Estimating Treatment Effects with Multiple Proxies of Academic Aptitude

Andrew Grodner, Assistant Professor,
Economics Department, East Carolina University, agrodner@andrewgrodner.com

Wayne A. Grove, Associate Professor,
Economics Department, Le Moyne College, grovewa@lemoyne.edu

NOVEMBER 2007

Abstract

Existing studies of college student learning gains due to pedagogical practices typically rely on a wide variety of proxy variables for academic aptitude. This study questions the wisdom of this approach because each proxy, or particular combinations of proxies, measures latent ability with distinctly different substantial errors. We consider four econometric approaches to identify the best proxies for latent academic aptitude: principle components, factor analysis, post hoc estimator, and instrumental variables. Our estimates suggest that collegiate GPA best controls for the students' ability to learn economics and the existing literature overestimates the treatment effect of faculty pedagogical practices.

Key words: Academic Aptitude, Proxy variables, Student learning, Methodology, Undergraduate economics,
JEL Code: A22, **C81**, I21

We would like to acknowledge the valuable comments by Dan Black, Kevin Rask, Darren Lubotsky, Martin Wittenberg. We are grateful to the administrators at Syracuse University for their cooperation with our research and to Le Moyne College for a Research & Development support.

A substantial body of literature estimates how different pedagogical practices or techniques in the classroom influence college student learning of economics. Researchers estimate measures of student learning (such as course grades) in an educational production function with a dichotomous variable for participation in a treatment group (such as the use of PowerPoint lectures), along with a variety of control variables to account for the non-random selection of students into classes. Despite William Becker's (1997) finding that academic aptitude variables are the only consistently significant and meaningful explanatory variables of undergraduate student learning of economics, we show elsewhere (Grove et al., 2006) that scholars use a wide array of control measures for academic ability and that doing so meaningfully influences estimated student learning treatment effect.[1] Since each proxy or particular combinations of proxies used probably measures latent ability with distinctly different substantial errors, our focus in this paper is on the academic aptitude covariates.

We reconsider the literature of student learning studies with the goal of estimating which of the typically available academic ability measures (some variant of high school GPA or rank, college entrance exam scores, or college GPA) offer the best proxy for latent academic ability. Four econometric approaches are employed to identify the best proxies for latent academic aptitude from a large variety of student performance variables combining multiple proxies via principle components or factor analysis, using the additional proxies as instruments, and a regression-based post hoc estimator recently proposed by Lubotsky and Wittenberg. To identify the best proxies for student academic aptitude, we first extract an estimate of latent academic ability from our rich set of proxy measures using factor analysis,

[1] Seldom do scholars explain the logic of their proxy choices or how alternative ability measures affect their estimates of cognitive achievement. For an exception, see Kennedy and Siegfried (1997, p. 388) who argue for using the composite SAT score instead of math- and verbal-SAT scores separately, ACT scores, or GPA. We

principal components, IV, and "post hoc" estimation. The best proxies, then, are those most highly correlated with estimated latent aptitude variables. We find that collegiate GPA best predicts students' ability to learn economics, whereas researchers have relied most heavily upon standardized college entrance exam i.e., the SAT or ACT, scores. Our estimates suggest that the existing literature overestimates the treatment effect of faculty pedagogical practices and illustrates the value of carefully selecting proxies in this and in other similar contexts.

II. Estimation - unified description of Principal Components, Factor Analysis, and "Post Hoc" estimator[2]

Our baseline econometric model can be described as follows:

$$Y = \alpha + \gamma T + \beta A + \in \qquad (1)$$

where *Y* is the academic outcome such as exam grade, *T* is the treatment effect such as problem set assignment, and *A* is the latent ability which is not observed. The fundamental problem is to estimate the coefficient of interest *γ,* which is biased if *T* and *A* are correlated and *A* is not observed. Thus, further interest is on describing how to best control for the unobserved ability, *A*.

We now present a unified description of the model in (1) when there are unobserved proxies that affect the exam grade. In a general form, our full structural model is:

$$Y_i = \beta A_{1i} + \varepsilon_i \qquad (2)$$

---

found that the variability of the proxy choice alone caused estimated learning gains to range from a C+ to less than a B- or to a B (see Grove, Wasserman and Grodner, 2006, p. 131).

[2] The following discussion is based on Johnson and Wichern (1992), Stata Reference for the procedure "factor",

$$\begin{cases} X_{1i} = \rho_{11} A_{1i} + \rho_{12} A_{2i} + \ldots + \rho_{1m} A_{mi} + u_{1i} \\ X_{2i} = \rho_{21} A_{1i} + \rho_{22} A_{2i} + \ldots + \rho_{2m} A_{mi} + u_{2i} \\ \quad\quad\quad\quad\quad . \\ \quad\quad\quad\quad\quad . \\ \quad\quad\quad\quad\quad . \\ X_{pi} = \rho_{p1} A_{1i} + \rho_{p2} A_{2i} + \ldots + \rho_{pm} A_{mi} + u_{pi} \end{cases} \tag{3}$$

where Y is the value of the dependent variable (mean test score) for observation i, A_i is the true but unobserved independent variable (aptitude) affecting the dependent variable Y, variables $X_1,\ldots,X_p$ are p proxies for A_1, the variables $A_{2,\ldots,}A_m$ are other variables that do not affect Y but do affect proxies X_j, β is the parameter of interest, ε is the error term, ρ_{jk} are the parameters representing the effect of A_k on the proxy variables X_j. In our setting, Y is the mean value of the test score for student i, A_i is unobserved student academic ability, X_j are the student ability proxies, such as GPA measures, high school grades, SAT scores, etc.

