ARTICLE IN PRESS

Available online at www.sciencedirect.com

SCIENCE DIRECT

Journal of Econometrics ▌▌▌ (▌▌▌▌) ▌▌▌–▌▌▌

JOURNAL OF Econometrics

www.elsevier.com/locate/econbase

The econometrics of higher education: editor's view

Lawrence C. Marsh*

Department of Economics and Econometrics, University of Notre Dame, 436 Flanner Hall, Notre Dame, IN 46556, USA

Abstract

This special issue examines the role of econometrics in controlling for unobserved factors that could bias our understanding of key issues in higher education. The usual identification assumption of linear selection on observables is frequently found lacking. The authors show their determination and ingenuity in devising various strategies to avoid the inconsistency and, generally, to reduce the bias typically forthcoming in such cases. These researchers have made significant progress in overcoming these obstacles.
© 2003 Elsevier B.V. All rights reserved.

JEL classification: C14; C51; I21; J31

Keywords: Cognitive ability; Earnings; Biased sample selection; Omitted variable bias

1. Introduction

Ehrenberg provides both the background and context for this volume by reviewing the literature in the economics of higher education as it relates to the contributions in this special issue. He systematically reviews progress by economists in recent years on the following four key issues in higher education: (1) rates of return to higher education, (2) academic labor market, (3) institutional behavior, and (4) higher education as an industry. He notes the key econometric problems of sample selection in studies using individual-level data and disentangling demand and supply shocks when using market-level data. While most of the articles in this issue use individual data, some make use of data at city, county and/or state levels to make across-country comparisons

* Tel.: +1-574-287-0458 or +1-574-631-6210; fax: +1-574-287-2173.
E-mail addresses: lawrence.c.marsh.1@nd.edu, lmarsh@nd.edu (L.C. Marsh).

0304-4076/$ - see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2003.10.007

to understand certain aspects of the market for college graduates and their implications for public policy.

Many of these papers grapple with unobserved heterogeneity often manifested as omitted-variable bias and/or sample selection bias where endogeneity effects distort key behavioral relationships. These problems could potentially result in econometric estimates that misinform analysts who, in turn, could produce misleading policy recommendations. The sample selection issue is particularly important in several of these papers. For example, Rothstein predicts freshman grade point average but wants to generalize to the universe of SAT-takers. His unique strategy is discussed below. The Stinebrickners want to understand how study time affects grade point average but have to devise a method to correct for measurement error because study time data were only collected on certain days. In some contexts sample selection translates into the more generic problem of making the most effective use of sometimes rather limited common support when trying to infer counterfactual outcomes. For example, Black and Smith work out a method for attributing the earnings differential associated with attending a high-quality rather than a low-quality college when ability levels are unevenly represented across the college quality spectrum. Hansen, Heckman and Mullen sort out the relationship between ability and schooling as both these variables affect test scores, and, ultimately, a person's wage rate. Several of the other papers also focus on unobserved or partially observed ability differentials. The Cappelli paper argues that tuition reimbursement programs help employers detect and retain workers with higher than average unobserved ability. The Arcidiacono paper deals with unobserved heterogeneity in both preferences and ability and separates out acquired ability that transfers across college majors from acquired ability that is specific to a particular major.

2. Using semiparametric and nonparametric methods to account for ability

Hansen, Heckman and Mullen develop both nonparametric and semiparametric models of the effect of schooling on test scores when both schooling and test scores depend on latent ability. They estimate models using NLSY data for four mutually exclusive groups: (1) high school dropouts, (2) high school graduates, (3) persons with some college, and (4) college graduates. The model is estimated using the Bayesian MCMC method to determine the joint posterior distribution of test scores and schooling. They analyze directly a model of tests at different schooling levels where $T(s)$ is the test score at schooling level s, f is latent ability,

$$T(s) = \mu(s) + \lambda(s)f + u(s) \tag{2.1}$$

and the causal effect of schooling is

$$\frac{\partial T(s)}{\partial s} = \frac{\partial \mu(s)}{\partial s} + \frac{\partial \lambda(s)}{\partial s} f, \tag{2.2}$$

where $\lambda(s)$ is the effect of schooling on revealing or transforming latent ability f and $\mu(s)$ is the effect of schooling that is constant across latent ability levels. Since people who were older when they started school on average tend to have lower cognitive ability but higher test scores than their classmates, Hansen, Heckman and Mullen also take into account the possible selection bias arising from whether a person started school at least by age six (normal-ahead) or after age six (late-behind). They use the method of control functions previously developed by Heckman (1976, 1980) and Heckman and Robb (1985, 1986) to identify schooling effects on tests and extend it to a new nonparametric version. They establish nonparametric new identification results for control functions. They compare control function estimates with a choice-theoretic semiparametric approach, which takes into account ceiling effects using the truncated mixtures of normal distributions.

They find that schooling raises measured achievement more in high school than in college, but latent ability is more important in affecting college choices. In some cases a one standard deviation increase in latent ability can raise test scores by as much as 20 percent. They conclude that high school dropouts and high school graduates do not differ in latent ability, but that latent ability is an important factor in explaining some college and college graduation outcomes. They also note that the verbal components of the AFQT score show a decreasing marginal effect of latent ability as schooling increases while the mathematics components show a slightly increasing marginal effect. Previous estimates including those reported in *The Bell Curve* by Herrnstein and Murray (1994) overestimated the effect of latent ability on years of schooling because they do not account for the effects of schooling on test scores. They also underestimate the effect of schooling on test scores. Hansen, Heckman and Mullen found that the biggest schooling effects on test scores occur for students with very low ability. They find that the marginal effects of schooling on test scores declines with increases in ability. However, schooling only slightly equalizes test scores.

Another part of their analysis analyzes log wages as a function of schooling, S, and test scores, T, with an appropriate error term:

$$\ln W = \alpha_0 + \alpha_1 S + \alpha_2 T + \varepsilon. \tag{2.3}$$

Here they explicitly incorporate the effect of latent ability through its influence on test scores as follows:

$$\frac{\partial \ln W}{\partial s} = \alpha_1 + \alpha_2 \frac{\partial T}{\partial s} f. \tag{2.4}$$

Conventional methods of "controlling" for ability underestimate the effect of schooling on wages because they ignore the effect of schooling on raising ability.

