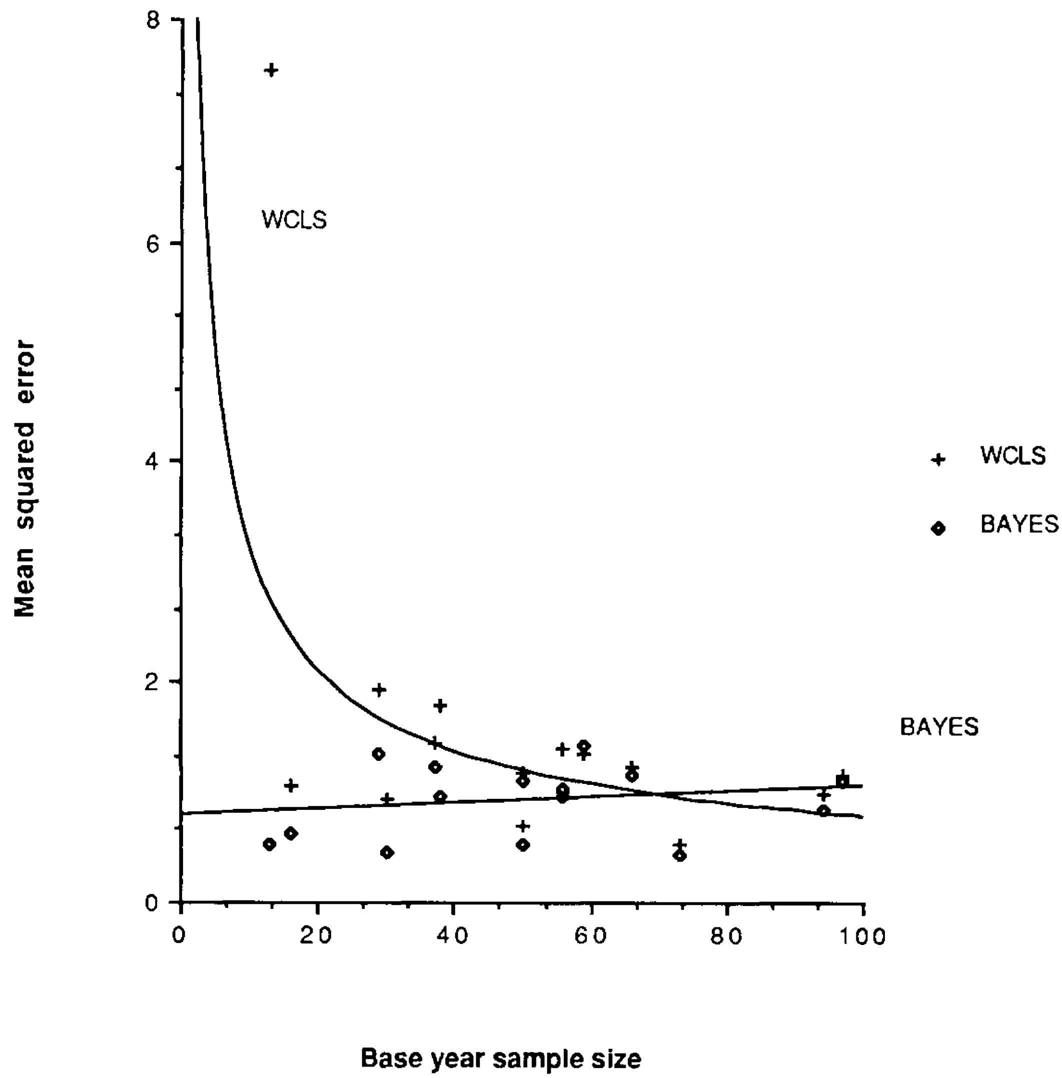


Figure B.5 Crossvalidated MSE for WCLS and BAYES Models by Base Year Sample Size. (Analysis Group: Biology-Male)