Although researchers usually focus to find the best estimate of β given proxies X_j, in our analysis the estimate β is of secondary interest. Instead, we focus on selecting the best subset of proxies X_j in order to control for, but not necessary estimate, the effect of unobserved ability A_1 on test scores Y. We use three multivariate techniques: principal components analysis, factor analysis, and the post hoc estimator. Principal components analysis ignores equation (2) and assumes no error terms u_j in (3). Factor analysis also ignores the relationship in (2) but allows randomness of uncorrelated $u_j^{'}$s. The post hoc estimator takes into account the relationship in (2) and allows for $cov(u_{ji}\, u_{kj}) \neq 0$. However, post hoc estimator assumes $\rho_{j2},\ldots,\rho_{jm} = 0$, which means that either there are no independent variables affecting proxies X other than A_1, or other A_j variables are included in the error terms u_j. Below we formally present the differences between methods.

---

and Lubotsky and Wittenberg (2006).

A. Principal Components Analysis

Principal component analysis assumes a set of p observable and highly correlated random variables **X'**=$\lfloor X_1 X_2 ... X_p \rfloor$ (i.e. the aptitude proxies) with the means **u**'=$\lfloor u_1 u_2 ... u_p \rfloor$ and variance-covariance matrix ∑. The principal components model postulates that we can summarize the variance in all **X** by generating an uncorrelated principal components **A'**=$\lfloor A_1 A_2 ... A_p \rfloor$ which are a linear combinations of the original variables in **X**. Formally, the model in (2) and (3) reduces to:

$$
\begin{aligned}
A_1 &= v_{11}x_1 + v_{11}x_2 + ...v_{1p}x_p \\
A_2 &= v_{21}x_1 + v_{22}x_2 + ...v_{2p}xp \\
&. \\
&. \\
&. \\
A_p &= v_{p1}x_1 + v_{p2}x_2 + ...v_{pp}x_p
\end{aligned}
$$

where v_{ij} is the coefficient of the ith principal component on the jth variable with the matrix of coefficients **V**, and x_j is the standardized variable with mean 0 and variance 1 (i.e. $x_j = (X_j - u_j)/\sigma_j$), stored in the matrix **x'**=$\lfloor x_1 x_2 ... x_p \rfloor$. In the matrix notation we can write the model as:

A=Vx

The principal components analysis seeks to find a **V** such that the resulting variables in **A** are uncorrelated and their variances are maximized. The solution is found by using the eigenvalue decomposition of the covariance matrix

∑=VSV'

where **S** is a diagonal matrix whose diagonal elements are the eigenvalues of ∑, and **V** is an orthogonal matrix whose columns form the set of eigenvectors. By convention, the eigenvalues in are arranged in a decreasing order. Therefore, the covariance matrix for **A** becomes

$$\text{var}(\mathbf{A})=\mathbf{V'}\sum\mathbf{V}=\mathbf{V'}(\mathbf{VSV'})\mathbf{V}=\mathbf{S}$$

Intuitively, the principal components procedure generates p new variables (**A**) from original p variables (**X**) such that new variables (**A**) are uncorrelated and have the highest variance. The first A_1 variable has the highest variance of all principal components and summarizes the greatest amount of the variation in **X**. Notice that for the principal components analysis to be effective in our study the first principal component needs reflect the unobserved ability. Therefore, we implicitly assume that the ability determines all our proxy variables and explains most variation between them.

B. Factor Analysis

Factor analysis assumes a set of p observable random variables $\mathbf{X'}=\lfloor X_1 X_2 \ldots X_p \rfloor$ (i.e., the aptitude proxies), with the means $\mathbf{u'}=\lfloor u_1 u_2 \ldots u_p \rfloor$ and variance-covariance matrix $\sum$. The factor model postulates that **X** is linearly dependent upon a few unobservable random variables $\mathbf{A'}=[A_1 A_2 \ldots A_m]$, called common factors, and p additional unobservable sources of variation $\mathbf{u'}=\lfloor u_1 u_2 \ldots u_p \rfloor$, called errors, with the variance-covariance matrix $\mathbf{\Psi}$. In particular, the model in represented in equations (2) and (3) reduces to:

$$x_1 = \frac{X_1 - u_1}{\sigma_1} = l_{11}A_1 + l_{12}A_2 + \ldots l_{1m}A_m + u_1$$
$$x_2 = \frac{X_2 - u_u}{\sigma_2} = l_{21}A_1 + l_{22}A_2 + \ldots l_{2m}A_m + u_2$$
$$\ldots$$
$$x_p = \frac{X_p - u_p}{\sigma_p} = l_{p1}A + l_{p2}A_2 + \ldots l_{pm}A_m + u_p$$

where l_{ij} is called the loading of the *i*th variable on the *j*th factor with the matrix of factor loading **L**, and x_i is the standardized variable with mean 0 and variance 1, stored in the matrix **X**. In the matrix notation we can now write the model as:

$$\mathbf{x}=\mathbf{LA}+\mathbf{u}$$

The model is based on the following assumptions (besides the linearity of the relationship between random variables and factors):

