DOCUMENT RESUME

ED 392 815 TM 024 453

AUTHOR	Longford, Nicholas T.
TITLE	Negative Coefficients in the GRE Validity Study Service. GRE Research. GRE Board Professional Report No. 89-05P.
INSTITUTION	Educational Testing Service, Princeton, N.J.
SPONS AGENCY	Graduate Record Examinations Board, Princeton, N.J.
REPORT NO	ETS-RR-91-26
PUB DATE	Nov 91
NOTE	56p.
PUB TYPE	Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	*Bayesian Statistics; *College Entrance Examinations; *Estimation (Mathematics); *Goodness of Fit; Graduate Study; Higher Education; Models; Regression (Statistics); *Test Validity
IDENTIFIERS	*Graduate Record Examinations

ABSTRACT

Operational procedures for the Graduate Record Examinations Validity Study Service are reviewed, with emphasis on the problem of frequent occurrence of negative coefficients in the fitted within-department regressions obtained by the empirical Bayes method of H. I. Braun and D. Jones (1985). Several alterations of the operational procedures are proposed that would reduce the frequency of negative coefficients, and, if desired, completely eliminate them. It is argued, however, that there are no a priori reasons for assuming that all the coefficients are nonnegative. Reports of the fitted within-department regressions should be based on a single model, that would be found by model exploration. The estimation procedures could be improved by employing more flexible software for modeling between-department variation. Appendixes describe extended shrinkage in the empirical Bayes estimation and give common within-department variance. (Contains 2 tables, 5 figures, and 13 references.) (Author/SLD)

GRE®
RESEARCH

Negative Coefficients in the GRE Validity Study Service

Nicholas T. Longford

November 1991

GRE Board Professional Report No. 89-05P
ETS Research Report 91-26

ETS

Educational Testing Service, Princeton, New Jersey

Negative Coefficients in the GRE Validity Study Service

Nicholas T. Longford

GRE Board Report No. 89-05P

November 1991

This report presents the findings of a research project funded by and carried out under the auspices of the Graduate Record Examinations Board.

Educational Testing Service, Princeton, N.J. 08541

The Graduate Record Examinations Board and Educational Testing Service are dedicated to the principle of equal opportunity, and their programs, services, and employment policies are guided by that principle.

Abstract

Operational procedures for the Graduate Record Examinations Validity Study Service (GREVSS) are reviewed, with the emphasis on the problem of frequent occurrence of negative coefficients in the fitted within-department regressions obtained by the empirical Bayes method of Braun and Jones (1985). Several alterations of the operational procedures are proposed that would reduce the frequency of negative coefficients, and, if desired, completely eliminate them. It is argued, however, that there are no a priori reasons for assuming that *all* the coefficients are nonnegative. Reports of the fitted within-department regressions should be based on a single model, that would be found by model exploration. The estimation procedures could be improved by employing more flexible software for modelling between-department variation.

Acknowledgements

Computing and secretarial assistance were ably provided by Minhwei Wang and Liz Brophy, respectively. Discussions with Henry Braun, Paul Holland, Charles Lewis, and Lisa Schneider were very helpful. Critical comments of both formal and informal reviewers were instrumental in improving the manuscript.

Table of Contents

Appendices

Tables

Table of Contents (Continued)

Figures

1. Background

The GRE Validity Study Service (GRE VSS) provides participating graduate school departments with an array of infomation about the association of the first-year grade average (**FYA**), a measure of academic performance, with the GRE verbal (**V**), quantitative (**Q**), and analytical (**A**) scores, and the undergraduate grade-point average (**U**). The report to a department consists of two parts:

1. A regression formula with the estimates of the regression coefficients of **FYA** on **V**, **Q**, **A**, and **U**.

2. Expectancy table - estimated distributions of **FYA** based on the **predicted FYA (pFYA).**

The departments can use the estimated regression formula and the expectancy table to assess the relative importance of various admission measures and to predict the success of the applicants. Some departments may use these formulas to adjust their admission rules.

Under normal circumstances it is expected that the regression formula would have all four coefficients nonnegative, and that the distribution of the outcomes for each feasible value of **pFYA** would be unimodal. Examples to the contrary are:

Regression formula:[1]

$$\mathbf{pFYA} = 2.8 + .09\mathbf{V} + .14\mathbf{Q} - .02\mathbf{A} + .04\mathbf{U}$$

Expectancy table for **pFYA** = 3.0 (one row of the Table):

[1] In order to simplify the presentation, the scores **V**, **Q**, and **A** are defined on the scale 1 - 4, obtained by dividing the original GRE scores by 200.

pFYA	< 2.5	2.5 - 3.0	3.0 - 3.2	3.2 - 3.4	3.4 - 3.6	3.6 - 3.8	> 3.8
3.0	29	21	21	9	10	6	41

In this example the negative coefficient on **A** has no straightforward interpretation and can be explained only by a reference to the complex processes of selection and self-selection of students into departments and the idiosyncratic influences of the department/school environment on the students. The example of the expectancy table, containing a percentage breakdown of students with a specific value of **pFYA** into bands of **FYA** scores, appears to suggest that a student with predicted **FYA** of 3.0 is slightly more likely to have an eventual **FYA** in the range 3.4 - 3.6 than in the range 3.2 - 3.4, even though the latter range is closer to the actual prediction. This is clearly a contradictory outcome, indicating problems at some stages of the statistical analysis.

2. Purpose of the Study

The main purpose of the study reported here is to explore the sources/reasons for these aberrant features of the statistical analyses on which the GRE VSS reports are based and to devise alterations to the currently used procedures that would integrally produce nonnegative coefficients for all the departments.

The currently used procedures are based on a hierarchy of empirical Bayes models. For each data set, 16 empirical Bayes models are fitted. The software for model fitting employs the method of Braun and Jones (1985). In the fitted models each of the four regression coefficients is either constrained to zero for every department or department-specific coefficients are estimated. The estimates of the sets of five coefficients (the four variables and the intercept) may vary across the departments. The distribution of a coefficient across the departments is characterized by its mean and variance. Estimating department-specific regression formulas is the underpinning of the Validity Study Service, since its purpose is to describe department-specific characteristics.

For each department, estimated regression coefficients are reported from one of the 16 models, in which each regression coefficient is nonnegative. Such a procedure may involve bias due to selection of the reported formula. The size and importance of this model selection bias depend on the relative sizes of the estimated means and variances of the regression slopes across the departments. The analyses reported in Section 6 imply that in general the means of the slopes are small relative to their standard deviations, and so the associated bias is not ignorable. However, this problem is confounded with that of multicollinearity; the means of the slopes are relatively small or the standard deviations of the slopes relatively large, partly because of multicollinearity among the estimated parameters. The issue of multicollinearity is discussed in detail in Section 6. We propose a method that would rely on a single model fit for all the departments, and we describe simple procedures for selection of this model.

Sections 3 and 4 provide a summary of the empirical Bayes methods and of their relevance to the GRE VSS. In Section 5 an "extended" shrinkage method is described, and its application for the GRE VSS is discussed. The extended shrinkage can remove most of the negative coefficients, and, in judiciously selected models, all the negative coefficients. The empirical Bayes models have certain optimality properties, and so the coefficients estimated by the extended shrinkage are likely to have poorer statistical properties than those obtained by the original empirical Bayes method. Therefore, it is, important that extended shrinkage be used sparingly. In Section 6 multicollinearity is identified as one reason for frequent occurrence of negative estimated coefficients. We propose two approaches to combatting this problem: use of simpler models and enhancement of software to extend the model choice. The scope of possible improvements in statistical modelling of the GRE VSS data is summarized in Section 7. In Section 8 a simple method for calculation of the expectancy tables is described. It is almost identical to the currently used procedure, but it avoids repeated (numerical) multidimensional integration. Analogues of the ordinary regression R^2 "proportion of the variation explained" are given in Section 9. The report concludes with a discussion of admissibility of negative coefficients (Section 10) and a list of recommended changes (Section 11).

3. Empirical Bayes Models

The first prerequisite for an empirical Bayes analysis is the clustering of the observations. In the case of the GRE VSS we have students clustered within graduate school departments. Modelling of further clustering of the departments within schools, or department years within departments, has not been considered in the operation of the GRE VSS because the clusters at the higher levels contain very few units; for example, no department has provided data for more than three years, and no school has contributed to the data with more than six departments.

The largest available dataset contains records of the students who communicate best in English (dataset CE). It consists of 9,200 records of students from 606 departments, collected over the eight cycles of the study. The most recent cycle[2], cycle 18, has 2,230 students, and the previous cycles, 11 - 17, contain 6,970 students. The data from these cycles are pooled in order to make full use of between-department information. Records from the same department at different cycles are regarded as separate units, and in this report we refer to them as *different* departments. The departments have provided data for between 5 and 106 students. Throughout the report we refer to this dataset for illustration.

Most graduate departments have a very small number of students in any particular year, or even over several years, and so estimates of the regression coefficients based solely on the data from a department would have very large standard errors. For example, in the CE dataset there are two departments with more than 100 students: Department No. 1 has 102 records, and department No. 229 has 106 records in the dataset. The within-department ordinary regressions for these departments are given in Table 1. We see that only the variable U is significantly different from zero (at the 5% level), and that modelling of nonlinear regression is impractical because the standard errors become vastly inflated. Even

[2] At the time of writing of the report, September 1989.

for *linear* regression models any comparison of the regression coefficients across the two departments is meaningless because only unrealistically large differences would be statistically significant.

This provides a rationale for application of the empirical Bayes (EB) regression. The EB estimates of the department-specific regression coefficients are formed as a mixture of two estimates: (a) the estimates of the regression coefficients based solely on the data from the department and (b) the coefficients from the pooled regression using data from all the departments.

The within-department regression a, is unbiased but statistically inefficient in that the collateral information, contained in the data from the other departments, is not used. The pooled regression (b), is biased but has certain consistency properties. The optimal mixing weights are established by the EB procedure. A detailed exposition of the empirical Bayes models is given in Braun and Jones (1985) and Braun (1989), and here we provide only a minimal summary. Readers interested in further background are referred to the review paper by Morris (1983) and the references therein.