Black and Smith want to determine the earnings effects of attending a high-quality rather than a low-quality college for those who do so. They note two problems not adequately addressed by models whose identification depends on the linear-selection-on-observables assumption. Such models dominate the literature. First, the data sets commonly used in this literature may suffer from common support problems, due to a lack of individuals who attend low-quality colleges but whose characteristics resemble those of students at high-quality colleges. Such support problems suggest that the data may not justify precise conclusions regarding the counterfactual effects. Linear models cover up this problem by relying on the linear structure to project beyond regions of common support. Second, linearity may not adequately characterize the selection behavior, especially when combined with relatively parsimonious model specifications. Together, these problems may lead to bias in regression-based estimates and to a false precision that comes from a strong reliance on the linear functional form.

Black and Smith address both of these problems using the propensity score matching methods originated in Rosenbaum and Rubin (1983) and further developed in Heckman et al. (1997, 1998a) and Heckman et al. (1998b). The propensity score in the probability of treatment conditional on observed covariates; in this case the treatment represents attending a high-quality college rather than a low-quality one. College quality consists of a quality index equal to the first principal component of three different observable measures of quality: mean SAT scores of the entering class, mean faculty salaries and the freshman retention rate. Propensity score matching relies on the so-called conditional independence assumption (CIA). In this context, the conditional independence assumption requires that the decision to attend a high- or low-quality school is unrelated to the labor market outcome associated with attending a low-quality school, conditional on a set of observed covariates. Black and Smith argue that the CIA is plausible in their data set due to the rich set of available observable characteristics related to both college quality choice and labor market outcomes such as test scores, high school characteristics, parental characteristics, home environment characteristics and so on. Dehejia and Wahba (1999, 2002) and Smith and Todd (2004) provide recent applications and discussions of propensity score matching methods. Dearden et al. (2002), and Heckman and Vytlacil (2001) undertake related analyses in the context of estimating the labor market effects of high school characteristics and years of schooling.

Black and Smith examine the common support problem using their estimated propensity scores. They find that their data—National Longitudinal Survey of Youth 1979 cohort—contains only a handful of individuals who attend a low-quality college but have a high probability of attending a high-quality college. Put differently, they find that the support condition only weakly holds. As expected, this results in matching estimates having much larger estimated (bootstrap) standard errors than linear regression estimates that condition on the same covariates.

The general form of the expected counterfactual for treated observation "i" in the context of matching is given by

$$\hat{E}(Y_0 \mid \hat{P}(X_i)) = \sum_{j=1}^{J} w(\hat{P}(X_i), \hat{P}(X_j))Y_{0j} \tag{2.5}$$

where Y_0 denotes earnings in the low college quality (untreated) state, $\hat{P}(X)$ is the estimated propensity score, X is the vector of observed conditioning variables and w is the weight assigned to each untreated observation "j" in constructing the counterfactual for treated observation "i".

Matching estimators differ in the way in which they construct the weighting function, although in all cases it depends on the distance between the estimated propensity score of the treated observation whose counterfactual is being constructed and the surrounding untreated observations. Black and Smith consider three matching estimators: single nearest-neighbor matching with replacement, Gaussian kernel matching with no caliper and Epanechnkov kernel matching. The latter has the feature that it imposes the common support condition because it does not assign positive weights beyond a certain distance that depends on the bandwidth. As a result, treated observations with no untreated observation with a propensity score within that distance get dropped from the analysis. Using leave-one-out cross-validation, Black and Smith find that Epanechnikov kernel matching outperforms the other two methods in terms of its mean squared error. They also use leave-one-out cross-validation to pick the optimal bandwidth. See, for example, Racine and Li (2004) for a general discussion of leave-one-out cross-validation methods.

Black and Smith find similar linear regression and matching estimates of the effect on wages of attending a high-quality college for men, but find that the matching estimates exceed the linear regression estimates for women. Moreover, for both men and women, they find that the matching estimates of the effect of attending a high-quality college are larger in the "thick support" region defined by $0.33 < \hat{P}(X) < 0.67$. While this could be evidence of heterogeneous treatment effects, Black and Smith argue that theoretical concerns render this explanation implausible. More likely, in their view, this finding results from measurement error in the reported college attended or lingering selection on unobservables, both of which should play a large role for values of the estimated propensity score outside the thick support region. This finding suggests that standard estimates may understate the labor market effect of college quality. Finally, Black and Smith find evidence that a parsimonious specification selected using cross-validation methods provides larger estimated effects with much smaller standard errors, indicating a potential trade-off in finite samples between the plausibility of the conditional independence assumption and the variance of the resulting estimates.

3. Applications, admissions and attendance

Jesse Rothstein's article decomposes the predictive power of the SAT in determining college performance. He notes that socioeconomic status of both applicants and their high schools tend to be positively correlated with their SAT scores. Admissions decisions do not generally award points directly to applicants for high socio-economic status (SES), and so-called "validity studies" rarely take account of SES. Consequently, SAT scores' predictive coefficients may partially reflect the influence of SES, and may be biased indications of the SAT's ability to measure preparedness. In addition, Rothstein confronts the problem of biased sample selection in that his population of interest is potential college applicants while his primary sample consists of students with a college freshman grade point average. Rothstein devises a new omitted variables estimator to deal with this problem. In particular, Rothstein considers the equation:

$$\mathrm{E}(y|X) = \mathrm{E}[\mathrm{E}(y|X,S)|X] = \alpha_1 + X\beta_1 + \mathrm{E}(S|X)\gamma_1, \tag{3.1}$$

where S is the applicant's SAT scores and X is a set of other variables characterizing the applicant. This produces consistent estimates only when $\mathrm{E}(S|X) = \mathrm{E}(S)$ which is generally not the case. To correct for this inconsistency Rothstein uses the linear projection

$$\mathrm{E}(S|X) \approx \mathrm{E}(S) + \rho_{SX}\,\mathrm{sqrt}[\mathrm{var}(S)/\mathrm{var}(X)](X - \mathrm{E}(X)), \tag{3.2}$$

where ρ_{SX} is the correlation between S and X while var(.) is variance and sqrt[.] is square root. By substituting in this expression for $\mathrm{E}(S|X)$ Rothstein, in effect, replaces α_1 with $\alpha_1 + \{\mathrm{E}(S) - \mathrm{E}(X)\rho_{SX}\,\mathrm{sqrt}[\mathrm{var}(S)/\mathrm{var}(X)]\}\gamma_1$ and β_1 with $\beta_1 + \rho_{SX}\,\mathrm{sqrt}[\mathrm{var}(S)/\mathrm{var}(X)]\gamma_1$ and provides the corrected conditional variance:

$$\mathrm{var}(y\,|\,X) = \sigma^2 + \gamma_1^2(1 - \rho_{SX}^2)\,\mathrm{var}(S). \tag{3.3}$$