$$\mathrm{E}(\mathbf{A})=\mathbf{0},\ \mathrm{Cov}(\mathbf{A})=\mathrm{E}[\mathbf{AA'}]=\mathbf{I}$$

$$\mathrm{E}(\mathbf{u})=\mathbf{0},\ \mathrm{Cov}(\mathbf{uu'})=\boldsymbol{\Psi}=\begin{bmatrix}\psi_1 & 0 & \cdots & 0\\ 0 & \psi_2 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \psi_p\end{bmatrix}$$

$$\mathrm{Cov}(\mathbf{u},\mathbf{A})=\mathbf{0}$$

We can see that all A_j factors are independent random variables with mean zero and variance one, where and the errors are independently, identically distributed (iid). The covariances can be decomposed as

$$\mathrm{Var}(\mathbf{X})=\sum=\mathbf{LL'}+\boldsymbol{\Psi} \tag{4}$$

$$\mathrm{Cov}(\mathbf{X},\mathbf{L})=\mathrm{L}$$

The model assumes that $p(p+1)/2$ variances and covariances for **X** can identify pm factor loadings l_{ij} and p specific variances ψ_i. Thus, there is a limit on the number of factors m that can be identified from a subset of variables p based on the relation $\frac{p(p+1)}{2} \geq p(m+1)$. For example, with two factors there should be at least five independent variables to identify the two factors.

Operationally, Stata solves the system in (4) by first estimating **Ψ**, and then the columns of **L** are computed as the eigenvectors and scaled by the square root of the appropriate eigenvalue (similar to the principal components procedure). The vector of scores for a single observation is computed as

$$\mathbf{t} \text{ [illegible] } \mathbf{L} \text{ [illegible] } \mathbf{x}$$

Where $\hat{\mathbf{L}}$ is a sample estimate of $\mathbf{L}$ and $\hat{\Sigma}$ is a sample estimate of Σ. As a result of this formulation each factor is distributed with mean 0 and variance 1 and factors are orthogonal to each other (have correlation 0).

The intuition of factor analysis is that if we run a regression of dependent variable y on a set of independent variables **X**, create factors **A** from independent variables in **X**, and then run a regression of y on **A**, the fit should be approximately the same in the two regressions (a smaller number of factors is supposed to summarize most of the information in**X**). As with the principal components analysis, the validity of using factor analysis in our empirical work is based on the fact that the first, highest scoring factor is the unobserved ability, A_1.

C. Lubotsky-Wittenberg's Post hoc estimation

The post hoc estimator assumes a set of proxies that can be correlated with each other which together determine a part of the unobserved aptitude A. The model in equations (2) and (3) takes the form:

$$Y_i = \beta A + \varepsilon_i$$

$$\begin{cases} X_{1i} = \rho_1 A_i + u_{1i} \\ X_{2i} = \rho_2 A_i + u_{2i} \\ \qquad \vdots \\ X_{pi} = \rho_p A_i + u_{pi} \end{cases}$$

where ρ_j is the coefficient relating proxies X_j with the latent variable A, with ρ_1 being standardized to equal one (standardization eases the interpretation of results). The model allows for correlation between proxies ($\mathrm{cov}(u_j, u_k) \neq 0$) but not a correlation of proxies with the dependent variable and the error term in A1 equation ($\mathrm{cov}(Y, u_j) \equiv 0$ and $\mathrm{cov}(\varepsilon, u_j) \equiv 0$). Notice that in contrast to principal components analysis and factor analysis the model for the post hoc estimator assumes that there is only one factor affecting all the proxies. Other potential variables ($A_{2,\dots,} A_m$) are assumed not to affect the proxies X_J and they are implicitly included in the error terms u_j.

One approach to estimate β is to generate an index variable

$$A^{\delta} = \delta_1 X_1 + \delta_2 X_2 + \cdots + \delta_p X_p$$

and use it in the original regression

$$Y_i = \widehat{\beta} A_i^5 + \zeta_i$$

Where ζ_i is the error term for some $\delta_{1,\dots,}\ \delta_{\mathrm{p}}$. The coefficient $\hat{\beta}$ suffers from the attenuation bias - a measurement error due to u_j's in equation (3) that causes $\mathrm{cov}(A^5, \zeta) \neq 0$. Lubotsky and Wittenberg (2006), however, show that the attenuation bias is minimized if the δ_j's are constructed as

$$\delta_j = \frac{b_j}{\sum_{j-1}^{k} b_j \left[\mathrm{cov}(Y, X_j) / \mathrm{cov}(Y, X_1)\right]} = \frac{b_j}{b_p}$$

where b_p is called the post hoc estimator. Parameters b_j are obtained from the direct regression

$$Y_i = b_1 X_{1i} + b_2 X_{2i} + \cdots b_p X_{pi}$$

Intuitively, the post hoc procedure gives the optimal weights to each individual proxy in constructing the latent index, A^5, so that the information from each proxy is proportional to its contribution in explaining the latent variable A. In our setting the constructed index A^5 is the proxy for the unobserved scholastic ability variable.