We assume the linear model[3]

$$FYA_{ij} = c_j + v_j^{*} V_{ij} + q_j^{*} Q_{ij} + a_j^{*} A_{ij} + u_j^{*} U_{ij} + \varepsilon_{ij}, \qquad (1)$$

where the lowercase letters c_j, v_j, q_j, a_j and u_j denote the coefficients for the department j (j = 1, 2, ..., J), and the uppercase letters denote the scores for the student i (i = 1, 2, ..., n_j) on the relevant variables. The random terms ε_{ij} represent the composite of the measurement error for FYA and model inadequacy (lack of fit). The linear model (1)

[3] Typographical note: Throughout the report the following notation is used; statistical parameters are denoted by lowercase characters, vectors of parameters by bold lowercase characters, and matrices of parameters by bold uppercase characters. Students' scores are denoted by capitals with double subscript ij denoting the student i in department j (e.g. V_{ij}). For department mean scores the "dot" notation is used (e.g., $V_{.j}$). Estimates of parameters and of conditional expectations are denoted by the ^ (e.g., $\hat{\mathbf{b}}$ denotes an estimate for b).

relates the individual scores to department-level coefficients, which are then further related to a set of population parameters:

$$
\begin{aligned}
c_j &= g_{cc} + g_{cv}\, V_{\cdot j} + g_{cq}\, Q_{\cdot j} + g_{ca}\, A_{\cdot j} + g_{cu}\, U_{\cdot j} + \delta_{c,j} \\
v_j &= g_{vc} + g_{vv}\, V_{\cdot j} + g_{vq}\, Q_{\cdot j} + g_{va}\, A_{\cdot j} + g_{vu}\, U_{\cdot j} + \delta_{v,j} \\
q_j &= g_{qc} + g_{qv}\, V_{\cdot j} + g_{qq}\, Q_{\cdot j} + g_{qa}\, A_{\cdot j} + g_{qu}\, U_{\cdot j} + \delta_{q,j} \qquad (2) \\
a_j &= g_{ac} + g_{av}\, V_{\cdot j} + g_{aq}\, Q_{\cdot j} + g_{aa}\, A_{\cdot j} + g_{au}\, U_{\cdot j} + \delta_{a,j} \\
u_j &= g_{uc} + g_{uv}\, V_{\cdot j} + g_{uq}\, Q_{\cdot j} + g_{ua}\, A_{\cdot j} + g_{uu}\, U_{\cdot j} + \delta_{u,j}
\end{aligned}
$$

where $V_{\cdot j}$, $Q_{\cdot j}$, $A_{\cdot j}$, $U_{\cdot j}$ are the department means for the explanatory variables V, Q, A, and U, respectively, g are the population parameters, and $\delta_{x,j}$ are (department-level) residual terms. The model (2) refers to a specific choice of departmental covariates; in principle any variable defined for the departments can be used as a covariate. It is advantageous to introduce more compact notation for (1) and (2), such as

$$
\begin{aligned}
FYA_{ij} &= \mathbf{X}_{ij}\mathbf{b}_j + \varepsilon_{ij} \\
\mathbf{b}_j &= \mathbf{X}_{.j}\mathbf{g} + \delta_j, \qquad (3)
\end{aligned}
$$

where $\mathbf{X}_{ij} = (1, V_{ij}, Q_{ij}, A_{ij}, U_{ij})$, $\mathbf{X}_{.j} = (1, c_j, v_j, q_j, a_j, u_j)^T$, $\mathbf{X}_{.j} = (1, V_{\cdot j}, Q_{\cdot j}, A_{\cdot j}, U_{\cdot j})$ and $\delta_j = (\delta_{c,j}, \delta_{v,j}, \delta_{q,j}, \delta_{a,j}, \delta_{u,j})^T$. Braun and Jones (1985) use a more compact notation:

$$
\begin{aligned}
\mathbf{FYA}_j &= \mathbf{X}_j\mathbf{b}_j + \boldsymbol{\varepsilon}_j \\
\mathbf{b} &= \mathbf{X}.\mathbf{g} + \delta \qquad (4)
\end{aligned}
$$

where $\mathbf{FYA}_j$, $\boldsymbol{\varepsilon}_j$ and $\mathbf{X}_j$ are the department vectors and matrix, respectively, corresponding to the outcome, the random terms, and the explanatory variables, and $\mathbf{b}$, $\boldsymbol{\delta}$, and $\mathbf{X}$ are the respective vectors and matrix of the department coefficients, the department-level random terms, and the within-department means. In principle, the department means in these models can be replaced, or augmented, by other covariates defined for the departments, although in the GRE VSS only the department means are used. We assume that the vectors of random terms $\boldsymbol{\varepsilon}$ and δ are mutually independent and that

$$\delta_j \sim \mathbf{N}_5(\mathbf{0}, \boldsymbol{\Sigma}) \quad \text{(i. i. d)}, \tag{5}$$

i.e. $\{\delta_j\}_{j=1,2,...,J}$ form a random sample from a five-variate normal distribution with zero means and a (nonnegative definite) variance matrix $\boldsymbol{\Sigma}$. The student-level random terms ε_{ij} are assumed to be themselves mutually independent and distributed according to $\mathbf{N}(0, \sigma_j^2)$. Alternatively, the model (1) - (2) can be described as a random coefficients model, in which the within-department regressions form a set of independent normally distributed random variables with a common structure for the mean, and common variance,

$$\mathbf{b}_j \sim \mathbf{N}_5(\mathbf{X}_{\cdot j}\mathbf{g}, \boldsymbol{\Sigma}). \tag{6}$$

Flexibility of model choice is achieved by deleting department- and/or student-level variables from the model (1) - (2). For example, for the CE dataset the following submodel of (1) - (2) is considered:

$$\mathbf{FYA}_{ij} = c_j + v_j^{*} V_{ij} + q_j^{*} Q_{ij} + a_j^{*} A_{ij} + u_j^{*} U_{ij} + \varepsilon_{ij},$$

$$\begin{aligned} c_j &= g_{cc} + g_{cv} V_{\cdot j} + \delta_{c,j} \\ v_j &= g_{vc} + g_{vv} V_{\cdot j} + \delta_{v,j} \\ q_j &= g_{qc} + g_{qv} V_{\cdot j} + \delta_{q,j} \\ a_j &= g_{ac} + g_{av} V_{\cdot j} + \delta_{a,j} \\ u_j &= g_{uc} + g_{uv} V_{\cdot j} + \delta_{u,j} \end{aligned} \tag{7}$$

Exclusion of a student-level variable - say A_{ij} - from the student-level model (1) corresponds to setting $a_j \equiv 0$ or, equivalently, $g_{ac} = g_{av} = 0$ and $\delta_{u,j} \equiv 0$. In the operation of the GRE VSS each of the coefficients for the four variables U, V, Q, and A, is either constrained to zero or is estimated. This gives rise to the 16 models that are routinely fitted for each dataset in the GRE VSS. Note that exclusion of a variable in (1) implies not only deletion of the associated row in (7), but also deletion of the corresponding row and column of Σ (or setting each element of this row and column to 0).

A computational algorithm for fitting the EB model is described in the technical appendix of Braun and Jones (1985).

4. Validity of the EB Model

The empirical Bayes model (1)-(2) provides an idealized description for the available data. Firstly, the assumption of normality of the random terms ε_{ij} is grossly violated because of the "lumpiness" of the data: The outcome score FYA is the average of a small number of (integer) grades, which are themselves highly correlated. As a result, more students tend to have FYA scores near the values 3, 3.5, and, 4 and in several departments only a very limited number of possible scores can be achieved. The scale of the FYA is too coarse for any assumptions of normality to be satisfied. Also, the scoring of FYA may reflect different standards of the institutions, or even of the departments.

Similarly, the predictor score U is not objectively scaled, and students in a graduate department usually come from a variety of undergraduate colleges. For the observed predictor scores V, Q, A, and U we have to consider the underlying latent traits as the appropriate explanatory variables, and in this perspective the observed scores represent the latent traits subject to measurement error. In the EB analysis this component of contamination is ignored; no practical methods for its incorporation are available.

The department-level variables are included in the model to represent the "context" of the department. In this respect the within-department means represent *proxies* for some department-level traits. The reliability of such proxies cannot be assessed since we do not have a definition of the underlying traits.

Search for additional predictors for the EB model is likely to be futile unless it is based on information about the underlying educational processes. Operationalizing these additional predictors would lead to a number of difficulties, including developing a rigorous definition of the predictor, devising reliable means of eliciting additional information from the departments without loss of cooperation, and so on.

A large proportion of the students have achieved the perfect score on the outcome FYA and/or on the predictor U; about 10% have achieved the perfect FYA score, 4, and 5% have a U score of at least 3.9. This may significantly diminish the validity of the underlying scales and is an additional threat to the assumptions of normality.

The main advantage of the EB models for the GRE VSS is in their compromise between parsimony and adequacy. We prefer to use models with as few parameters as possible, while insisting on having all the salient features of the data (and of the processes that generate them) explicitly represented in the models. In the absence of complete information about these processes the analyst would be inclined to represent in the model as much collateral information, in the form of explanatory variables, as possible. This improves the chances of generating an adequate model, at the cost of possible redundancy and loss of efficiency.

In standard statistical models several tools for arbitrating between model adequacy and statistical efficiency are available. In ordinary regression (ordinary least squares) the well-known t- and F-tests are often employed to find variables that make unimportant contributions toward description of variation of the outcomes. In the implementation of Braun and Jones (1985) the analogues of the t- and F-tests cannot be performed because standard errors for the estimated parameters are not available. The likelihood ratio test could be used for comparing the quality of fit for two models, one of which is a special case of the other.

The extreme case of model redundancy is multicollinearity, or linear dependence, of the predictors. For example, if the scores V, Q, and A were linearly dependent, one of these three variables could be excluded from the models without any loss of adequacy. Standard statistical packages implement various measures for collinearity such as the "Measure for Collinearity" in F4STAT, the square of the partial correlation of the variable with the outcome, given all the other predictor variables. Another simple indicator of collinearity is the condition number (see Section 6). In practical situations, the more explanatory variables are used, the greater the threat of collinearity. This is certainly the case in many educational research applications where many predictors are highly correlated with the general ability *g*.

Consider an alternative representation of the EB model (1) - (2)

$$\begin{aligned} FYA_{ij} = {} & g_{cc} && + g_{cv}V_{.j} && + g_{cq}Q_{.j} && + g_{ca}A_{.j} && + g_{cu}U_{.j} \\ & + g_{vc}V_{ij} && + g_{vv}V_{.j}V_{ij} && + g_{vq}Q_{.j}V_{ij} && + g_{va}A_{.j}V_{ij} && + g_{vu}U_{.j}V_{ij} \\ & + g_{qc}Q_{ij} && + g_{qv}V_{.j}Q_{ij} && + g_{qq}Q_{.j}Q_{ij} && + g_{qa}A_{.j}Q_{ij} && + g_{qu}U_{.j}Q_{ij} \\ & + g_{ac}A_{ij} && + g_{av}V_{.j}A_{ij} && + g_{aq}Q_{.j}A_{ij} && + g_{aa}A_{.j}A_{ij} && + g_{au}U_{.j}A_{ij} \\ & + g_{uc}U_{ij} && + g_{uv}V_{.j}U_{ij} && + g_{uq}Q_{.j}U_{ij} && + g_{ua}A_{.j}U_{ij} && + g_{uu}U_{.j}U_{ij} \\ & + \gamma_{ij}, \end{aligned} \qquad (8)$$

where $\gamma_{ij} = \delta_{c.j} + \delta_{v.j}V_{ij} + \delta_{q.j}Q_{ij} + \delta_{a.j}A_{ij} + \delta_{u.j}U_{ij} + \varepsilon_{ij}$ accumulates all the random terms. The threat of the collinearity in (8) is obvious. The regressor variables A_{ij}, $V_{.j}A_{ij}$, $Q_{.j}A_{ij}$, $A_{.j}A_{ij}$ and $U_{.j}A_{ij}$ [a row of (8)] are very closely related. The range of values of the department-means of the four scores V, Q, A, and U is very narrow, and many departments have students with very similar scores, thus strengthening the redundancy in the EB models that use several department-level means. The analysis reported in Section 6 indicates acute collinearity not only among the 25 predictors, but also within the set of 10 predictors that would be considered in the present operation of the GRE VSS as relatively simple models. It turns out that the cross-level interactions $V_{.j}V_{ij}$, $V_{.j}Q_{ij}$, $V_{.j}A_{ij}$ and $V_{.j}U_{ij}$, included to take account of the context of the department, are the main causes of collinearity. The model (7) contains five parameters for the students' scores ($g_{x,c}$) and five parameters for the context ($g_{x,v}$). Description of dependence of the outcome on the student background could be supplemented by quadratic terms if more adequacy was required, but the description for the context contains a lot of redundancy.

The variance matrix Σ that describes the department-level variation contains 15 parameters - 5 variances and 10 covariances. The parametrization for Σ may also contain some redundancy, but with the software used in the operation of the GRE VSS no constraints on Σ can be imposed. Of particular importance would be setting the variances to zero (common regression slope for all the departments) and certain covariances to zero, so as to enhance statistical efficiency. We note that even in data with a large number of clusters (departments) the data may not contain plentiful information about the multivariate nature of variation of the regression coefficients. Therefore, it is essential in model selection to have the option of constraining the covariance structure in Σ and to have suitable criteria for selection among the available options. Clear evidence of redundancy in the parametrization of Σ is that in most cases the estimate of this matrix is almost singular, and would probably be singular if perfect convergence were achieved (in order to control costs in the operation, the EB procedure is stopped after a preset number of iterations).