Rothstein obtains a consistent estimate of $\rho_{SX}\,\mathrm{sqrt}[\mathrm{var}(S)/\mathrm{var}(X)]$ by regression S on X using data on the universe of SAT-takers. In a manner somewhat similar to that of Heckman (1990) and Altonji et al. (2000), he uses a trimmed sample to check for endogenous matriculation bias and instrumental variables to correct for heterogeneous grading patterns on different campuses. The results do not change much so little endogenous bias is evident. Rothstein's analysis shows that SAT scores are proxies for socio-economic background. He finds that the SAT does have some independent predictive power for freshman grade point average but that failure to control for demographic variables overstates its role. He concludes that admissions officials seeking SES-blind admissions rules may wish to downplay the importance of SAT scores, while those intent on maximizing performance, as measured by freshman grade point average, may wish to include socio-economic variables that more directly predict that performance. In either case, SAT scores warrant less emphasis than is currently the practice at most schools.

Bridget Long first models the individual's choice among colleges conditional on college attendance and then models the decision to attend college. In addressing the choice among colleges, she modifies McFadden's (1973) conditional logit model to explain which of over 2500 colleges was chosen. Long includes individual-specific measures of institutional characteristics and interaction terms such as $\beta_4(X_{i4}Y_{j4})$ where, Y_{j4} is specific to the jth college, X_{i4} is individual-specific but common across colleges, and the coefficient β_4 is common across colleges and individuals. In comparing 1972 results with those for 1992, she found that college quality, measured as higher expenditures per student, was a positive inducement to attend a particular college only for high-income students in 1972 but had become an inducement for all students by 1992. Also, she found that while tuition price and distance to college still negatively affected the probability of enrollment at any particular college in 1992, they were less important than in 1972. However, low-income students were even more adversely affected by tuition price in their relative probability of choosing a particular college in 1992 than in 1972.

In addressing the issue of whether to attend any college at all, Long restricts her analysis to individuals that are either low income, have low-SAT or have parents without a college education. With this group she used a binary logit model to explain attendance using the college characteristics of either the closest 2-year public college or the college with the highest probability predicted by her conditional logistic college choice model discussed above. She found that both tuition price and distance to college had negative effects on attendance in 1972 but not in 1992. For the college with the highest predicted probability, she found that a 1 percent rise in the local unemployment rate increased attendance by 6 percent in 1972 and 9 percent in 1992. These results suggest that costs alone no longer explain differences in college access but the opportunity cost of job market alternatives is important.

Mark Long examines the impact of the elimination of affirmative action in college admissions. He uses the SAT score reports sent to six types of colleges to determine whether eliminating affirmative action influenced the number of score reports sent by high school students to each type of college.

His theoretical model expresses each student's expected indirect utility as the expected indirect utility arising from the outcome of the admission process net of the disutility cost associated with completing numerous applications. In general, in choosing between J college types ordered from highest quality to lowest quality with corresponding indirect utilities $V_1, \ldots, V_J$, respectively, and V_{J+1} as the indirect utility associated with not attending college, the expected indirect utility is

$$\mathrm{E}[V(N_1, \ldots, N_J)] = V_1(1 - \pi_1^{N_1}) + \sum_{j=2}^{J}\left[V_j(1 - \pi_j^{N_j})\left(\prod_{k=1}^{j-1} \pi_k^{N_k}\right)\right] + V_{J+1}\left(\prod_{j=1}^{J} \pi_j^{N_j}\right) - \alpha N, \tag{3.4}$$

where π_j is the probability of not being accepted by a college of the jth college type, while N_j is the number of such colleges to which SAT scores were sent, N is the sum of the N_j's and α is the disutility cost of each application. Long finds the optimal values for N_1 through N_J that maximize his expression for indirect utility.

To empirically estimate the effects of changes in affirmative action policy, Long stacks the count data models $N_j = \exp(x\beta_j) + \upsilon_j$ for $j = 1, \ldots, J$ into the $JT \times 1$ vector $u(\beta)$ with covariance matrix Ω where T is the number of student applicants. Long estimates his model using the semiparametric generalized least-squares method of Delgado (1992). First he applies nonlinear least squares to each of these J equations separately to get $\tilde{\beta}$ as a consistent estimate of β. He then minimizes the following quadratic form with respect to β:

$$T^{-1}\sum_{t=1}^{T} [u_t(\beta)]' \hat{\Omega}_t^{-1} [u_t(\beta)] \text{ using } \hat{\Omega}_t = \sum_{s=1}^{T} [u_s(\tilde{\beta})][u_s(\tilde{\beta})]' W_s(X_t), \tag{3.5}$$

where $W_s(X_t)$ equals $1/k$ for the k observations X_s nearest to X_t and equals zero for all other observations, and k is the integer closest to the square root of T. The variables in Long's X matrix include the student's academic characteristics, personal and family characteristics, college and state characteristics, and a specially designed probability of acceptance predicted by probit analysis. Long uses National Education Longitudinal Study (NELS) data in the probit analysis to estimate the direct effect of eliminating affirmative action in admission decisions on the probability of acceptance for underrepresented minorities (URM).

Not surprisingly the predicted probabilities of acceptances at the various types of colleges are highly correlated with one another. He then uses his specially-obtained SAT data set to estimate the number of SAT score reports (i.e. N_j) sent to each of the $J = 6$ college types using the Delgado (1992) estimation method described above. Unfortunately, the high degree of correlation among the predicted acceptance rates for the different college types prevented him from including the predicted probabilities for college types other than the one for which the number of score reports is being predicted in that equation. He finds that a lower probability of admission at type J would lower the number of score reports sent by URMs to that type. Further, since the acceptance probabilities have high positive correlations, the omitted variables suggests a downward bias in the negative coefficient estimates which, in turn, implies that the true effects may be even more strongly negative than the ones obtained by Long.