III. Data: A Quasi-Natural Experiment

Our data comes from both university administrative student records and from one of the author's class grade records. During the fall of 1998, 239 undergraduate students enrolled in and completed four sections of introductory microeconomics taught by one of the authors at Syracuse University, a large, private residential university in the northeast (Carnegie Classification: Doctoral Research Universities II—Extensive). A "natural experiment" separated students into one group (of 143) whose course grades were based on problem set performance and another (of 96) whose course grades were not. Students with graded problem sets received a course grade that included the average of the best four of five possible problem set grades. All lectures, handouts, exams, and review sessions were as identical as possible. All students received the problem sets at the same time and encouragement to practice economics by solving them as important preparation for the exams. When the problem sets were handed in, students in both groups received the answer keys. Mastery of the course material was measured by performance on three required exams of equal value, each of which was assessed with a 0 to 100-point scale. To ensure as much uniformity in grading as possible, the same grader evaluated each exam question for all 239 students. Thus, the most discernible difference between the four sections of introductory microeconomics was that students in three sections had a direct grade-based incentive to practice economics problems throughout the semester (the treatment group), whereas those in the control group received neither reward nor penalty for completing problem sets.

All the data used in this study come from university records, not from student surveys which have been shown to overstate actual performance. Due to missing SAT scores and high school data, a common set of academic aptitude variables exists for 117 students, of whom 71 were in the experimental group and 46 in the control group. In Table 1 we provide descriptive statistic (means and standard deviations) for each variable used for each group. Mean exam scores and college GPA measures were not equivalent in the experimental and control groups (see Table 1). We have three types of academic aptitude variables: college GPA, SAT scores, and high school measures. According to the correlation matrix presented in Table 2, the collegiate GPAs measures are highly correlated with each other (.72-.96) as are high school GPA and rank (.7) but each of those types is not strongly correlated with variables in the other groups (less than .45). SAT math and verbal scores have a much lower correlation (0.51) with each other and even less with other aptitude variables.

Academic Aptitude Proxies

College entrance examination scores, typically thought to measure raw intelligence, a stock of knowledge, and/or a general aptitude for learning, have the virtue of being uniform and methodical but, when used to control for aptitude in college courses, have the disadvantage of providing a measure at a point in time in the past.[3] We use three SAT scores: SAT math (MSAT), SAT verbal (VSAT), and SAT combined (TSAT).

College grade point average (GPA), an institution of higher education's indicator of academic success, directly measures success in course work at the same college or university. Performance in college coursework also reflects the application, throughout each academic term, of good study skills, motivation, organization, industriousness, perseverance, and

[3] Rothstein (2004) argues that the SAT is a proxy for individual and high school demographic characteristics.

consistency.[4] We use two temporal measures of collegiate GPA: grades earned during the semester under study and cumulative GPA including the semester under study.[5] Same semester GPA, a coincident measure of student academic success during the semester being studied, includes information about positive or negative shocks that may have affected a student's potential scholastic achievement. From a practical perspective, much of the economic education research addresses the principles of economics course even though during the fall term freshman have no prior college cumulative GPA. If the dependent variable in learning studies is the course grade or is highly correlated with it, the appropriate GPA measure would be the concurrent semester's GPA minus the economics grade (SemGPA-ECN) or the cumulative GPA minus the economics grade (CumGPA-ECN).[6]

Academic success as measured by grades, though, is obscured by students' heterogeneous mix of courses and the variability of grade distributions by professors, courses, and departments, since a B+ might represent a low grade in one course but a top score in another. To improve the comparison of between-class grades, we created z-score GPA measures which calculate a student's grade deviation from the distribution mean for

[4] Jencks (1979) demonstrates that, net of background, formal schooling, and cognitive skills, personal traits such as industriousness, perseverance, and leadership have noteworthy associations with earnings and occupational status. With similar controls and housework time, Dunifon, Duncan and Brooks-Gunn (2001) establish that a "clean-home measure" is predictive of own and children's earnings 25 years later and of children's schooling. On this point, note that the only student with a perfect math-SAT score from our full sample, not the sub-sample of 117 students used for this study, failed to hand in the required four problem sets.

[5] Caudill and Gropper (1991) view the prior semester's cumulative GPA as "probably a better measure than the GPA at the time of the course because the former measures the student's performance over a longer time period" (305). In the fall term freshman have no cumulative GPA, whereas sophomores, juniors and seniors have two, four or six previous semesters of grades, respectively, so that for freshman semester and cumulative grades are the same.

[6] For example of scholars who exclude the economics grade from GPA, see Evensky et al (1997) and Chizmar and Ostrosky (1998).

each course.[7] Construction of such "standardized GPA" data requires access to the transcripts of every student enrolled in each course taken by a member of the sample group.

Some researchers, and college admissions committees, have expressed skepticism about using cumulative high school GPA to measure cognitive ability because of long time lags and large variations in the standards and quality of schools, school districts and states (Caperton, 2001; Ferber et al, 1983, 36; Chaker, 2003, D2). Georgetown University and Haverford College, for example, use high school rank, rather than GPA, in their admissions' decisions (Chaker, 2003).