Another element of model redundancy is in allowing separate student-level variances σ_j^2 for each department. As an alternative, a common variance σ^2 should be considered. The need for differing student-level variances could be established only if the data contained a lot of large departments. At present, the parsimonious model with common σ^2 is preferable.

From the formulation (7) we can see the EB model is a special case of the random regression model in which each regression coefficient is either constant across all the clusters or varies from cluster to cluster according to the normal law. In (7) the regression coefficients in the first line are declared as varying from department to department. Software for these general models is available, using the EM algorithm[4] (Raudenbush and Bryk, 1986), the Fisher-scoring algorithm (Longford, 1987), or the iteratively reweighted least squares (Goldstein, 1986). The EM algorithm is generally very slow, especially with complex models (as many as 500 iterations may be required), while the other two algorithms require usually fewer than 15 iterations, and provide standard errors for all the estimated parameters. Convergence of the EM can be substantially speeded up by simple acceleration routines (such as reported in Lindstrom and Bates, 1988).

[4] EM stands for Expectation - Maximization; see Dempster, Laird, and Rubin (1977) for details.

5. Negative Coefficients and Extended Shrinkage

In the model formula (7), rewritten in the form

$$\begin{aligned} \mathrm{FYA}_{ij} = & \; g_{cc} + g_{cv}V_{\cdot j} + \delta_{c,j} \\ & + (g_{vc} + g_{vv}V_{\cdot j} + \delta_{v,j})V_{ij} \\ & + (g_{qc} + g_{qv}V_{\cdot j} + \delta_{q,j})Q_{ij} \\ & + (g_{ac} + g_{av}V_{\cdot j} + \delta_{a,j})A_{ij} \\ & + (g_{uc} + g_{uv}V_{\cdot j} + \delta_{u,j})U_{ij} + \varepsilon_{ij}, \end{aligned} \tag{9}$$

the parentheses in lines 2 - 5 contain the respective regression coefficients on the scores V, Q, A, and U. We can see from (9) how the regression coefficient, say,

$$g_{vc} + g_{vv}V_{\cdot j} + \delta_{v,j}$$

for V, depends on the department means $V_{\cdot j}$. A condition for these "true" coefficients to be nonnegative for all the departments is that their expectations $g_{vc} + g_{vv}V_{\cdot j}$ be nonnegative for all values of $V_{\cdot j}$ that occur in the data, and substantially larger than the standard deviation (the square root of the variance) of $\delta_{v,j}$:

$$g_{vc} + g_{vv}V_{\cdot j} > (\Sigma_{vv})^{1/2},$$

and analogously for the other variables,

$$\begin{aligned} g_{qc} + g_{qv}V_{\cdot j} > (\Sigma_{qq})^{1/2}, \\ g_{ac} + g_{av}V_{\cdot j} > (\Sigma_{aa})^{1/2}, \\ g_{uc} + g_{uv}V_{\cdot j} > (\Sigma_{uu})^{1/2}. \end{aligned} \tag{10}$$

In practice we consider the estimates of the parameters g_{xy} and of the variance matrix Σ in place of the parameters in (10), and the posterior expectations of the random terms $\delta_{x,y}$ in place of the random terms in (9), see (11). The condition (10) is less likely to be satisfied the more overparametrization there is for the variance matrix $\mathbf{\Sigma}$. In particular, when the estimate of $\mathbf{\Sigma}$ is singular, its diagonal elements (the estimated variances) tend to be inflated.

The conditional expectations for the within-department regression coefficients

$$b_j = (c_j, v_j, q_j, a_j, u_j)$$

are obtained by the formula

$$\begin{aligned} \hat{b}_j &= \mathrm{E}(b_j \mid \mathbf{FYA}, \mathbf{X}; \{\sigma_j^2\}_j, \mathbf{G}, \mathbf{\Sigma}) \\ &= (\mathbf{P}^* + \mathbf{P}_j)^{-1}(\mathbf{P}^*\mathbf{G}^T\mathbf{X}_{.j} + \mathbf{P}_j\mathbf{B}_j), \end{aligned} \tag{11}$$

where $\mathbf{P}^* = \mathbf{\Sigma}^{-1}$, $\mathbf{P}_j = \mathbf{X}_j\mathbf{X}_j^T/\sigma_j^2$, $\mathbf{B}_j = (\mathbf{X}_j\mathbf{X}_j^T)^{-1}\mathbf{X}_j\mathbf{y}_j^T$ is the within-department ordinary least squares solution, $\mathbf{X}_{.j}$ the vector of within-department means of the scores on $\mathbf{X}$, and $\mathbf{G}$ is the matrix of the parameters $g_{x,y}$. The unknown parameters in (11) are replaced by their maximum likelihood estimates. The vector of coefficients (11) is a mixture of two estimates, the pooled regression estimate $\mathbf{G}^T\mathbf{X}_{.j}$ and the within-department regression $\mathbf{B}_j$. An alternative interpretation is that the within-department regression estimates are being shrunk toward the overall regression, and the amount of shrinkage is determined so as to optimally combine the within-department and the pooled-data information. The mixing weight for the latter is given by the (estimated) within-department information, $\mathbf{P}_j$. Since the within-department regressions have a lot of sampling variability, a necessary condition for obtaining nonnegative coefficients for all the departments is that the "stable" component, the pooled regression estimate $\mathbf{G}^T\mathbf{X}_{.j}$, be positive for all the departments. Even if this condition is satisfied. a negative coefficient for a department can arise when the within-department regression has a negative coefficient so large in absolute value that it remains negative even after the shrinkage to the pooled regression.

The straightforward solution to this aberration is to extend the shrinkage until the coefficient is shrunk to zero. Such an extended shrinkage is guaranteed by the positive regression coefficients of the pooled-data regression. Of course, no theoretical justification for this procedure can be given, other than prior information that all the departments *have* nonnegative regression coefficients. The procedure can be partly justified on the grounds of poor resampling properties of the estimators of the variance matrix $\mathbf{\Sigma}$ (because of over-parametrization) and of the variances σ_j^2 (based on too few data-points).

The consequences of this extended shrinkage have to be carefully weighed. First, if sampling variation of the estimators for $\mathbf{\Sigma}$ and σ_j^2 is ignored, the extended shrinkage is less optimal, in terms of statistical efficiency, than the original shrinkage determined by the EB procedure. Therefore, this adjustment method should be used sparingly. Second, if we insist on nonnegative estimated coefficients, then for successful application of the extended shrinkage we require an EB solution for which $\mathbf{G}^T\mathbf{X}_{.j}$ has nonnegative components for *all* values of the departmental covariate(s) $\mathbf{X}_{.j}$. The matrix of parameters $\mathbf{G}$ can be consistently estimated by ordinary least squares, that is, by analyzing the entire dataset without department identification (treating all students as a single department). Thus, the burden of model selection is shifted to the pooled-data regression.

Model selection based on the pooled-data regression may turn out to be advantageous for the GRE VSS. If model selection is based on ordinary pooled-data regression, the use of the computationally intensive EB procedures can be postponed to the fitting of only a very small number of models. At the first stage, ordinary regression models would be fitted for the pooled data using the student-level scores, their department-level means, and the cross-level interactions. One or a small number of parsimonious models would be adopted for which all the departments have nonnegative pooled regression coefficients. The corresponding EB models would be fitted, with the addition of the extended shrinkage. The amount of extended shrinkage would be monitored to provide an additional criterion for selection among the EB model fits. Various diagnostic procedures for multicollinearity in ordinary regression can be directly applied, as discussed in Section 6.

6. Multicollinearity in EB Regression

For the analysis of the CE dataset in the operation of the GRE VSS the EB model (7) with the covariate $V_{.j}$ is used. The expectation of an outcome is

$$E(\mathrm{FYA}_{ij}) = (1\ \mathrm{V}_{ij}\ \mathrm{Q}_{ij}\ \mathrm{A}_{ij}\ \mathrm{U}_{ij})\mathbf{G}(1\ \mathrm{V}_{\cdot j})^{\mathrm{T}},$$

or in matrix notation

$$E(\mathrm{FYA}_{ij}) = \mathbf{X}\boldsymbol{\beta}, \tag{12}$$

where **X** is a matrix with N=9,200 rows (students) and 10 columns representing the regressor variables **1** (intercept), V_{ij}, Q_{ij}, A_{ij}, U_{ij}, $\mathrm{V}_{\cdot j}$, $\mathrm{V}_{ij}\mathrm{V}_{\cdot j}$, $\mathrm{Q}_{ij}\mathrm{V}_{\cdot j}$, $\mathrm{A}_{ij}\mathrm{V}_{\cdot j}$ and $\mathrm{U}_{ij}\mathrm{V}_{\cdot j}$. The parameter vector $\boldsymbol{\beta}$ is uniquely defined only if **X** is of full rank, r(**X**)=10. Otherwise, if **X** is singular, a column of the matrix **X** could be reconstructed as a linear combination of the other columns, say,

$$\mathbf{x}_{10} = \mathbf{X}_{-10}\mathbf{B}, \tag{13}$$

where $\mathbf{X}_{-10}$ is the N×9 matrix formed from **X** by deleting the 10th column, and **B** is a 9×9 matrix of full rank. Then the regression formula becomes

$$E(\mathrm{FYA}_{ij}) = \mathbf{X}_{-10}\boldsymbol{\beta}^{*}, \tag{14}$$

with $\boldsymbol{\beta}^{*} = \mathbf{B}\boldsymbol{\beta}$. Therefore, the 10th variable could be deleted from the model and the data description simplifie '. Some elements of $\boldsymbol{\beta}^{*}$ may be substantially different from the corresponding elements of $\boldsymbol{\beta}$, thus making these parameters difficult to interpret.

Often in regression problems the matrix **X** is of full rank, but is almost singular. Proximity of the matrix **X** to singularity is referred to as multicollinearity, or ill-conditioning.

The extent of multicollinearity of the design matrix **X** can be established by an eigenvalue analysis of the corresponding matrix of crossproducts, $\mathbf{X}^{\mathrm{T}}\mathbf{X}$. Let

$$\mathbf{X}^{\mathrm{T}}\mathbf{X} = \sum_{k=1}^{10} \lambda_k \mathbf{a}_k \mathbf{a}_k^{\mathrm{T}},$$

be the eigenvalue decomposition for $\mathbf{X}^T\mathbf{X}$, with the (positive) eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{10}$ in descending order. The ratio of the largest and smallest eigenvalues is referred to as the *condition number,* and it is a useful indicator of multicollinearity in ordinary regression. A large condition number implies almost linear dependence of a column of the design matrix **X**, such as (13). Let $\hat{\boldsymbol{\beta}}$ be the estimate of $\boldsymbol{\beta}$ obtained by EB analysis. Then the vector of fitted expected values, $\mathbf{X}\hat{\boldsymbol{\beta}}$, is very close to the vector of fitted expected values from the model with the abbreviated list of regressors (14), $\mathbf{X}_{-10}\hat{\boldsymbol{\beta}}^*$. Thus, we have two vectors of estimates for $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\beta}}$ and $(\hat{\boldsymbol{\beta}}^*,0)$, which lead to very similar fitted expected values, but the corresponding elements of these vectors may be substantially different, and may even have different signs. Multicollinearity in ordinary regression is associated with highly inflated standard errors for some of the estimates. Also, the estimates tend to be very unstable; a small change in the data, or in the model specification, may cause a profound change in the estimates. The EB estimates share these undesirable properties; the inflation of the standard errors could be demonstrated if the standard errors of the estimates were available, and instability can be observed by the substantial changes in the estimates that occur after even a modest change of the model specification.