Thus, his analysis demonstrates a strong indirect effect on URM enrollments through the significant decline in the number of their applications, especially in California's top-quality public universities. Long concludes that the indirect application effects of eliminating affirmative action may be significantly greater than the immediate direct admission effects of such policy changes.

The graduate admissions article by Marsh and Zellner formalizes the decision process often faced by admissions officials in MA, MBA and Ph.D. programs. It provides a general approach for selection decisions where acceptance or rejection of offers of admission has a random aspect. For example, a dean might decide on how many offers of admission to make based on an unspecified, but implicit, loss function and unknown acceptance probabilities. Marsh and Zellner show that the optimal number of offers falls as the variance associated with the probabilities of acceptances rises. They allow for asymmetric loss with a linex loss function of the form

$$L(\tilde{N},\hat{N}) = b[\mathrm{e}^{a(\hat{N}-\tilde{N})} - a(\hat{N}-\tilde{N}) - 1], \tag{3.6}$$

where $\tilde{N}$ is the actual (random) number of offers accepted, $\hat{N}$ is a point prediction for that number, b adjusts the units of measurement, and a controls the degree of asymmetry. They show that the larger the value of a, the more asymmetric the loss function, and then the smaller will be the optimal number of acceptances, $\hat{N}$. In particular, for the linex loss function the optimal point forecast can be based on heterogeneous acceptance probabilities which are obtained following Zellner and Rossi (1984) by estimating the posterior distribution for $\underline{\beta}$ given by

$$h(\underline{\beta}|D) = f(\underline{\beta}|I)\prod_{i=1}^{q} F(\underline{z_i'}\underline{\beta})^{x_i}[1 - F(\underline{z_i'}\underline{\beta})]^{1-x_i}, \tag{3.7}$$

where $f(\underline{\beta}|D)$ is a prior distribution for $\underline{\beta}$ and $F(\underline{z_i'}\underline{\beta})$ is a transformation function (e.g. normal or logistic cdf) based on a vector of the applicant's characteristics, z_i. Using this methodology Marsh and Zellner determine the optimal number of offers of admission given the characteristics of a given set of applicants.

4. Student choices during college

Stinebrickner and Stinebrickner tackle an interesting measurement error problem associated with missing data in their study of the effect of study time on grade point average:

$$O_i = \alpha_s\,\mathrm{STUDY}_i + \alpha_X X_i + \varepsilon_i, \tag{4.1}$$

where O_i is the freshman grade point average(GPA), STUDY_i is the average study time, X_i is a set of other factors that influence GPA, and ε_i is an error with zero mean and variance σ_ε^2, all for the ith student. Their unique data set provides student self-reported data on study hours during the last 24 h, the last 7 days, and the last weekend day, among other things. Defining T as the total possible number of study days and $s_{i,t}$ as the number of hours of study by person i on day t, they expressed the average amount of studying by person i as

$$\text{STUDY}_i = \frac{1}{T}\sum_{t}^{T} s_{i,t}, \quad \text{where } s_{i,t} = C + \beta X_i + \mu_i + v_{i,t} \tag{4.2}$$

with μ_i representing permanent, unobservable aspects of the average study time of the ith person and $v_{i,t}$ representing transitory shocks that are mean zero and have variance σ_v^2. However, they recorded the number of hours of study only on N_i days so their measured average amount of studying for person i was:

$$\text{ST}\hat{\text{U}}\text{DY} = \frac{1}{N_i}\sum_{i=1}^{N_i} s_{i,t}. \tag{4.3}$$

Measurement error was generated by the fact that $N_i < T$. However, Stinebrickner and Stinebrickner derive an unbiased estimator of the coefficient of study hours on grade point average as

$$\hat{\alpha}_{\text{s, measurement}} = \hat{\alpha}_{\text{s, OLS}} \frac{\text{Var}(\text{ST}\hat{\text{U}}\text{DY})}{\text{Var}(\text{ST}\hat{\text{U}}\text{DY}) - \text{Var}((1/N)\sum_{i=1}^{N} v_{i,t})}. \tag{4.4}$$

Using this relationship in their model they produced a coefficient estimate for the effect of study time on GPA of 0.160. To take into account missing data, they also assume normality for ε_i and $v_{i,t}$ and formulate a maximum likelihood approach to this problem. Assuming that conditional on μ_i and X_i the variables $s_{i,1}, \ldots, s_{i,Ni}$, O_i are independent, they specify the likelihood of the ith person as

$$L_i = \int g_1(s_{i,1}|\mu_i, X_i)\cdots g_1(s_{i,N_i}|\mu_i, X_i) g_2(O_i|\mu_i, X_i) h(\mu_i)\,\mathrm{d}\mu_i \tag{4.5}$$

and estimate by simulation by maximizing:

$$L = \prod_i \frac{1}{D}\sum_{d=1}^{D} g_1(s_{i,1}|\mu_i^d, X_i)\cdots g_1(s_{i,N_i}|\mu_i^d, X_i) g_2(O_i|\mu_i^d, X_i), \tag{4.6}$$

where μ_i^d is the dth draw from the distribution $h(\mu_i)$. Using this MLE approach Stinebrickner and Stinebrickner found the effect of an hour of study time on GPA to be 0.171. Using a multiple imputation approach similar to that in Brownstone and Valletta (1996), they adjust their estimation for reporting error $m_{i,t}$ to produce the likelihood function:

$$L = \prod_i \frac{1}{D}\sum_{d=1}^{D} g_1(s_{i,1} - m_{i,1}^d|\mu_i^d, X_i)\cdots g_1(s_{i,N_i} - m_{i,N_i}^d|\mu_i^d, X_i) g_2(O_i|\mu_i^d, X_i), \tag{4.7}$$

where $m_{i,1}^d, \ldots, m_{i,Ni}^d$ is the dth draw from the distribution of $m_{i,1}, \ldots, m_{i,Ni}$. This results in a reporting-error-corrected estimate of their study time coefficient of 0.221. Either way they found that study time has a positive and significant association with GPA.