IV. Estimates–Selection of Proxies

We hypothesize that assigning and grading a regular problem set throughout the semester improves student's cognitive achievement, controlling for their academic ability and demographic characteristics.[8] Thus, our baseline estimation equation is:

$$\begin{aligned} MeanExamScore = \alpha_0 + \alpha_1 Freshman + \alpha_2 ProblemSet \\ + \sum_{j=1}^{k} \alpha_j AcademicAptitudeProxy(ies)_j + \in \end{aligned} \tag{2}$$

In Table 3 we provide regression results for five specifications of this model (for full results see Grove et al, 2006). Since the dependent variable is exam performance on a 100 point scale, the coefficient on the graded problem set variable, when controlling for academic ability with zSemGPA-ECN (second column), indicates that students in the treatment group improved their performance by 3.74 points, i.e., by more than a third of a letter grade. The results of the other four specifications of the model in Table 3, each using a different aptitude

---

[7] Z-scores are calculated as the difference between the raw course grade and the sample mean course grade divided by the standard deviation of the course grades. We thank Kevin Rask for this suggestion.

[8] Student learning is typically modeled as a production function with exam performance resulting from student

proxy, reveal estimated learning gains ranging from 3.38 points to 5.5 points. Freshman status is not significant.

Because there is evidence that the proxy choice meaningfully influences the estimated treatment effect (Grove at al, f2006) a natural question to ask is which proxies should a scholar use to control for student academic aptitude? Our strategy is to use principal components, factor analysis, and so called "post hoc" estimation to extract students' underlying academic aptitude from the proxies and then to determine the correlations between the estimated latent ability index and each individual proxy.

Lubotsky and Wittenberg (2006) provide the proxy choice criteria for deciding which academic ability control variables scholars should include in a regression of student learning. They show that a tradeoff exists between how well one estimates the latent variable versus how efficiently one estimates the treatment effect. For the best estimate of the latent variable all proxies should be included even though adding additional proxies that may add little information about the underlying unobserved variable may affect the accuracy of the coefficients on correctly measured variables in the model. Our goal, though, and the objective of student learning studies is to determine how to estimate the treatment effect most accurately. Thus, our proxy search criteria is to include proxies that are most highly correlated with the estimated latent variable but not with each other, to avoid correlated measurement errors (14-15).

Table 4 presents results of regressing each generated aptitude proxy on the mean exam grade. The best estimate is 3.83 from the posthoc method, however, estimates from principal components (3.87) and factor analysis (3.90) are close enough to warrant those

---

human capital inputs, demographic characteristics, student effort, and treatment effects (see Becker, 1997).

methods credibility. By ignoring proxies the effect of the problem set is upward biased by about 2 grade points.

In table 5 we display the correlations between the individual academic aptitude proxies and each of the three generated latent aptitude indexes. No pairs of the three highly-correlated indexes have significantly different correlations with each other. While the use of three techniques serves as a specification check of our analysis, we emphasize the post hoc estimates when differences emerge.

All three methods uniformly find that collegiate GPAs are the most highly correlated and SAT verbal scores are the least correlated with the generated latent aptitude indexes. The post hoc estimator, though, identifies, more definitively than with the other techniques, that college grades have significantly higher correlations with the post hoc index than do all other proxies. The main difference that emerges between the three techniques is that principal components analysis and factor analysis cluster math SAT scores and both high school variables together (as similarly correlated to the latent variable), while the more precise post hoc estimator groups high school rank with the weakest set of control variables, SAT verbal scores. The post hoc estimator, then, identifies three clusters of proxies with similar correlations to the generated ability indexes: (1) collegiate GPA, (2) SAT math scores and high school GPA, and (3) high school rank and SAT verbal scores. In addition, the post hoc results find a more pronounced distinction between each of the three clusters of proxy variables. Fisher z tests of the post hoc results indicate that SAT math and high school grades differ significantly from SAT verbal scores. Similar evidence suggests that SAT math scores (but not high school grades) are significantly better proxies than high school rank.

We present the effect of adding proxies in a specific order in Table 6. We include the variables that prove significantly correlated with the latent aptitude and observe the change in problem set effect on the mean exam score. As expected, more proxy variables let us estimate the effect that is closer to the best estimate from table 4. In fact, with only three GPA-derived variables we can significantly decrease the bias to the problem set effect.

Thus, we find that academic aptitude measures constitute three proxy clusters: collegiate grades are twice as highly correlated with the generated aptitude indexes as are SAT verbal scores and high school rank, while SAT math and high school GPA fall in the middle (Table 5). As Lubotsky and Wittenberg (2006) suggest, the best set of aptitude proxies has the highest correlation with the generated index variables but without being correlated with each other. High correlations exist between each of the college GPA variables and between high school GPA and rank (see Table 2). Thus, our results, from this single-institution data set, indicate that a scholar interested in controlling for academic ability in a study of student learning should include one of the collegiate GPA variables and either SAT math (or SAT sum) or, in the absence of SAT data, high school GPA. Beyond those control variables, high school rank and SAT verbal scores appear to so poorly proxy for academic aptitude that little would be gained by including them. Thus, a reasonable summary of our empirical results is for researchers of undergraduate student learning studies to proxy for academic aptitude with collegiate GPA and SAT math.

V. Estimates–Measuring Bias on the Treatment Coefficient

So far we assumed that the lower coefficient on the problem set variable means that the treatment effect is estimated with less bias and closer to the true value. Thus, including more proxies should give us a better estimate for the treatment effect. However, from table 1 we can see that students sorted themselves into each section based on the GPA-related proxy variables. Inclusion of those proxies into regression creates an endogeneity problem which will affect the treatment effect.