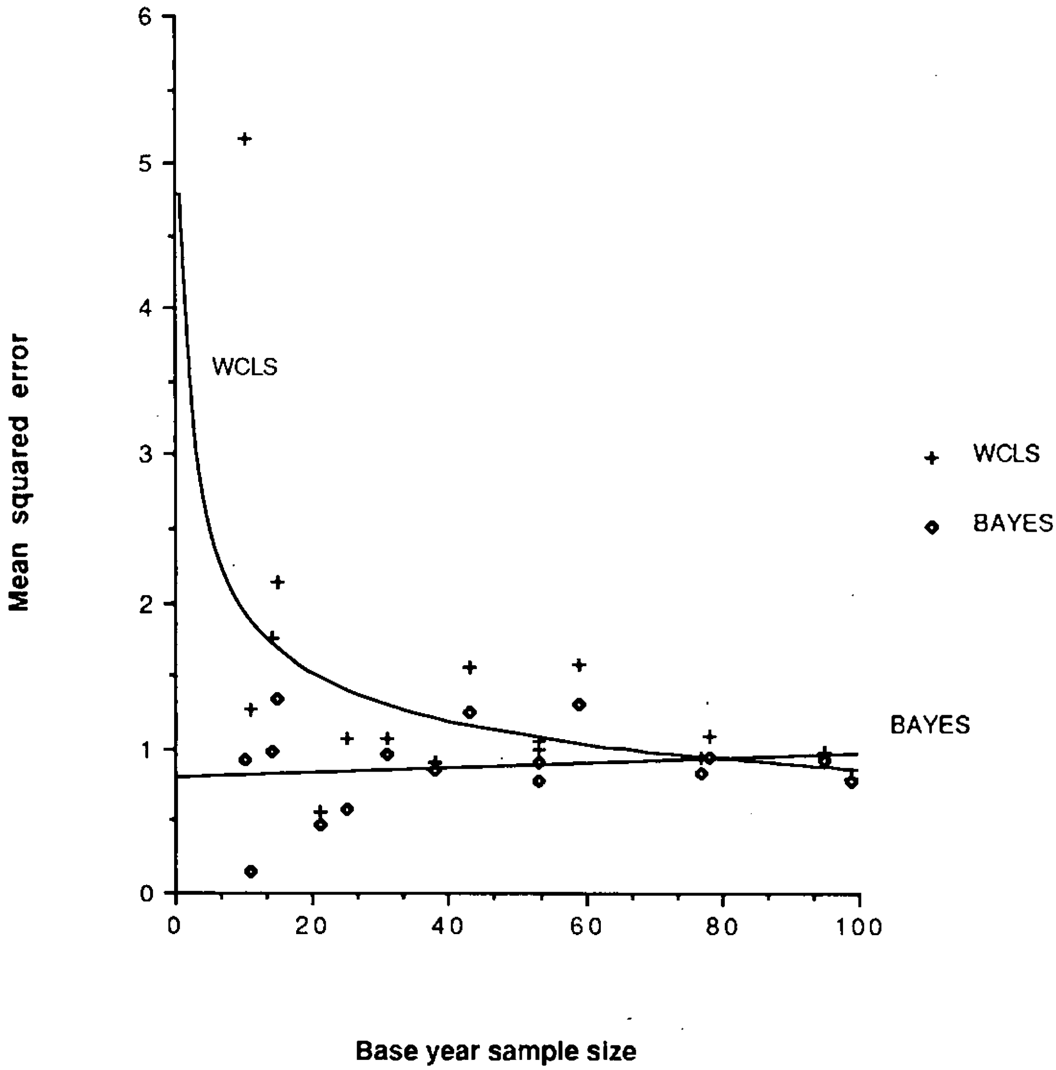


Figure B.6 Crossvalidated MSE for WCLS and BAYES Models by Base Year Sample Size. (Analysis Group: Biology-Female)