Peter Arcidiacono looks at how ability sorting and labor market earnings relate to choice of various post-secondary schooling combinations. His dynamic model separates the influence of monetary from nonmonetary rewards in both the choice of college and the choice of major. His three-period model shows how learning about ability through grades helps students sort themselves into appropriate majors.

In the first two periods, individuals choose between various school-major combinations, recognizing the option values associated with the first period choice. Learning then occurs between the periods before the final school-major combination is chosen.

He breaks learned ability into two-parts: $A_{uk} = \eta_1 + \eta_{2k}$, where η_1 is learned ability that transfers across majors and η_{2k} is learned ability that is specific to the kth major where the subscript "u" means that the ability is unobserved and is learned about through the realization of grades.

Arcidiacono estimates and simulates his model both with and without allowing for the unobserved heterogeneity in preferences and abilities. The unobserved heterogeneity follows a finite mixture distribution along the lines of Keane and Wolpin (1997) and Heckman and Singer (1984). The dynamic discrete model is estimated along the lines of Rust (1987). Conditional on the unobserved type (the mixture distribution), errors are independent across preferences, grades and earnings. Estimates of the model show that ability sorting across majors is not due to differences in monetary returns but instead appears to be a function of individual preferences for particular majors quite apart from differences in labor market returns. He also found that math ability was significant while verbal ability was insignificant in influencing both choice of college major and eventual monetary returns. Choice of college major was found to have a substantially greater impact than college quality on labor market earnings.

There are times when it may not be possible to obtain directly even a roughly adequate approximation for some important factor that is correlated with other key factors. In sorting out the effect of the presence of an undergraduate teacher certification program (UTCP) on the probability of becoming a teacher using the Baccalaureate and Beyond (B& B) data set, Randall Reback confronts the problem of controlling for the unobserved pre-college desire for a teaching career. Reback taps econometric research by Yatchew and Griliches (1985) and his own ingenuity to estimate the bias on the coefficient of the UTCP dummy variable in predicting the decision to teach. Using P^*_{ij} as the latent probability of the ith graduate of the jth college taking a teaching job, Reback uses probit analysis to estimate the model

$$P^*_{ij} = \cdots + \beta_3 X^*_{3i} + \beta_4 X_{4i} + \cdots + \varepsilon_{ij}, \tag{4.8}$$

where X^*_{3i} is the ith graduates pre-college desire to teach and X_{4i} is a dummy variable for the presence of a UTCP. Reback then expresses the unobserved pre-college desire to teach as a linear function of the other variables in his model including the UTCP variable:

$$X^*_{3i} = \cdots + \gamma_4 X_{4i} + \cdots + u_{ij}. \tag{4.9}$$

Substituting this into his first equation produces:

$$P^*_{ij} = \cdots + (\beta_4 + \beta_3\gamma_4) X_{4i} + \cdots + \beta_3 u_{ij} + \varepsilon_{ij}. \tag{4.10}$$

The composite residual, $\beta_3 u_{ij} + \varepsilon_{ij}$, has a mean of zero and a variance of $\beta_3^2\sigma_u^2 + \sigma_\varepsilon^2$. Therefore, leaving out the unobserved X_{3i}^* from the teaching probability model results in a biased and inconsistent estimator for β_4.

Since an estimator generally goes somewhere asymptotically, it might be said that every estimator consistently estimates something, just not always the right something. In this case the estimator of the coefficient of X_{4i} is consistently estimating $(\beta_4 + \beta_3\gamma_4)/\sqrt{\beta_3^2 + \sigma_u^2 + \sigma_\varepsilon^2}$ instead of β_4. Reback's strategy to correct for this bias is to estimate γ_4 in an auxiliary equation using a separate data set that contains information related to X_{3i}^* but does not contain all of the other variables needed for the complete model. In particular Reback creates a dummy variable, X_{3i} , indicating a high school senior's expressed interest in a teaching career in a probit model containing the UTCP explanatory variable, X_{4i}, in order to estimate its coefficient, γ_4. Reback imposes the normalizations $\sigma_u^2 = 1$ and $\sigma_\varepsilon^2 = 1$ as is standard in probit analysis.

Although primarily using the B& B data, Reback makes use of the National Education Longitudinal Study (NELS) to correct for (or at least substantially mitigate) the bias caused by the omission of the pre-existing preference to teach, X_{3i}^*. The NELS data contain an indication of whether or not high school seniors have a desire to teach, and, 2 years later, whether those same individuals still have a desire to teach when asked in their sophomore year in college. Using the latter responses as an actual decision to teach, Reback estimates the predictive power of their high school response in generating their sophomore response, which serves as an upper bound approximation for β_3, which, in turn, allows a lower bound calculation for β_4. Reback concludes that creating an UTCP at an elite college or university could increase the flow of undergraduates into public school teaching by at least 50 percent.

5. The market for college educated labor

Peter Cappelli asks why do employers provide tuition assistance to their employees? Given human capital, h, Cappelli depicts the value of the marginal contribution of the low (L) ability worker as $F(L,h) = f(h)$ while that of the high (H) ability worker is $F(H,h) = H + f(h)$. Since initially a high ability worker's high ability is not observed, both types of workers are paid the same wage, $W(h) = f(h)$. Tuition reimbursement attracts high ability workers to the firm by self-selection and helps retain them at least until their eligibility period (N years) and education (M years) are completed. If p is the per unit price of human capital, then the self-selected high-ability workers increase their human capital by T/p where T is the fixed tuition cost. By completing their education these workers reveal their higher ability, and their wage then increases to