Therefore, following Black and Smith (2006), who showed that the GMM estimation is superior to the OLS when either only proxies or latent variables[9] are included the regression, we use an instrumental variables estimator with SemGPA-ECN assumed as endogenous. We use remaining proxies as instruments and test three groups: CumGPA-derived variables, SAT-related variables, and HS-originated variables. The choice of the instrument set is motivated from high correlations within selected variable groups (see table 2). We exclude zSemGPA because of its very high correlation with endogenous SemGPA and a potential for non-ability related unobserved error being correlated.

We present results of IV estimation in table 7. As expected, the coefficient on GradeProblemSetGroup (GPSG) variable decreases significantly in (3), as opposed to OLS in (1) and (2) experiencing severe positive bias. The specification test for validity of instruments (Sargan / J-statistic) cannot reject only the model (3) when we use as instruments CumGPA-ECN and zCumGPA-ECN. Also in that case we cannot reject the test for endogeneity of these instruments (C-statistic). Estimations (4) and (5) have slight upward bias in GPSG variable and it is reflected in both the Sargan test and C-statistic. Thus we choose (3) as our preferred specification with the true coefficient on the GradeProblemSetGroup around 2.94. (footnote? Note that coefficient on SemGPA-ECN shows the same pattern of bias as the GPSG variable with OLS underestimating the effect to the greatest extent and IV having the highest value).

---

[9] Black and Smith (2006) considers generated proxies from factor analysis and posthoc estimation (Lubotsky and Wittenberg, 2006) to estimate latent college quality.

As expected, relative to other methods, IV estimator is superior and has the lowest bias. However, because posthoc estimator is the closest to our best estimate we believe that the latent variable created by the method is closest to true latent ability. Thus, we have the confidence that the correlations between posthoc latent variable and independent variables are the best approximations to the correlations of independent variables with the true unobserved ability. That conclusion is the basis for our recommendation of which variables should be first controlled for when controlling for ability.

V. Conclusions and Implications

Using a unique data on academic performance in an Principles of Economics course, and rich set of proxy variables for academic aptitude obtained from university records, we study which proxies best control for latent academic ability. Four econometric approaches are employed to identify the best proxies for latent academic aptitude from a large variety of student performance variables: principle components, factor analysis, using the additional proxies as instruments for endogenous GRE-based indicators, and a regression-based post hoc estimator recently proposed by Lubotsky and Wittenberg.

We show that there is significant gain in including multiple proxies into the regression, however, it has to be done with caution. Each included proxy decreases the bias on the treatment effect, however, highly correlated proxies (between themselves) may lower the benefit of more information by decreasing efficiency of the estimates. We find that collegiate GPA best predicts students' ability to learn economics and our estimates uniformly suggest that the existing literature overestimates the treatment effect of faculty pedagogical practices.

In general, in order to best estimate the effect of the problem set on the grade score in the exam, an instrumental variable approach is the most appropriate. On the other hand, in order to best control for the academic aptitude in the regression, a post-hoc method proposed by Lubotsky and Wittenberg is the most useful. We find evidence that as long as proxies' errors are uncorrelated most of them can be included in the regression without significant bias to the treatment effect.

Table 1

Descriptive Statistics:
Means, Standard Deviations, and P-Values for T-tests and Chi-square
(N = 117)

	Means and in Parenthesis Standard Deviation		
Dependent and Independent Variables	**Graded Group (N=71)**	**Non-Graded Group (N=46)**	**P-values** [1]
Mean Exam Score	83.0 (8.12)	77.0 (8.94)	0.000*
SemGPA	3.12 (0.57)	2.78 (0.64)	0.003*
SemGPA-ECN	3.14 (0.64)	2.79 (0.78)	0.009*
zSemGPA-ECN	0.20 (0.63)	-0.10 (0.84)	0.023*
CumGPA	3.21 (0.49)	3.05 (0.47)	0.087
CumGPA-ECN	3.18 (0.55)	2.99 (0.53)	0.056
zCumGPA-ECN	0.23 (0.55)	0.04 (0.64)	0.091
MSAT	590.1 (81.4)	579.8 (70.2)	0.480
VSAT	567.2 (81.5)	559.3 (66.1)	0.586
TSAT	1157.3 (143.5)	1139.1 (114.1)	0.470
HS%	75.3 (14.90)	72.0 (14.63)	0.238
HSGPA	3.40 (0.43)	3.35 (0.37)	0.570
White	0.79 (0.41)	0.80 (0.40)	0.838[2]
Male	0.63 (0.49)	0.65 (0.48)	0.840[2]
Freshman	0.35 (0.48)	0.22 (0.42)	0.102[2]

[1] P-values for continuous variables from t-tests and for dichotomous variables from chi-square.
[2] Chi-square analysis.
* Means differ at the 5-percent level of significance.