When the estimated regression parameters are unstable, they are less likely to satisfy certain inequalities believed to hold for the parameters, such as

$$\begin{aligned}
&\dot{g}_{vc} + \dot{g}_{vv}V_{\cdot j} > 0\\
&\dot{g}_{qc} + \dot{g}_{qv}V_{\cdot j} > 0\\
&\dot{g}_{ac} + \dot{g}_{av}V_{\cdot j} > 0\\
&\dot{g}_{uc} + \dot{g}_{uv}V_{\cdot j} > 0, \qquad (15)
\end{aligned}$$

[compare with (10)] for the values of $V_{\cdot j}$ that occur in the dataset, or in the underlying population of departments. The left-hand sides of the inequalities in (15) are the estimated average slopes of the respective scores V, Q, A and U in a department with the mean V score $V_{\cdot j}$.

Suppose (f=) $\dot{g}_{qc} + \dot{g}_{qv}V^* < 0$ for a specified value of V^*. The fitted department-regression coefficient on Q for a department with $V_{\cdot j} = V^*$ (denoted by q_j)

differs from the mean regression coefficient f by the posterior mean of the deviation $\delta_{j,q}$. The mean of the posterior deviations $\delta_{j,q} = q_j - f$ for the departments with the department-mean V score V^* is very close to zero (unless the model is seriously misspecified), and therefore at least half the departments with department-mean V scores in the vicinity of V^* have negative fitted coefficients on Q. Nevertheless, owing to the multicollinearity of **X**, it is feasible that the fit for the data could be almost exactly reproduced by a completely different set of regression parameters $\hat{\beta}$. Although this would imply substantially different department-regression coefficients $\hat{\mathbf{b}}_j$, there might be only insubstantial changes in the fitted values for the students. In other words, the currently applied EB model may have very good crossvalidation properties for prediction of the outcome scores $\{y_{ij}\}$ (see Braun and Jones, 1985, Section 3.6), but not so as good properties for prediction of the regression coefficients.

The eigenvalues of the matrix of crossproducts for $(1\ V_{ij}\ Q_{ij}\ A_{ij}\ U_{ij}\ V_{\bullet j})\ V_{ij}V_{\bullet j}\ Q_{ij}V_{\bullet j}\ A_{ij}V_{\bullet j}\ U_{ij}V_{\bullet j}$, the regressors implied by the model (7), are

$$\begin{array}{ccccc} 2.71\times10^6, & 2.02\times10^4, & 1.28\times10^4, & 7.58\times10^3, & 2.93\times10^3, \\ 679, & 26.5, & 14.8, & 9.25, & 1.34, \end{array}$$

and so the condition number is about 2×10^6. If we simplify the structure of the model (12) by excluding the regressors $V_{\bullet j}$, $V_{ij}V_{\bullet j}$, $Q_{ij}V_{\bullet j}$, $A_{ij}V_{\bullet j}$ and $U_{ij}V_{\bullet j}$, that is, by deleting the covariate $V_{\bullet j}$, the corresponding matrix of crossproducts has the eigenvalues

$$3.23\times10^5, \quad 2.50\times10^3, \quad 1.49\times10^3, \quad 921, \quad 118,$$

with the condition number of about 2,750. Deletion of these five regressors, corresponding to deletion of the covariate $V_{\bullet j}$ from (7), may lead to a substantially poorer fit to the data. It would be more appropriate to delete one regressor at a time and assess the loss of adequacy of the resulting regression fit. Deletion of a variable - say A - corresponds to deletion of the regressors A_{ij} and $A_{ij}V_{\bullet j}$ (a row of the matrix **G**), and the eigenvalues of the corresponding matrix of crossproducts are

2.09×10^6, 1.78×10^4, 1.37×10^4, 2.38×10^3, 658, 23.6, 14.6, 1.36,

(condition number 1.55×10^6), and so multicollinearity remains present. Deletion of a variable (of a GRE test score, since U is the predictor with the largest estimated slope and smallest standard error) would also be undesirable on theoretical grounds; it is believed that the GRE analytical score A makes a contribution toward description of the performance in graduate school above and beyond the verbal and quantitative scoies. Sources of multicollinearity can be explored in more detail by considering submatrices of the functionally related regressors. For example, the eigenvaiues corresponding to the regressors 1, V_{ij}, $V_{\cdot j}$, $V_{ij}V_{\cdot j}$ are

6.51×10^5, 2.71×10^3, 497, 3.91

(condition number 1.66×10^5). This indicates that the regressor $V_{ij}V_{\cdot j}$ could be deleted with minimal loss to the quality of the fit. A similar pattern of the eigenvalues is observed for the other sets of regressors 1, Z_{ij}, $V_{\cdot j}$, $Z_{ij}V_{\cdot j}$, where Z_{ij} is either U, Q, or A (the corresponding condition numbers are 6.75×10^5, 1.65×10^5 and 1.88×10^5).

We consider two methods for combatting multicollinearity. The first method involves model simplification and a description of the causes of multicollinearity. The second method, ridge regression (Hoerl and Kennard, 1970), is a general principle based on adjustment of the matrix of crossproducts $\mathbf{X}^T\mathbf{X}$.

Model Simplification

Within the framework of the algorithm of Braun and Jones (1985) and the operational software based on it, two kinds of model simplification are possible: (1) deletion of a covariate (a column of **G**), and (2) deletion of a variable (a row of **G**). For the CE dataset the former leads to deletion of five regressors and possibly a substantially poorer fit for the data. By deleting a variable two regressors are removed, but, as the eigenvalue analysis indicates, this would not reduce the acute multicollinearity among the regressors.

Clearly, multicollinearity is caused by the cross-level interactions $V_{ij}V_{\cdot j}$, $Q_{ij}V_{\cdot j}$, $A_{ij}V_{\cdot j}$ $U_{ij}V_{\cdot j}$ some, or all of which should be deleted from the model, while retaining the department mean $V_{\cdot j}$, or if necessary even adding another mean (e.g., $U_{\cdot j}$) to the list of regressors. These models cannot be fitted by the operational software based on Braun and Jones (1985), although in other implementations of the EB models, such as Bryk, Raudenbush, Seltzer, and Congdon (1988), Rasbash, Prosser, and Goldstein (1988) and Longford (1988) they can be fitted routinely.

Ridge Regression

Ridge regression is a standard method for combatting multicollinearity in ordinary regression. If the matrix of crossproducts $\mathbf{X}^T\mathbf{X}$ has a small eigenvalue, then its inverse $(\mathbf{X}^T\mathbf{X})^{-1}$ and the ordinary least squares solution $(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$ are unstable. Stability of the solution can be enhanced by replacing the matrix of crossproducts by $\mathbf{X}^T\mathbf{X} + h\mathbf{I}$, where $\mathbf{I}$ is the unit matrix and $h>0$ a tuning constant. The choice for h should be such as to induce little bias (the smaller h the lesser the bias) and to promote stability (the higher h the higher the eigenvalues of $\mathbf{X}^T\mathbf{X} + h\mathbf{I}$, and the more stable the ridge regression solution $(\mathbf{X}^T\mathbf{X} + h\mathbf{I})^{-1}\mathbf{X}^T\mathbf{X})$. In the EB approach we may consider applying ridge regression for the within-department regressions as well as for estimation of the matrix $\mathbf{G}$. The within-department regressions involve ordinary least squares, and so the application of the ridge regression is straightforward as long as we have an intelligent method of choosing the constant h_j; each department may have a different ridge constant. The matrix $\mathbf{G}$ is estimated by the multivariate regression

$$\hat{\mathbf{G}} = (\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{R},$$

where $\mathbf{Z}$ is the matrix of covariates, with rows $(1\ V_{\cdot j})$, and $\mathbf{R}$ is the matrix of the estimated department-regression coefficients, consisting of rows $\hat{\mathbf{b}}_j$. An opportunistic choice for the ridge constant would be the smallest value of h for which the components of $(\mathbf{Z}^T\mathbf{Z} + h\mathbf{I})^{-1}\mathbf{Z}^T\mathbf{R}$ satisfy the inequalities (15). The choice of the ridge constant h affects the estimates of the regression parameters $\hat{\mathbf{G}}$, which in turn influence the estimate of the matrix of department-regression coefficients **R**. Thus, another layer of iterations of the EB procedure would have to be implemented, which would iteratively calculate the matrices $\hat{\mathbf{G}}$ and **R** and update all the variance and covariance parameters in the process. A "short-cut" solution would involve finding a suitable value of h for the fixed set of within-department regression coefficients **R** obtained after convergence of the operational EB algorithm.

7. Scope for Improvement of the Regression Model

The importance of the covariate $V_{\cdot j}$ can be explored by fitting the EB model in two stages. First we fit the EB model (1) with no covariates ("shrinking to a point" in the terminology of Braun and Jones, 1985) and obtain the within-department regression coefficients. In the second stage, these within-department regression coefficients are regressed on the covariate $V_{\cdot j}$. The results of this stage are displayed in the "Submodel" column of Table 2. For comparison, we fit the operational model, (7) (using $V_{\cdot j}$ as a covariate), and regress the resulting department-regression coefficients on $V_{\cdot j}$. Having accounted for the covariate $V_{\cdot j}$ in the model (7), the estimated slopes on $V_{\cdot j}$ should be equal to zero. But the estimated slopes, displayed under the column heading "Operational model" in Table 2, are of comparable size with the corresponding estimated slopes for the submodel. Some of the simple regressions on $V_{\cdot j}$ are even significant, using the traditional t-ratio test, say, at the 5% level of significance. We see that the intended role of the covariate in the operational model, to account for systematic variation due to $V_{\cdot j}$, has not been fulfilled. This is most likely due to the combination of acute multicollinearity and imperfect convergence of the employed algorithm.

Another area of possible model improvement is in nonlinear regression. In general we can consider a polynomial regression in the variables V, Q, A, and U. It turns out that a small number of quadratic terms significantly improve the fit of the model, and these additional variables contribute only marginally to multicollinearity. In the operational software these variables would have to be associated with between-department variation, but in other software the associated variances could be constrained to zero. We note that nonlinear regression would substantially complicate the discussion of negative coefficients, and would involve substantial changes in the presentation of the regression formulas in the GRE VSS reports.

Variances and Covariances in Σ

The variances in the matrix Σ can be interpreted as a measure of variation of the within-department regressions. A randomly selected department with the verbal score mean $V.$ has the slope on Q equal to $q = g_{qc} + g_{qv}V. + \delta_v$, where $\delta_v \sim N(0, \Sigma_{vv})$, or equivalently, $q \sim N(g_{qc} + g_{qv}V., \Sigma_{vv})$. Therefore, the probability that q is positive is equal to $\Phi\{(g_{qc} + g_{qv}V.)/(\Sigma_{vv})^{1/2}\}$, where Φ is the distribution function for the standard normal distribution $N(0,1)$. The estimates of the variances in Σ, in conjunction with the regression parameter estimates, indicate the frequency of negative department-regression coefficients.

Multicollinearity can also arise among the estimated variances and covariances. The procedures for detecting multicollinearity can be based on the estimated information matrix (and its eigenvalue decomposition) for the variances and covariances or the standard errors associated with these parameters. These are not available in the operational software, but are readily available in the software based on the iteratively reweighted least squares procedure (Goldstein, 1986) and the Fisher-scoring algorithm (Longford, 1987). Multicollinearity is present to some extent among the variance and covariance parameters, because information about the within-department slopes is very scarce, but the problem is not as acute as for the regression parameters. To alleviate multicollinearity, several covariances could be constrained to 0, and so the number of (co-)variance parameters in Σ would be reduced from 15 to 12 or even lower. Moreover, constraining the coefficients on A to a constant, and those of V to a different constant - which implies constraints on two more variance and seven more covariance parameters - would, in the CE dataset, be justified. For the CE dataset the corresponding likelihood ratio statistic is equal to 10.1 (χ^2 null-distribution with 9 degrees of freedom), indicating insignificant loss of model adequacy. Such model simplification would contribute toward reduction of the number of departments with negative department-regression coefficients since less multicollinearity in Σ would lead to smaller estimates of the variances and more pronounced shrinkage.