$$W(H, h + T/p) = H + f(h + T/p). \tag{5.1}$$

In effect the tuition assistance has compensated them for their higher ability during the $N + M$ years when their high ability was unobserved by the market. In his empirical analysis Cappelli addresses the key concern of inconsistency due to biased sample selection by employing a two-step maximum likelihood procedure proposed by Magee et al. (1998). The traditional weighted least squares estimator for situations where subpopulations appear in the sample in different proportions than they exist in the population has a rather large variance when the weights are wide ranging. In the Census Bureau's National Employer Survey II, the weights vary by industry and size of establishment with the latter ranging from 20 to over 5000 employees. Hausman tests are performed to show the inconsistency of the unweighted least-squares procedure in this context. The two-step procedure for this biased sample selection problem produces a consistent estimator with a substantially reduced variance relative to the weighted least squares estimator. Cappelli regresses education on tuition assistance and a host of control variables including the average education of new employees and finds a positive significant relationship between tuition assistance and employee education levels. In a separate regression he finds that the education level of new hires is also positively associated with offers of tuition assistance. Another regression shows that wage rates are a positive function of tuition assistance, education levels, work hours, all benefits, sales and a union workforce. Conversely, a probit of tuition assistance on many of these same variables plus the residuals from the previous wage regression yields a positive and significant coefficient for the wage residuals, thus confirming the distinct relationship between wages and tuition assistance. In particular it undermines the alternative theory that wage rates are held down in order to pay for the tuition assistance and is consistent with Cappelli's theory that tuition assistance is associated with at least partially unobserved higher ability, and, therefore, higher productivity which, in turn, helps pay for both tuition assistance and the somewhat higher wages. Finally, Cappelli performs a Tobit analysis on turnover, defined as the percentage of the permanent workforce that left for either voluntary or involuntary reasons. The Tobit analysis revealed a significant negative relationship between turnover and tuition assistance, education levels, and pay levels while controlling for a host of other variables. Given high turnover costs this last analysis reinforced Cappelli's argument that employers benefit both directly and indirectly from offering and providing tuition assistance.

John Bound et al. investigate how degree production in higher education is related to the eventual concentration of university-educated workers in a state. The theoretical motivation is based on the relationship between relative wages (w), the flow or production of new college degrees (f), the stock or proportion of the state's population with college degrees (s), and the demand (d) for college degree workers. Bound et al. summarize these relationships in the three equations:

$$(1)\ f = \gamma w + \xi, \quad (2)\ s = \theta w + f, \quad (3)\ d = -\sigma w + \zeta, \tag{5.2}$$

where γ is the supply elasticity, θ is the migration elasticity, σ is the elasticity of substitution between labor with a college degree and labor with only a high school degree, ξ is an exogenous shock to flow, and ζ is an exogenous shock to demand. The primary parameter of interest links stock and flow in response to exogenous changes in degree production. In the context of the above model, the ratio of $\partial s/\partial \xi$ to $\partial f/\partial \xi$ is equal to $\sigma/(\sigma + \theta)$ and interpreted as the elasticity of stocks with respect to flows. An implication of the model is that degree types concentrated in the production of goods and services that are not traded across states will face a lower elasticity of demand in the labor market and, thus, a smaller elasticity between stock and flows in the context of the above model. At the BA level, Bound et al. estimate the long-term elasticity between stocks and flows to be no greater than 0.3. At the MD degree level, there is little relationship between where physicians are trained and where they practice. Since MDs produce services that are generally produced and consumed, it is no surprise that their stock-flow coefficient estimates were not statistically significant. A state with a medical school that increases its output of MDs cannot expect a significant number of those additional MDs to stay in that state after graduation. An important empirical challenge in this research is to distinguish exogenous variation in the output of the higher education sector from responses of colleges and universities to local labor market conditions. Evidence on the relationship between the relative wages of college-educated workers are not necessarily the states with a comparative advantage in the use of college-educated labor. In the end, state policy in the higher education sector is predicted to have a limited effect on the long-term human capital levels in the state work force and the magnitude of these effects are predicted to vary with the degree to which local demand for university-educated workers adjusts to changes in supply.

Jeffrey Groen develops a discrete-choice model at the individual worker level that explains how going to college in a particular state might affect the decision to work in that state. Starting with the simple model:

$$Y_{ij} = s_j + \beta_1 H_{ij} + \beta_2 C_{ij} + \varepsilon_{ij}, \tag{5.3}$$

where s_j is a fixed effect controlling generally for the attractiveness (or lack thereof) of each state, while Y_{ij}, H_{ij} and C_{ij} equal one when the ith individual works in the jth state, grew up in the jth state, and went to college in the jth state, respectively, and equal zero otherwise. He notes the inadequacy of least squares in this context which (among other things) produces inconsistent estimates of β_2, where the extent of this inconsistency is commonly expressed as

$$\text{plim}(\beta_2) = \beta_2 + \text{Cov}(C_{ij}, \varepsilon_{ij})/\text{Var}(C_{ij}). \tag{5.4}$$

The bias represented by the second term is positive since a student is likely to have an a priori preference for a state before attending college in that state. The bias may be so large as to inflate the estimate of β_2 to produce a value big enough to imply that a policy designed to increase college attendance in the state might be cost effective in retaining educated workers. However, Groen provides a strategy to reduce the bias and get a more accurate assessment of the effect of college attendance on worker retention. In particular, Groen introduces A_{ij}, which equals one if the ith individual applied to at least one college in state j and equals zero otherwise. Essentially, the application information accounts for differences in location preferences before students started college. Adding this variable results in a new specification:

$$Y_{ij} = s_j + \gamma_1 H_{ij} + \gamma_2 C_{ij} + \gamma_3 (H_{ij} C_{ij}) + \gamma_4 A_{ij} + \gamma_5 (H_{ij} A_{ij}) + v_{ij}, \tag{5.5}$$

where $\gamma_2 + \gamma_3$ is the impact of attending college in one's home state while γ_2 is the impact of attending college in another state. Groen then incorporates this into a conditional logit model and calculates the corresponding college effects using conditional means of predicted probabilities for the four home-state/college-state combinations. He concludes that the resulting impact of college attendance on the probability of working in a state is insufficient to justify policies that help pay college tuition as a way of attempting to induce a substantial number of college-educated workers to stay in the state.