Table 2

Correlation Matrix (N-117)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. MeanExamScore														
2. Male	.18													
3. White	.15	.06												
4. Freshman	.16	.14	-.04											
5. SemGPA	.76	.10	.19	.20										
6. SemGPA-ECN	.65	.11	.27	.18	.96									
7. z-SemGPA-ECN	.66	.09	.16	.16	.92	.91								
8. CumGPA	.60	.01	.24	.13	.76	.75	.72							
9. CumGPA-ECN	.60	.05	.24	.19	.86	.87	.84	.94						
10. z-CumGPA-ECN	.65	.08	.19	.15	.86	.86	.92	.88	.94					
11. TSAT	.45	.22	.29	.20	.40	.36	.33	.48	.43	.42				
12. MSAT	.48	.33	.30	.17	.40	.38	.34	.44	.41	.40	.87			
13. VSAT	.30	.05	.19	.18	.29	.24	.24	.39	.33	.32	.86	.51		
14. HS%	.35	-.01	.13	.09	.35	.34	.34	.38	.35	.38	.41	.44	.28	
15. HSGPA	.44	.08	.14	.15	.36	.34	.37	.36	.34	.39	.39	.43	.24	.70

Table 3

Estimated Learning Gains for the Student Learning Model with the Four Most Common Individual Aptitude Proxies

Dependent Variable: Mean Exam Score (in points); N =117

Independent Variables	SemGPA-ECN	zSemGPA-ECN	CumGPA-ECN	SATsum	HSGPA
GradedProblemSetGroup	3.38* (2.60)	3.74** (2.96)	4.22** (3.16)	5.47** (3.77)	5.53** (3.80)
Academic Aptitude Proxy (see column heading)	7.57** (8.40)	7.50** (8.85)	9.16** (7.57)	0.03** (5.29)	9.07** (5.18)
Freshman	0.40 (0.29)	0.52 (0.70)	0.36 (0.25)	0.54 (0.34)	1.01 (0.64)
Adjusted R^2	0.45	0.47	0.4	0.28	0.27

Note: t-statistics are in parentheses.
* Mean is different from zero at the 5-percent level of significance.
** Mean is different from zero at the 1-percent level of significance.

Table 4

Estimated Learning Gains for the Student Learning Model with the Four Most Common Individual Aptitude Proxies

Dependent Variable: Mean Exam Score (in points); N =117

Independent Variables	No Apptitude proxy	Principal Components Proxy	Factor Analysis Proxy	Posthoc proxy
GradedProblemSetGroup	5.93** (1.57)	3.87** (1.17)	3.90** (1.17)	**3.83**** (1.13)
Academic Aptitude Proxy (see column heading)	-	2.87** (0.29)	6.07** (0.61)	9.79** (0.92)
Freshman	1.81 (1.70)	-0.47 (1.26)	-0.26 (1.26)	-0.66 (1.22)
Adjusted R^2	0.15	0.55	0.54	0.57

Note: t-statistics are in parentheses.
* Mean is different from zero at the 5-percent level of significance.
** Mean is different from zero at the 1-percent level of significance.

Table 5

Correlations Between Academic Aptitude Variables and Index Variables Created from Three Multivariate Estimation Methods
(N =117)

	Multivariate Estimation Methods		
	Principal Components	**Factor Analysis**	**Posthoc Estimator**
zCumGPA-ECN	.87	.88	.89
zSemGPA-ECN	.81	.82	.89
SemGPA-ECN	.81	.81	.87
CumGPA-ECN	.84	.84	.83
MSAT	.71	.72	.65
TSAT	.75	.77	.62
HSGPA	.61	.55	.61
HS%	.61	.56	.48
VSAT	.59	.61	.43
Principal Components	-	.992	.972
Factor Analysis	-	-	.965

Note: See text for details.

Table 6

Estimated Learning Gains for the Student Learning Model with the Four Most Common Individual Aptitude Proxies

Dependent Variable: Mean Exam Score (in points); N =117

Independent Variables	No Apptitude proxy	SemGPA-ECN	SemGPA-ECN, zSemGPA-ECN	SemGPA-ECN, zSemGPA-ECN, CumGPA-ECN	SemGPA-ECN, zSemGPA-ECN, CumGPA-ECN, MSAT	SemGPA-ECN, zSemGPA-ECN, CumGPA-ECN, MSAT, HSGPA
GradedProblemSetGroup	5.93**	3.49**	3.69**	3.76**	3.75**	3.81**
	(1.57)	(1.30)	(1.29)	(1.24)	(1.20)	(1.18)
Freshman	1.81	0.06	0.12	0.07	-0.32	-0.50
	(1.70)	(1.38	(1.35)	(1.31)	(1.27)	(1.26)
Adjusted R^2	0.15	0.45	0.48	0.51	0.54	0.055

Note: t-statistics are in parentheses.
* Mean is different from zero at the 5-percent level of significance.
** Mean is different from zero at the 1-percent level of significance.

Table 7

Estimated Learning Gains for the Student Learning Model
Using Instrumental Variables Estimation (SemGPA-ECN assumed endogenous)

Dependent Variable: Mean Exam Score (in points); N =117

Independent Variables	OLS without any proxies SemGPA-ECN (1)	OLS with SemGPA-ECN (2)	IV with endogenous SemGPA-ECN (3)	IV with endogenous SemGPA-ECN (4)	IV with endogenous SemGPA-ECN (5)
GradedProblemSetGroup	5.93** (1.1.58)	3.49** (1.30)	2.94* (1.30)	2.95* (1.30)	2.96* (1.30)
Freshman	1.81 (1.70)	0.06 (1.38)	-0.34 (1.37)	-0.33 (1.36)	-0.32 (1.37)
SemGPA-ECN	-	7.48** (0.93)	9.20** (1.10)	9.16** (1.09)	9.11** (1.09)
Instruments	-	-	CumGPA-ECN, zCumGPA-ECN	zSemGPA-ECN, CumGPA-ECN, SATsum, VSAT	zSemGPA-ECN, CumGPA-ECN, SATsum, VSAT, HS%, HSGPA
Sargan test (p-value)	-	-	0.206	0.023	0.009
Instruments tested for exogeneity	-	-	CumGPA-ECN, zCumGPA-ECN	SATsum, VSAT	HS%, HSGPA
C statistic (p-value)	-	-	0.216	0.072	0.055
Partial F-statistic from first-stage regression	-	-	139.18	69.48	45.54