Modelling Within-department Variation

In Braun and Jones (1985) different within-department variances σ_j^2 are fitted. Their estimates are based on the (iteratively updated) within-department sums of squares of residuals, and so for small departments they have very poor resampling properties. Many departments with small numbers of students with a wide range of backgrounds or a small range of outcomes have very small estimated variances σ_j^2. In the formula for the within-department regression coefficients $\hat{\mathbf{b}}_j$, (11), with the parameters replaced by their estimates, the between-department and within-department regressions are weighted in the proportions of their *estimated* precisions. A small value of the estimate σ_j^2 then causes the within-department regression to be inaccurately regarded as very well determined ($\mathbf{X}_j^T\mathbf{X}_j/\sigma_j^2$ is very large relative to $\mathbf{\Sigma}$), and therefore minimal shrinkage takes place. The plots in Figure 1 demonstrate the association of the estimated regression coefficients with the *estimated* within-department variance. In these plots only the departments with fitted variance $\sigma_j^2 < .1$ are represented. Among the departments with larger fitted variance there are only three instances of negative fitted coefficients (each with respect to the score A). The EB algorithm could be adapted to estimate a common within-department variance σ^2 to hedge against this phenomenon, as well as to promote model parsimony. Technical details are given in Appendix B.

Thus, common variance σ^2 ensures more equitable shrinkage, but we note that overparametrized regression part of the model may cause some of the coefficients for some departments to shrink toward negative values. Therefore, application of extended shrinkage is suitable only in conjunction with careful choice of the EB model.

8. Expectancy Tables

In the current procedures, computation of the expectancy tables involves numerical integration with respect to a five-variate normal density. The number of random draws from the integrating distribution, set at 100, is most likely insufficient, and that causes aberrant features in the simulated expectancy tables. We propose a method that involves no numerical integration and guarantees unimodality of the row- and column-distributions in the expectancy tables.

The posterior distribution of the department-regression coefficients is

$$(\mathbf{b}_j \mid \mathbf{G}, \boldsymbol{\Sigma}, \sigma_j^2) \sim N[\mathbf{r}_j, (\mathbf{P}^* + \mathbf{P}_j)^{-1}],$$

where $\mathbf{r}_j$ is the vector of the posterior means for department j, $\mathbf{P}^* = \boldsymbol{\Sigma}^{-1}$ and $\boldsymbol{\Sigma}$ is the unconditional variance of $\{\mathbf{b}_j\}$, and $\mathbf{P}_j = X_j^T X_j/\sigma_j^2$ is the within-department information matrix.

The fitted regression formula for department j is

$$y_{ij} = \mathbf{x}_{ij}\mathbf{b}_j + \varepsilon_{ij},$$

so that the posterior distribution of the outcome y_{ij} given the vector scores $\mathbf{x}_{ij}$ is

$$y_{ij} \sim N(a_{ij}, d_{ij}), \tag{16}$$

where $a_{ij} = \mathbf{x}_{ij}\mathbf{r}_j$ and $d_{ij} = \sigma_j^2 + \mathbf{x}_{ij}(\mathbf{P}^* + \mathbf{P}_j)^{-1}\mathbf{x}_{ij}^T$. Note the different vectors of background scores $\mathbf{x}_{ij}$ may yield the same mean a_{ij} but different variances d_{ij}.

Calculation of the Expectancy Tables.

The expectancy tables contain the estimated conditional probabilities of the outcome score y_{ij} in a given department, given that predicted score a_{ij} is in a specified range. If we conditioned on the scores $\mathbf{x}_{ij}$, a standard confidence interval could be derived from (16) by ignoring the sampling variability due to estimation of a_{ij} and d_{ij}. If for a given posterior mean a_{ij} the variance d_{ij} as a function of the predictor vector $\mathbf{x}_{ij}$ has a wide range of values, estimation of the probabilities in the expectancy tables could be substantially improved by conditioning on the future scores $\mathbf{x}_{ij}$. For departments with small numbers of students, it would be meaningful to consider the average of the fitted posterior variances $d_j = \Sigma_i d_{ij}/n_j$, and, therefore, for the prediction in the expectancy tables we could use the approximation to (16),

$$y_{ij} \sim N(a_{ij}, d_j).$$

This implies equal variances corresponding to each row of the expectancy table.

For the departments with the largest representation in the GRE VSS data (say, 40 or more students) it may be advantageous to consider separate averages of the posterior variances for the students with fitted scores in each of the specified ranges.

The conditional probabilities are approximated by the formula

$$\Pr\{c_1 < y < c_2 \mid a, G, \sigma_j^2, \Sigma\} = \Phi\{(c_2 - a)/\sqrt{d_{j,2}}\} - \Phi\{(c_1 - a)/\sqrt{d_{j,1}}\},$$

where a is the predicted outcome score (column of the expectancy table), (c_1, c_2) the range of scores, and $d_{j,1}$ and $d_{j,2}$ the corresponding averages of the posterior variances (equal to a common value d_j for small departments). Note that in Section 7 we recommend that a common within-department variance σ^2 be estimated ($\sigma_j^2 \equiv \sigma^2$ for all j).

9. Measures of Quality of the Model Fit

In this section we propose an R^2-type coefficient, specific to a department, that would reflect the quality of the model fit for the data from each department. In the regression analysis of independent observations (e.g., when there are data from one department only) we use the familiar R^2, defined as

$$R^2 = 1 = \sigma^2/\sigma^2_{\text{raw}}, \tag{17}$$

where σ^2 is the residual variance in the assumed model (i.e., with regressors 1, V, Q, A, and U), and σ^2_{raw} is the raw variance of the outcomes FYA or, equivalently, the residual variance in the model with no explanatory variables (regressor 1 only). There are two natural extensions of the definition (1) for the two-level data (students within departments).

For the adopted model (e.g., variables 1 , V, Q, A, and U and the covariate) we have the within-department variances $\{\sigma_j^2\}$ (possibly equal to a common value) and the between-department variance matrix Σ. The variance of an observation in this model is $\sigma_j^2 + \mathbf{x}_{ij}\Sigma\mathbf{x}_{ij}^T$. The two-level analogue of the "empty" model is

$$y_{ij} = \mu + \delta_j + \varepsilon_{ij}, \tag{18}$$

where $\delta_j \sim N(0, \tau^2_{raw}$ and $\varepsilon_{ij} \sim N(0, \sigma^2_{raw})$. The variance of an observation in this model is $\tau^2_{raw} + \sigma^2_{raw}$. Now for the empirical Bayes R^2 we can consider two definitions:

$$\text{A.} \quad R^2 = 1 - (\sigma_j^2 + \mathbf{x}_{ij}\Sigma\mathbf{x}_{ij}^T)/(\sigma^2_{j\,raw} + \tau^2_{raw}). \tag{19a}$$

$$\text{B.} \quad R^2 = 1 - \sigma_j^2/\sigma^2_{j\,raw}. \tag{19b}$$

The definition A is based on the unconditional variance of an observation and the definition B on the within-department variances. To provide a single figure (percentage) for each department, using the definition A, $\mathbf{x}_{ij}\Sigma\mathbf{x}_{ij}^T$ in (19a) should be replaced by the department-mean of these quantities, $\Sigma_i\ \mathbf{x}_{ij}\Sigma\mathbf{x}_{ij}^T/n_j$. Both definitions provide measures of improvement of prediction due to the explanatory variables, and these measures are department-specific. In practice the variance matrix Σ and all the variances are replaced by their maximum likelihood estimates. The definition B will always yield R^2 in the interval (0,1), and it will be constant across the departments if common within-department variances are fitted in both the raw and the assumed models; the definition A will yield values of R^2 outside (0,1) only in the most pathological cases not expected to arise in the GRE VSS data.

An important advantage of these definitions over those in current use for GRE VSS reports is that they involve pooling of information across the departments. The sampling properties of the estimators A and B are only moderately affected by the department size and the within-department distribution of the GRE and U scores. Thus, the definition of predicted R^2 in Braun and Jones (1985), based on within-department half samples, would be suitable only for larger departments; for department sizes 10 and smaller, it has a very large resampling variation because the implied prediction is based on too few observations.

10. Discussion

Although the main purpose of this report is to design adjustments to the EB procedures that would guarantee nonnegative within-department coefficients, it should be emphasized that there are no profound reasons why all the coefficients should be nonnegative. On the one hand, owing to small department sizes strong evidence of a negative coefficient for any particular department is most unlikely. On the other hand, from the EB analyses we have evidence that a small (but significant) proportion of the departments does have negative coefficients. For example, in several EB analyses of the CE dataset both the estimated mean and the estimated standard deviation for the within-department coefficients on the quantitative score are about .06. That implies that about $100\Phi(-1) = 16\%$[5] of the departments have negative coefficients on the quantitative score. The unpredictability of the composition of backgrounds of the students of a department, and the imperfect explanation of the (graduate school) academic performance in terms of the predictors scores, provide a purely substantive explanation for negative coefficients. The complex processes of selection and self-selection of students may, purely by chance, lead to an apparent negative association of a predictor score with the graduate school performance in a small proportion of the departments.

The regression formula is derived from enrolled students, but its application is extrapolated to applicants, who may have much more varied background scores. In addition, the fact that FYA is not a perfect measure of academic performance in graduate school will cause a distortion of the relationship of the academic performance (as a latent variable) on the predictor scores, and since there are a large number of departments, evidence about negativeness of some of the coefficients may strengthen.

[5] Φ is the distribution function of a standard normal variate.

We believe that in the current procedures negative coefficients probably arise much more frequently than it would be reasonable to expect. The proposed procedures can substantially reduce and, if desired, even eliminate occurrences of negative coefficients. The motivation for avoiding negative estimated coefficients is based on the entirely understandable inability to provide a case-by-case (or comprehensive) explanation of why a particular negative coefficient has arisen, and probably on the belief that negative coefficients would be seen as evidence that GRE scores are not very useful predictors. Certainly, negative coefficients are difficult to interpret without reference to the complex and not very well understood processes of selection and self-selection of the student body, and as a consequence, such reports might be regarded by an uninformed client as not useful, or suspected to be incorrect.

However, there are realistic configurations of student background in a department for which the true coefficients *are* negative. After all, we should regard these configurations as outcomes of a random process, and so among the large number of departments there are bound to be a few with extreme or unexpected configurations that are associated with negative coefficients. Therefore, by establishing an unreserved committment to nonnegative estimated coefficients, the GRE VSS is threatened with systematic biases in its reports.

The reported regression formula cannot be used on its own to justify a substantial adjustment of the process of selection of students in the coming academic year. The formula reflects a mixture of two causes: how the background scores are "converted" into academic performance and how successful the selection and self selection processes are. Thus, any substantial change of the selection process will affect the relationship of the studied scores in the future. In the extreme case, if selection of students were based solely on this formula, the selection procedures might be changed over time so dramatically that a substantially different formula for the dependence of FYA on GRE scores would then apply. Also, in prediction formulas based on models with a covariate (say, $V_{.j}$) the "new" value of $V_{.j}$ (unknown at the time) should be applied. Reliance on small variation of the covariate across the years is not justified, and the current procedures do not have any means for adjustment due to uncertainty about the future value of the covariate, which for most departments is the average of a very small number of scores. This raises the issue of the covariate as a suitable representation of the context. The GRE VSS should search for a more valid representation of the context in the employed models. For example, the average of GRE score means, $(V_{.j} + Q_{.j} + A_{.j})/3$, could be considered, but as a contextual covariate it still suffers from the same ills as $V_{.j}$: instability, imperfect representation of the context, and high correlation with the predictor variables.