Enrico Moretti examines the extent of any positive externalities or spillovers that might be obtained in the form of increased wages from increasing the proportion of college graduates in a city. In the presence of human capital externalities, increases in the number of educated workers in a city may increase productivity and earnings over and above the private return to education. Moretti defines the external return to education as the derivative of the log of averages wages with respect to the number of college-educated workers, minus the private return. Identification of this external return is complicated by two factors. First, increases in college share may result in increases in average wages above the private return to education even in the absence of spillover effects, if high and low education workers are imperfect substitutes. Second, the presence of individual-level or city-level unobservable factors that are correlated with college share and wages across cities may bias simple OLS estimates. Using longitudinal and cross-sectional data from the NLSY and the Census of Population, Moretti accounts for unobserved ability of individuals and unobserved variation in the return to skills across cities by controlling for the interaction of individual and city fixed effects. To allow for the possibility that city-specific time-varying demand shocks might account for the relationship between college share and wages, he uses two instrumental variables that generate arguably exogenous shifts in the supply of college educated workers across cities. The first instrument is based on differences in the age structure of cities. The U.S. labor force is characterized by a long-run trend of increasing education, with younger cohorts better educated than older ones. To the extent that the relative population shares of different cohorts vary across cities, this will lead to differential trends in college share across cities. The second instrument is based on the presence of a land-grant college. To account for the possibility of imperfect substitution between high and low education workers, Moretti estimates the effect of changes in the fraction of highly educated workers on wages of different education groups. Standard demand and supply considerations suggest that the effect of an increase in college share should be positive for low education groups and that for college graduates its sign should depend on the size of the spillover. Moretti's findings indicate that a one percentage increase in the proportion of college graduates can be expected to raise the wages of high school drop-outs by 1.9 percent, high school graduates by 1.6 percent and college graduates by 0.4 percent. The increase in the wages of college graduates is especially interesting since an increase in the local supply of college educated workers should drive their wages downward.

6. Conclusions

Ehrenberg provides an important contribution to this special issue by placing it within the broader context of the evolving literature on the economics of higher education. He relates the research provided by each of these papers with a broad array of past and ongoing research. In fact his review covers 176 other papers. By providing an extensive discussion of the progress achieved in the study of higher education to date (especially regarding the findings of the econometric studies on higher education), he brings into focus the importance of each of the papers in this special issue.

An important issue in education generally is determining the role of cognitive ability in influencing educational outcomes. Latent cognitive ability is shown by Hansen, Heckman and Mullen to be a key factor as it relates to years of schooling and test scores. Their model provides a more complete picture of how ability works both directly, and indirectly through its effect on schooling, to influence test scores. On average they find that a one standard deviation increase in latent ability can increase test scores by 12–17 percent of the maximum possible score. This effect is stronger for those with low ability where the increase can be around 30 percent. However, they find that cognitive ability does not have as strong a relationship to dropping out of school as claimed by the previous literature. Black and Smith are also concerned with ability and the extent to which it accounts for the earnings differential associated with attending a high-quality instead of a low-quality college. Using the predicted propensity to attend a high-quality college from a logit model, they find a 12–14 percent increase in wages for men and a 6.7 to 7.8 percent increase in wages for women associated with attending a higher-quality rather than a lower-quality college. In the middle third of the propensity score range (the region of "thick" support), they find an even stronger effect of 20–25 percent for men and 12.4–15.7 percent for women.

Rothstein, Mark Long, Bridget Long, and Marsh and Zellner all produce models of college admissions. Marsh and Zellner focus on graduate admissions from the admission officer's point of view while the others address undergraduate admissions with an eye towards understanding the plight of underrepresented minorities. Marsh and Zellner show how different types of loss functions (e.g. symmetric versus asymmetric) with different cost constraints can significantly alter the optimal number of offers of admission. In particular, they find that an increase in the variance of acceptance probabilities will reduce the optimal number of offers. Mark Long finds that a reduced probability of acceptance for underrepresented minorities can result in a significant reduction in the number of SAT score reports they send to high-quality colleges. His simulations demonstrate that policies that guarantee admission to students in the top x-percent of their high school class do little to restore the number of URM applications especially at the top-quality colleges. The implications of these results are reinforced by those of Bridget Long who finds that while tuition price is no longer a significant constraint to attending college, it does have a substantial effect on the quality of the college attended. Lacking adequate independent financial resources, low-income applicants appear to be significantly constrained in reaching their fullest academic potential. Rothstein clarifies the role of SAT scores in explaining freshman grade point average. Among other things, finds that uncorrected SAT regression coefficients, such as those reported in traditional SAT validation studies, are significantly biased in favor of applicants with favorable socio-economic background characteristics. After correcting for biased sample selection and omitted-variables bias, he finds that SATs have a much smaller effect on first year performance. The overall message from these last three studies is clear. Equality of opportunity in higher education is still a goal that is far from fulfilled.

College attendance clearly constitutes one set of decisions, while decisions made during college involve another set. The Arcidiacono, Reback, and Todd and Ralph Stinebrickner papers primarily focus on what to study, or (in the Stinebrickners' paper) how much to study in college. In explaining grade point average as a function of study time, the population sampled by the Stinebrickners is the freshman class at Berea College. After controlling for sex, race, family income, parental education and American College Test (ACT) scores and correcting for measurement error, they find that study time does indeed have a significant impact on a student's grade point average. On the other hand, Reback looks at how the availability of teacher certification programs might increase participation of students at selective colleges. After dealing with heterogeneous entry costs and controlling for omitted variable bias, Reback found that allowing teacher certification within the standard 4 years of college at selective colleges could potentially increase the number of teachers coming from such colleges by at least 50 percent. Arcidiacono takes an even broader look at student decisions including whether to go to college, where to go to college, what to major in, whether to switch majors or drop out, and how much emphasis to put on math versus verbal skills. Some of Arcidiacono's conclusions include (1) math is more important than verbal skills for both success in school, particularly in the more demanding majors, and for success in the labor market after graduation, (2) college major choice is more important than quality of school in terms of monetary return with natural science and business offering the best returns, and (3) both preferences in school and expected job market returns play a role in college major selection, but preferences in school turns out to be the more important determinant of college major. The obvious conclusion from these studies is that students play a central role in their own fate. This would suggest that they might benefit from reading these research papers to get more precise information about their choices, including information about the consequences of where they go to college, what they major in, and to what extent more study time might translate into higher grades.