Note 1: standard errors are in parentheses.
Note 2: C statistic test performed in IV regression using as instruments: zSemGPA-ECN, CumGPA-ECN, zCumGPA-ECN, SATsum, VSAT, HS%, HSGPA
* Mean is different from zero at the 5-percent level of significance.
** Mean is different from zero at the 1-percent level of significance.

REFERENCES

Ballard, Charles L. and Marianne F. Johnson (2004), "Basic Math Skills and Performance in An Introductory Economics Class." *Journal of Economic Education*, 35(1), Winter.

Bauer, Thomas and Klaus F. Zimmermann. 1998. "Learning Efficiency of Economics Students," Discussion Paper No. 23, IZA (Institute for the Study of Labor), Bonn, Germany.

Becker, W. E. (1997): "Teaching Economics to Undergraduates," *Journal of Economic Literature*, 35(3), 1347-1373.

Black, Dan A; Smith, Jeffrey A (2006). Estimating the Returns to College Quality with Multiple Proxies for Quality, *Journal of Labor Economics*, 24(3)(July): 701-28.

Boivin, Jean and Serena Ng (2003) "Are More Data Always Better for Factor Analysis?" NBER Working Paper 9829, July 2003.

Caperton, G. (2001), "Toward a holistic definition of merit: a response to the UC proposal," College Board News.

Chaker, Anne Marie (2003), "Why Colleges Scoff at Your Kid's GPA," *Wall Street Journal*, July, 24, D1-2.

Chizmar, John F. and Anthony L. Ostrosky (1998), "The One-Minute Paper: Some Empirical Findings," *Journal of Economic Education*, 29(1), (Winter), 1-8.

Dewald, William G., Jerry G. Thursby and Richard G. Anderson (1986), "Replication in Empirical Economics: *The Journal of Money, Credit and Banking* Project," *American Economic Review*, 76(4), (Sept.), William G. 587-603.

Dunifon, Rachael, Greg J. Duncan and Jeanne Brooks-Gunn. 2001. "As Ye Sweep, So Shall Ye Reap," *American Economic Review*, 91(2), 150-154.

Evensky, Jerry, Duke Kao, Richard Fenner, Qing Yang, and Remi Falele. 1997. "Addressing Prerequisite Math Needs—A Case Study in Introductory Economics." *International Journal of Mathematical Education in Science and Technology*. 28(5): 629-639.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). "Evaluating the use of exploratory factor analysis in psychological research." *Psychological Methods*, 3, 272-299.

Forni, M. and L. Reichlin (1998), "Let's Get Real: a Factor-Analytic Approach to Disaggregated Business Cycle Dynamics," *Review of Economics Studies* 65, 453-473.

Grove, Wayne and Tim Wasserman. 2006. Incentives and Student Learning: Results from a Natural Experiment. *American Economic Review* 96:2 (May), 447-452.

__________ (2004), "The Life-Cycle Pattern of Collegiate GPA: Longitudinal Cohort Analysis and Grade Inflation," *Journal of Economic Education*, 35:2, 162-174.

Grove, Wayne, Tim Wasserman and Andrew Grodner. 2006. Choosing a Proxy for Academic Aptitude. *Journal of Economic Education* 37(Spring): 131-147.

Jencks, Christopher. 1979. *Who Gets Ahead?* New York: Basic Book.

Jensen, Elizabeth J., and Ann L. Owen. 2000. "Why Are Women Such Reluctant Economists? Evidence from Liberal Arts Colleges," *American Economic Review*, 90(2): 466-470.

Johnson Richard A. and Dean W. Wichern. 1992. *Applied Multivariate Statistical Analysis*, Englewood Cliffs, New Jersey: Prentice Hall.

Kennedy, P. and J. Siegfried (1997), "Class size and achievement in introductory economics," *Economics of Education Review*, 16(4), 385-394.

Lubotsky, Darren and Martin Wittenberg (2006) "Interpretation of Regressions with Multiple Proxies," *Review of Economics and Statistics*. 88(3) (August), 549-562.

McCloskey, Deirdre N., and Stephen T. Ziliak. 1996. "The Standard Error of Regressions," *American Economic Review*. 34:1, 97-114.

Rothstein, Jesse M. 2004, "College performance predications and the SAT," *Journal of Econometrics*. 121: 1-2 (July-August): 297-317.

Sabot, Richard and John Wakeman-Linn (1991), "Grade Inflation and Course Selection," *Journal of Economic Perspectives*, 5:1, 159-170.

Siegfried, J.J. and W.B. Walstad (1998), "Research on Teaching College Economics," in *The Principles of Economics Course: A Handbook for Instructors*, by William B. Walstad and Philip Saunders (McGrawHill), 141-166.

Stock, J.H. and M.W. Watson (2002), "Diffusion Indexes," *Journal of American Statistical Association* 97:460, 1167-1179.