As an ini‿al step, information about departments that have provided data for several years should be collected. For these departments the stability of the estimated coefficients as well as the quality of the prediction could be assessed.

The procedure of selecting separately for each department a model that has no negative coefficients is prone to serious biases. In EB analysis (potential) bias is a property of the model as applied to the entire dataset. As we select a subset of the data (departments with nonnegative coefficients), the estimators with no bias for the entire dataset may be substantially biased for the selected subset, especially if this subset is of moderate size and if the selection is based on the results themselves.

11. Recommended Changes in the GRE VSS

We recommend that the program staff investigate the feasibility and cost of the following changes in the operational analysis of the GRE VSS:

1. Limit the use of covariates to such an extent that acute collinearity would not arise. In analyses of large datasets, either only one or no covariate should be used. In analyses of smaller datasets, such as students with GRE Subject Test scores, the model should be substantially reduced; no interactions of covariates with the special subject indicators should be used. The guidelines for minimal data sizes for special subjects (at present 100 students from at least 10 departments) should be reviewed and increased substantially.

2. The departments that have provided data over several years should be used for cross-validation and to provide empirical evidence of stability of the regression coefficients

in consecutive years. Changes in the values of the covariates (such as $V_{\cdot j}$) should be recorded because they are a threat to the usefulness of the GRE VSS reports. The current practice of using the last year's value of $V_{\cdot j}$ in the prediction formula for the next year's should be reviewed. It would be more appropriate to use the latest available value for the mean of the verbal scores, or impute its estimate. As an alternative, the prediction formulas could be presented in the form requiring the department to substitute its current value of $V_{\cdot j}$. It would not be appropriate to substitute the mean verbal score of the applicants because of substantial variation in selectivity of the departments.

3. A common value of the within-department variance σ^2 should be used. See Appendix B for technical details.

4. The extended shrinkage should be implemented and the amount of shrinkage recorded. See Appendix A for technical details.

5. For new datasets, pooled ordinary regression models (ignoring between-department variation) with covariate-by-variable interactions should be used to establish the extent of the problem with negative coefficients and to assess multicollinearity of the regression parameters.

6. A single model should be used for the report to all the departments in a dat ıset. This would avoid the "report" bias.

7. The value of the log-likelihood should be used in the choice between candidate models.

8. The procedure for expectancy tables described in Section 8 should be implemented. It will produce results very similar to those obtained by the current procedure, except that errors due to numerical integration would be largely avoided.

9. The minimum numbers of departments and students for a subject test dataset should be reviewed and increased substantially.

References

Braun, H. I., and Jones, D. (1985). *Use of empirical Bayes methods in the study of the validity of academic predictors of graduate school performance.* GRE Professional Report No. 79-13P, Research Report No. 84-34. Princeton, NJ: Educational Testing Service.

Braun, H. I. (1989). Empirical Bayes methods: A tool for exploratory analysis. In R. D. Bock (Ed.), *Multilevel Analysis of Educational Data* (pp. 19-55). San Diego, CA: Academic Press, Inc..

Bryk, A. S., Raudenbush, S. W., Seltzer, M., and Congdon, R.T. (1988). *An introduction to HLM: Computer program and user's guide*. Chicago, IL: University of Chicago.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with Discussion). *JRSS B, 39*, 1-38.

Goldstein, H. (1986). Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika*, *73*, 43-56.

Hoerl, A. E., and Kennard, R. W. (1970). Ridge regression: Biased estimation for the nonorthogonal problems. *Technometrics*, *12*, 55-66.

LaMotte, L. R. (1972). Notes on the covariance matrix of a random nested ANOVA model. *Annals of Mathematical Statistics*, *43*, 659-62.

Lindstrom, M. J., and Bates, D. M. (1988). Newton-Raphson and EM algorithms for linear mixed effects models for repeated-measures data. *Journal of the American Statistical Association*, *83*, 1014-1032.

Longford (1987). A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects. *Biometrika*, *74*, 817-827.

Longford (1988). *VARCL Software for variance component analysis of data with hierarchically nested random effects (maximum likelihood).* [Manual] Princeton, NJ: Educational Testing Service.

Morris, C. N. (1983). Parametric empirical Bayes inference: Theory and applications (with Discussion). *Journal of the American Statistical Association*, *78*, 47-65.

References (Continued)

Rasbash, J., Prosser, R., and Goldstein, H. (1988). *ML2 Software for two-level analysis.* [User's guide] London, England, University of London, Institute of Education.

Raudenbush, S. W., and Bryk, A. S. (1986). A hierarchical model for studying school effects. *Sociology of Education*, *59*, 1-17.

Appendix A

Extended Shrinkage Empirical Bayes Estimation

The empirical Bayes estimates of the within-department coefficients are given by the formula

$$\mathbf{r}_j = (\hat{\mathbf{P}}^* + \hat{\mathbf{P}}_j)^{-1}(\hat{\mathbf{P}}^* \hat{\mathbf{G}}^T\mathbf{Z}_j + \hat{\mathbf{P}}_j B_j) \tag{A.1}$$

where

$\hat{\mathbf{P}}^*$ = $\hat{\Sigma}^{-1}$ is the inverse of the estimated between-department variance matrix,

$\hat{\mathbf{P}}_j$ = $\sigma_j^2 \mathbf{X}_j^T\mathbf{X}_j$ is the estimated within-department information matrix,

$\hat{\mathbf{B}}_j$ = $(\mathbf{X}_j^T\mathbf{X}_j)^{-1}\mathbf{X}_j^T\mathbf{y}_j$ is the least-squares estimate of the within-department coefficients,

$\hat{\mathbf{G}}$ is the estimate of the department-level regression coefficients,

and

$\mathbf{Z}_j$ is the vector of covariates for the department j,

(Braun and Jones, 1985). The estimator $\mathbf{r}_j$ is a mixture of the within-department estimate $\hat{\mathbf{B}}_j$, which is unbiased but inefficient, and the pooled regression estimate $\hat{\mathbf{B}}_j^* = \hat{\mathbf{G}}^T\mathbf{Z}_j$, which is suitable for the "average" department. The advantage of the empirical Bayes method is in optimal "trading" of unbiasedness for efficiency (lowest mean-square error). The optimality properties hold under the unrealistic assumption of known variance and covariance parameters, $\{\sigma_j^2\}$ and the elements of Σ. The optimality of the empirical Bayes estimates is under threat when inappropriate models are used and when the estimates of the variances and covariances are subject to substantial sampling variation. From the form of (A.1) we can deduce that these problems are particularly acute if the estimates of some of the within-department variances σ_j^2 are very small. As observed in Section 7 (see Figure 1) most of the negative coefficients occur for such departments.

We propose to adapt the empirical Bayes estimator to satisfy the additional constraint of nonnegativity of the regression coefficients at minimal loss of efficiency. We consider a more general class of estimators:

$$\mathbf{r}_j(c_j) = (\hat{\mathbf{P}}^* + c_j\hat{\mathbf{P}}_j)^{-1}(\hat{\mathbf{P}}^*\hat{\mathbf{B}}_j^* + c_j\hat{\mathbf{P}}_j\hat{\mathbf{B}}_j), \tag{A.2}$$

where $0 \leq c_j \leq 1$ is a department-specific constant (to be chosen by the analyst). The extreme choices are $\mathbf{r}_j = \mathbf{r}_j(1)$, the empirical Bayes solution, and $\mathbf{r}_j(0) = \hat{\mathbf{B}}_j^*$, the least-squares solution for the pooled dataset.

For the simplest department-level design, $\mathbf{Z}_j \equiv 1$, we have

$$\mathbf{r}_j(0) = \hat{\mathbf{G}}^T,$$

which is expected to be positive for each dataset. More complex choices for $\mathbf{Z}_j$, such as $\mathbf{Z}_j = (1, V_{\cdot j}, A_{\cdot j})$, where $V_{\cdot j}$ and $A_{\cdot j}$ are the department-means for verbal and analytical scores, respectively, should be admitted only when $\hat{\mathbf{G}}^T\mathbf{Z}_j = (\hat{g}_{11} + \hat{g}_{12}V_{\cdot j} + \hat{g}_{13}A_{\cdot j}, \hat{g}_{21} + \hat{g}_{22}V_{\cdot j} + \hat{g}_{23}A_{\cdot j}, \ldots)$ have nonnegative components for all values of $V_{\cdot j}$ and $A_{\cdot j}$ that occur in the data.

Let's assume that $\mathbf{r}_j(0) = \hat{\mathbf{G}}^T\mathbf{Z}_j$ is positive for all departments and suppose that the k^{th} component of $\mathbf{r}_j(1)$ is negative. Then there is a constant c_j such that the second component of $\mathbf{r}_j(c_j)$ is equal to zero. This constant can be found by a simple iterative procedure: As an initial approximation we set

$$c_{j,INI} = (\hat{\mathbf{G}}^T\mathbf{Z}_j)_k / \{(\hat{\mathbf{G}}^T\mathbf{Z}_j)_k - [\mathbf{r}_j(0)]_k\}, \tag{A.3a}$$

where the subscript k denotes the k^{th} component of the vector. This value of c_j is used to evaluate (A.2). If the k^{th} component of the new vector $\mathbf{r}_j(c_j)$ is not close enough to zero, we essentially iterate (A.3a) by updating

$$c_{j,NEW} = c_{j,OLD}(\hat{\mathbf{G}}^T\mathbf{Z}_j)_k \ / \ \{(\hat{\mathbf{G}}^T\mathbf{Z}_j)_k - [\mathbf{r}_j(c_{j,OLD})]_k\}. \tag{A.3b}$$

The iterative formula (A.3b) would be applied until $|[\mathbf{r}_j(c_{j,OLD})]_k| < .002$.

Applying no shrinkage corresponds to $c_j = 1$. Repeated application of the extended shrinkage (more than one negative coefficient for a department) corresponds to the product of the shrinkage coefficients. The amount of shrinkage could be effectively monitored by recording all the departments for which it was employed, together with the shrinkage coefficients, and a suitable summary would be the total shrinkage $\Sigma_j\, c_j$.

As an alternative the linear Taylor expansion for $\mathbf{r}_j(c_j)$

$$\mathbf{r}_j(c_j) \approx \mathbf{r}_j(1) + (1 - c_j)(\hat{\mathbf{P}}^* + \hat{\mathbf{P}}_j)^{-1}\hat{\mathbf{P}}_j\{\hat{\mathbf{B}}_j - \mathbf{r}_j(1)\} \tag{A.4}$$

at $c_j = 1$ could be used. This approximation can be used iteratively until a constant c_j is found for which the component of $\mathbf{r}_j(c_j)$ is close enough to zero (so that after rounding to the usual number of decimal places it would be reported as .0). If a different component of $\mathbf{r}_j(c_j)$ is negative, the procedure will be repeated for that component.

The procedure based on (A.3b) is much simpler and requires only a moderate number of iterations (usually less than 6).