The Cappelli, Groen, Moretti and Bound et al. papers focus on the effects of education (primarily college education) on labor market behavior. The econometric evidence provided by Cappelli is consistent with the claim that workers who self select to use tuition assistance are more productive, and that their employers pay for the tuition assistance from that (otherwise unobserved and uncompensated) higher productivity. He also finds that the availability of this type of benefit is associated with lower employee turnover rates. Bound et al. consider the retention of college-educated labor from a state's point of view. They find a rather modest elasticity of no more than 0.3 between the production of college graduates in a state and their retention by that state. For medical degree recipients (MDs) they find that production per se has no disproportional effect on retention. Groen's analysis at the individual student-worker level finds only a modest relationship between college location and work location after graduation. He suggests that other considerations beyond economics (e.g. politics) may provide a better explanation for the popularity of state-supported merit scholarships. However, using a partial equilibrium approach, Moretti concludes that an increase in a city's share of college graduates pays off in increased wages especially for the least educated members of society. For example, a 1 percent increase in college share raises the wages of high school dropouts by almost 2 percent. Thus, Moretti establishes the benefits to a city of increasing its proportion of college graduates while Groen and Bound et al. demonstrate the difficulties in effectively altering that proportion. Similarly, Cappelli notes the benefits to a firm of attracting and retaining more productive employees but suggests the exhaustion of any potential excess profits in equilibrium through payments for the tuition assistance and the widespread adoption of such tuition assistance plans. Once again, this suggests that Adam Smith was right. States and firms striving to improve themselves only end up making society as a whole better off while retaining little, if any, excess benefit for themselves.

Acknowledgements

I greatly appreciate the comments and suggestions of referees and authors on this overview and analysis of our special issue. Since each of these papers had as many as four referees, there were often alternative perspectives that had to be reconciled before moving ahead. I appreciate the patience of all of the authors and the fifty-four referees with this process. I am especially grateful to Arnold Zellner for this opportunity and for his inspiration and guidance over the years. His counsel on all matters econometric and statistical has been of great value. There were several scholars whose leadership in the economics of higher education has contributed directly and indirectly to this special issue. The leadership of Charles T. Clotfelter, Caroline M. Hoxby, John J. Siegfried and, of course, Ronald G. Ehrenberg has been especially noteworthy in this regard. Clotfelter's suggestions and recommendations for referees were particularly helpful. The NBER seminars on higher education have provided especially fruitful forums for recruiting authors and referees. I am grateful to the many participants who helped in various ways with this effort.

References

Altonji, J., Elder, T., Taber, C., 2000. Selection on observed and unobserved variables: assessing the effectiveness of catholic schools. NBER Working Paper 7831.

Brownstone, D., Valletta, R., 1996. Modeling earnings measurement error: a multiple imputation approach. Review of Economics and Statistics 78, 705–717.

Dearden, L., Ferri, J., Meghir, C., 2002. The effect of school quality on educational attainment and wages. Review of Economics and Statistics 84, 1–20.

Dehejia, R., Wahba, S., 1999. Causal effects in non-experimental studies: re-evaluations of training programs. Journal of the American Statistical Association 94 (448), 1053–1062.

Dehejia, R., Wahba, S., 2002. Propensity score matching methods for nonexperimental causal studies. Review of Economics and Statistics 84, 151–161.

Delgado, M., 1992. Semiparametric generalized least squares in the multivariate nonlinear regression model. Econometric Theory 8, 203–222.

Heckman, J., 1976. Simultaneous equation models with both continuous and discrete endogenous variables with and without structural shift in the equations. Studies in Nonlinear Estimation, Ballinger, Lexington, MA.

Heckman, J., 1980. Addendum to sample selection bias as a specification error. In: Evaluation Studies Review Annual, Vol. 5. Sage, Beverley Hills, CA.

Heckman, J., 1990. Varieties of selection bias. American Economic Review 80, 313–318.
Heckman, J., Robb, R., 1985. Alternative methods for evaluating the impact of interventions: an overview. Journal of Econometrics 30 (1–2), 239–267.
Heckman, J., Robb, R., 1986. Alternative methods for solving the problem of selection bias in evaluating the impact of treatments on outcomes. In: Wainer, H. (Ed.), Drawing Inferences from Self-Selected Samples. Lawrence Erlbaum Associates, New Jersey (reprinted 2000).
Heckman, J., Singer, B., 1984. A method for minimizing the impact of distributional assumptions in econometric models of duration data. Econometrica 52 (2), 271–320.
Heckman, J., Vytlacil, E., 2001. Identifying the role of cognitive ability in explaining the level and change in the return to schooling. Review of Economics and Statistics 83, 1–12.
Heckman, J., Ichimura, H., Todd, P., 1997. Matching as an econometric evaluation estimator: evidence from evaluating a job training programme. Review of Economic Studies 64, 605–654.
Heckman, J., Ichimura, H., Todd, P., 1998a. Matching as an econometric evaluation estimator. Review of Economic Studies 65, 261–294.
Heckman, J., Ichimura, H., Smith, J., Todd, P., 1998b. Characterizing selection bias using experimental data. Econometrica 66, 1017–1098.
Herrnstein, R., Murray, C., 1994. The Bell Curve. The Free Press, New York.
Keane, M., Wolpin, K., 1997. The selection and estimation of discrete choice dynamic programming models by simulation and interpolation: Monte Carlo evidence. Review of Economics and Statistics 76 (4), 648–672.
Magee, L., Robb, A., Burbridge, J., 1998. On the use of sampling weights when estimating regression models with survey data. Journal of Econometrics 84, 251–271.
McFadden, D., 1973. Conditional logit analysis of qualitative choice behavior. In: Zarembka, P. (Ed.), Frontiers in Econometrics. Academic Press, New York.
Racine, J., Li, Q., 2004. Nonparametric estimation of regression functions with both categorical and continuous. Journal of Econometrics, forthcoming.
Rosenbaum, P., Rubin, D., 1983. The central role of the propensity score in observational studies for causal effects. Biometrika 70 (1), 41–55.
Rust, J., 1987. Optimal replacement of GMC bus engines: an empirical model of Harold Zurcher. Econometrica 55 (5), 999–1033.
Smith, J., Todd, P., 2004. Does matching overcome LaLonde's critique of nonexperimental methods? Journal of Econometrics, forthcoming.
Yatchew, A., Griliches, Z., 1985. Specification error in probit models. Review of Economics and Statistics 67, 134–137.
Zellner, A., Rossi, P., 1984. Bayesian analysis of dichotomous quantal response models. Journal of Econometrics 25, 365–393.