Appendix B

Common Within-department Variance

The maximum likelihood estimator for the common within-department variance is given by the formula

$$\sigma^2 = \mathbf{e}\hat{\mathbf{V}}^{-1}\mathbf{e}^{\mathrm{T}}/N, \tag{A.5}$$

where **e** is the vector of student-level residuals, $\mathbf{e} = \mathbf{y} - \mathbf{X}\hat{\mathbf{G}}\mathbf{Z}$, $\hat{\mathbf{V}}$ is the $\hat{\sigma}^2$-multiple of the estimated variance matrix for the observations **y** (FYA-scores),

$$\hat{\mathbf{V}} = \mathbf{I}_N + \mathit{diag}_j\{\mathbf{X}_j^{\mathrm{T}}\,\hat{\mathbf{\Omega}}\mathbf{X}_j\},$$

$\mathbf{I}_N$ the N×N identity matrix, $\hat{\mathbf{\Omega}} = \hat{\mathbf{\Sigma}}/\hat{\sigma}^2$, $\mathbf{X}_j$ the segment of the design matrix **X** corresponding to the department j, and N the number of students in the dataset. The matrix $\hat{\mathbf{V}}$ depends on the estimates of the variances and covariances, and therefore it has to be updated at every iteration. Since $\hat{\mathbf{V}}$ is a block-diagonal patterned matrix, formulas for evaluation of (A.5) without inversion of any large matrices can be employed; see LaMotte (1972) or Longford (1987). We have

$$\hat{\mathbf{V}}^{-1} = \mathbf{I}_N - \hat{\sigma}^{-2}\,\mathit{diag}_j\{\mathbf{X}_j^{\mathrm{T}}(\hat{\mathbf{\Sigma}}^{-1} + \mathbf{X}_j\mathbf{X}_j^{\mathrm{T}}/\hat{\sigma}^2)^{-1}\mathbf{X}_j\}, \tag{A.6}$$

and hence

$$\hat{\sigma}^2 = \mathbf{e}\mathbf{e}^{\mathrm{T}}/N - \Sigma_j\,\mathbf{e}_j\mathbf{X}_j^{\mathrm{T}}(\hat{\mathbf{\Sigma}}^{-1} + \mathbf{X}_j\mathbf{X}_j^{\mathrm{T}}/\hat{\sigma}^2)^{-1}\mathbf{X}_j\mathbf{e}_j^{\mathrm{T}}, \tag{A.7}$$

where

$$\mathbf{e}\mathbf{e}^{\mathrm{T}} = \Sigma_j\,\mathbf{e}_j\mathbf{e}_j^{\mathrm{T}} = \Sigma_j(\mathbf{Y}_j\mathbf{Y}_j^{\mathrm{T}} - 2\mathbf{Y}_j\mathbf{X}_j^{\mathrm{T}}\hat{\mathbf{G}}^{\mathrm{T}}\mathbf{Z}_j^{\mathrm{T}} + \mathbf{Z}_j^{\mathrm{T}}\hat{\mathbf{G}}^{\mathrm{T}}\mathbf{X}_j\mathbf{X}_j^{\mathrm{T}}\hat{\mathbf{G}}\mathbf{Z}_j)$$

and

$$\mathbf{X}_j\mathbf{e}_j^T = \mathbf{X}_j\mathbf{Y}_j^T - \mathbf{X}_j\mathbf{X}_j^T\hat{\mathbf{G}}^T\mathbf{Z}_j^T.$$

Note that $\hat{\boldsymbol{\Sigma}}^{-1} + \mathbf{X}_j\mathbf{X}_j^T/\hat{\sigma}^2$ in (A.6) is equal to $\hat{\mathbf{P}}^{*} + \hat{\mathbf{P}}_j$.

In an iterative procedure the residual mean-squared error from the pooled ordinary regression can be used as the initial estimate for σ^2. For the CE dataset the estimate of the common variance σ^2 is about .10.

Table 1

Ordinary Regression Models for the Departments with Large Numbers of Students

Department 1 has 102 students, department 229, 106 students. Standard errors for the estimated regression parameters are in parentheses.

Department 1
$pFYA = 1.355 + \underset{(.103)}{.532U} + \underset{(.128)}{.068V} + \underset{(.115)}{.055Q} - \underset{(.114)}{.108A}$
$pFYA = 2.926 - \underset{(1.026)}{.662U} + \underset{(.128)}{.0723V} + \underset{(.116)}{.064Q} + \underset{(.492)}{.082A} + \underset{(.164)}{.192U^2} + \underset{(.091)}{.041A^2}$
Department 229
$pFYA = 1.944 + \underset{(.089)}{.378U} + \underset{(.072)}{.121V} - \underset{(.085)}{.010Q} - \underset{(.063)}{.007A}$
$pFYA = 4.498 - \underset{(1.499)}{1.203U} + \underset{(.074)}{.105V} - \underset{(.086)}{.026Q} + \underset{(.064)}{.020A} + \underset{(.234)}{.248U^2}$

Table 2

Regression of the Operational Model and Submodel Coefficients on the Covariate $V_{\cdot j}$

The operational model is given by (7), that is using the department mean verbal score as a covariate. The submodel is obtained from (7) by deleting the covariate, that is, by setting $g_{cv} = g_{vv} = g_{qv} = g_{av} = g_{uv} = 0$. The operational software was used to fit these two models (500 iterations). The resulting department coefficients were then regressed, using the ordinary regression with equal weights, on the department mean score $V_{\cdot j}$. The standard errors corresponding to the ordinary regression are given in parentheses.

Coefficient	Submodel	Operational Model
U	$.127 + .035V_{\cdot j}$ (.012)	$.280 + .020V_{\cdot j}$ (.013)
V	$.102 - .008V_{\cdot j}$ (.004)	$.280 - .008V_{\cdot j}$ (.004)
Q	$.065 - .002V_{\cdot j}$ (.004)	$.193 - .005V_{\cdot j}$ (.004)
A	$-.006 + .013V_{\cdot j}$ (.003)	$.039 - .003V_{\cdot j}$ (.003)

Figure 1.a Negative coefficients on the undergraduate grade-point average and fitted within-department variance.

The horizontal axis is the fitted within-department variance. The vertical axis is the fitted EB coefficient on the undergraduate grade-point average. Each asterisk * represents one department. The plot *i*, refers to the results with the model (7), covariate $V_{.j}$, the plot *ii*, to the model with no covariate. The plots contain only departments with fitted within-department variance smaller than .10. Among the omitted departments there are no negative fitted EB coefficients for either model.

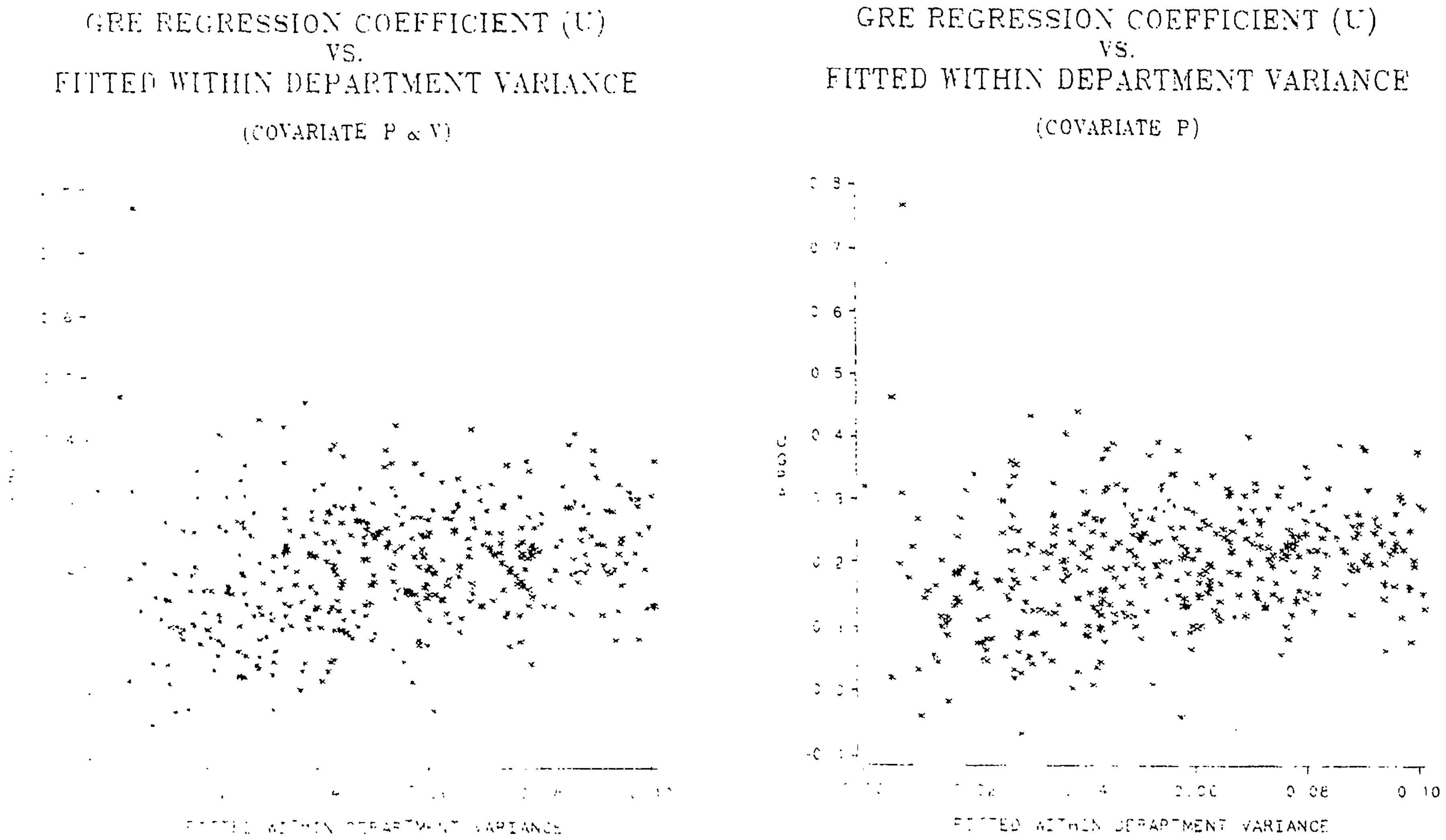
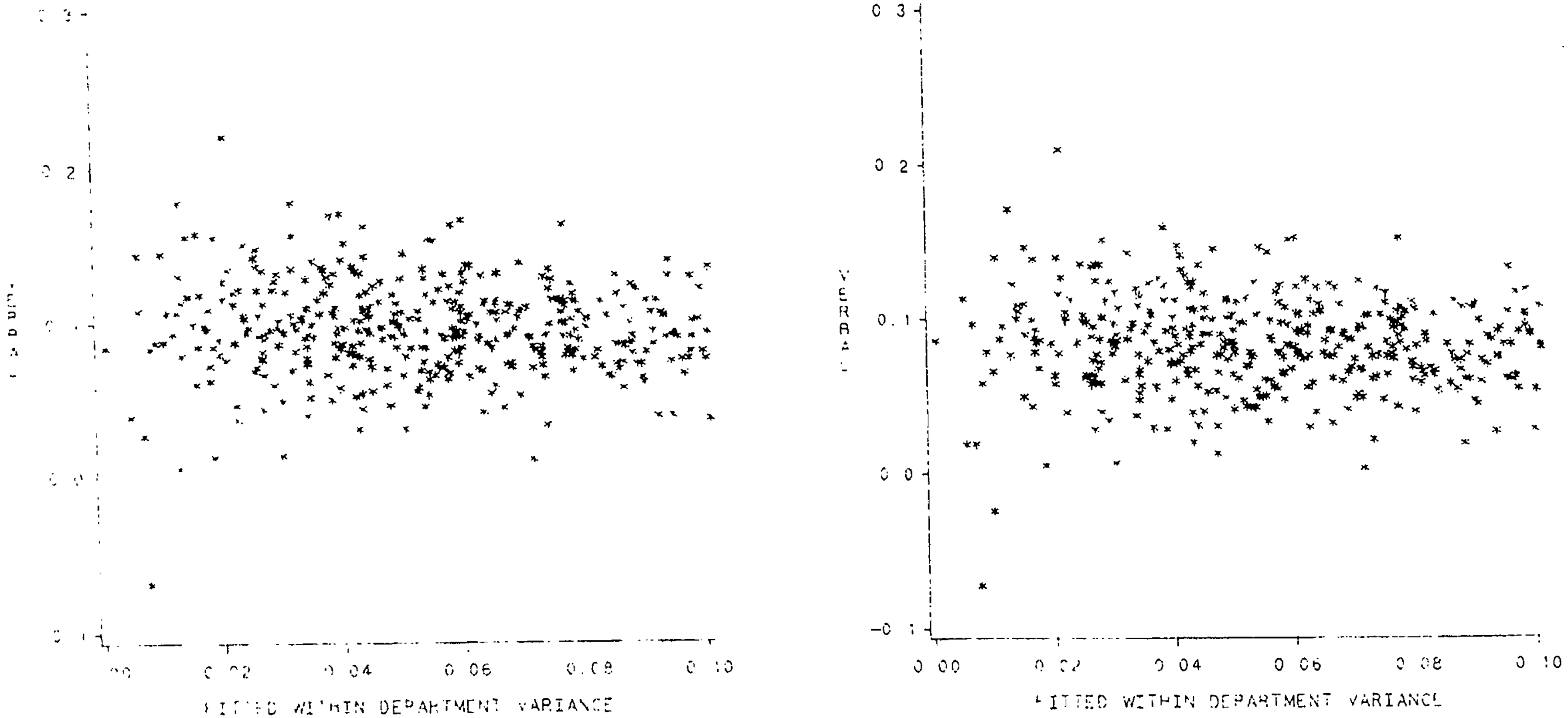


Figure 1.b Negative coefficients on the verbal score and fitted within-department variance.

The horizontal axis is the fitted within-department variance. The vertical axis is the fitted EB coefficent on the verbal score. Each asterisk * represents one department. The plot *i*, refers to the results with the model (7), covariate $V_{\cdot j}$, the plot *ii*, to the model with no covariate. The plots contain only departments with fitted within-department variance smaller than .10. Among the omitted departments there are no negative fitted EB coefficients for either model.

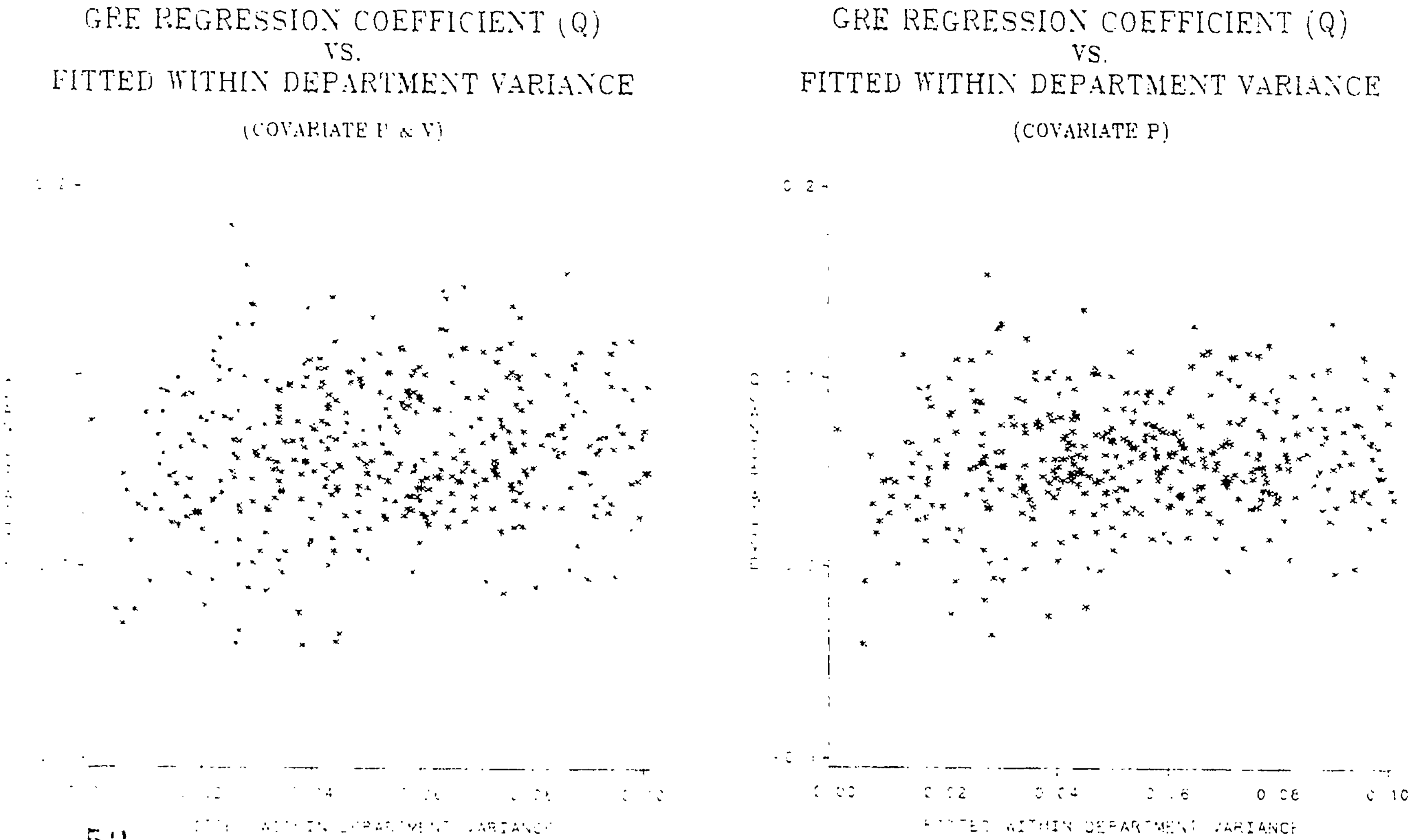


GRE REGRESSION COEFFICIENT (V)
VS.
FITTED WITHIN DEPARTMENT VARIANCE
(COVARIATE P & V)

GRE REGRESSION COEFFICIENT (V)
VS.
FITTED WITHIN DEPARTMENT VARIANCE
(COVARIATE P)

Figure 1.c Negative coefficients on the quantitative score and fitted within-department variance.

The horizontal axis is the fitted within-department variance. The vertical axis is the fitted EB coefficent on the quantitative score. Each asterisk * represents one department. The plot *i*, refers to the results with the model (7), covariate $V_{\cdot j}$, the plot *ii*, to the model with no covariate. The plots contain only departments with fitted within-department variance smaller than .10. Among the omitted departments there are no negative fitted EB coefficients for either model.

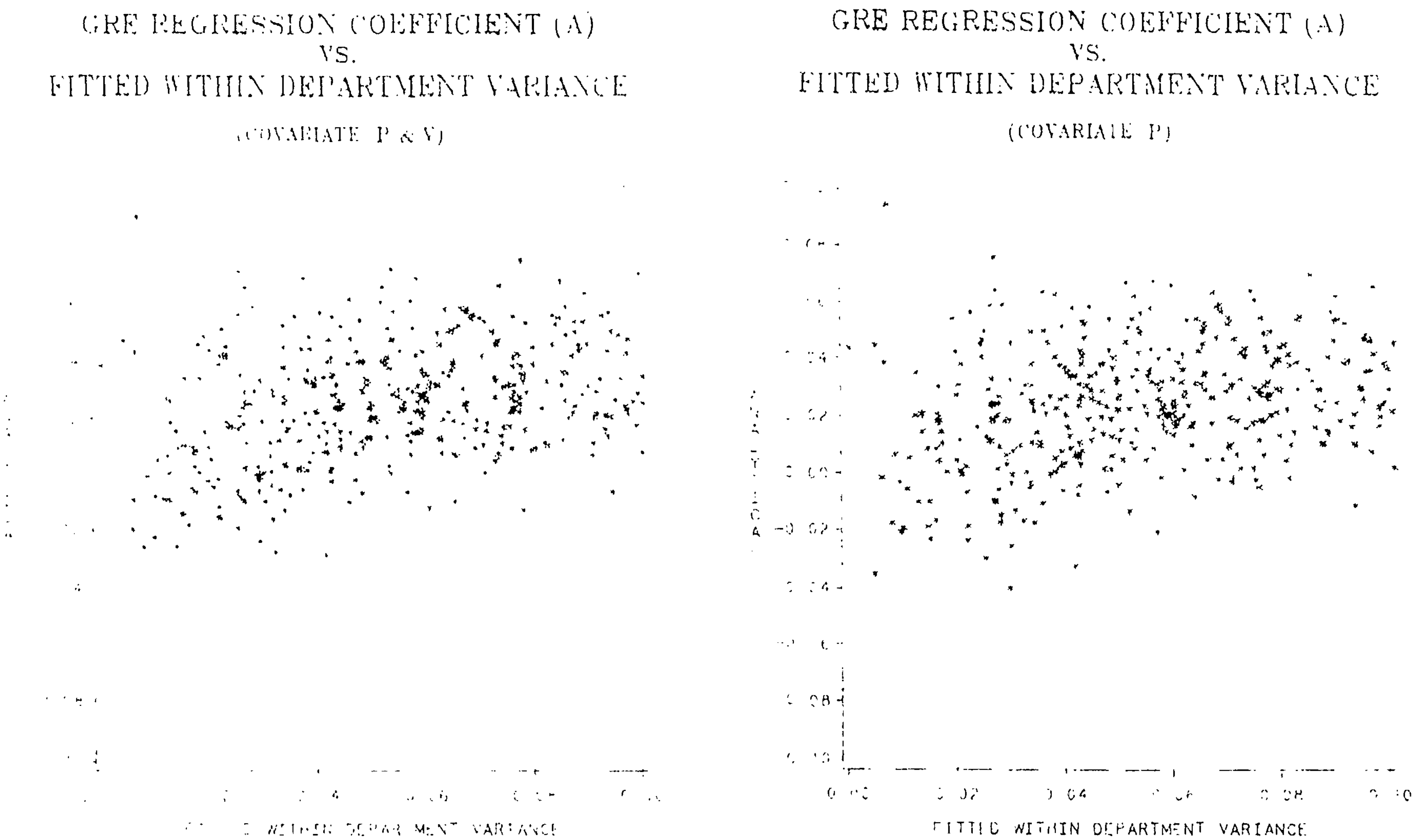


GRE REGRESSION COEFFICIENT (Q)
VS.
FITTED WITHIN DEPARTMENT VARIANCE
(COVARIATE P & V)

GRE REGRESSION COEFFICIENT (Q)
VS.
FITTED WITHIN DEPARTMENT VARIANCE
(COVARIATE P)

Figure 1.d Negative coefficients on the analytical score and fitted within-department variance.

The horizontal axis is the fitted within-department variance. The vertical axis is the fitted EB coefficent on the analytical score. Each asterisk * represents one department. The plot *i*, refers to the results with the model (7), covariate $V_{.j}$, the plot *ii*, to the model with no covariate. The plots contain only departments with fitted within-department variance smaller than .10. Among the omitted departments there are three negative fitted EB coefficients for either model, these departments have fitted within-department variances in the range .12 - .25. The largest fitted within-department variance is .75.


GRE REGRESSION COEFFICIENT (A)
VS.
FITTED WITHIN DEPARTMENT VARIANCE
(COVARIATE P & V)
WITHIN DEPARTMENT VARIANCE
GRE REGRESSION COEFFICIENT (A)
VS.
FITTED WITHIN DEPARTMENT VARIANCE
(COVARIATE P)
FITTED WITHIN DEPARTMENT VARIANCE