Research in Higher Education, Vol. 45, No. 1, February 2004 (© 2004)

ACCOUNTING FOR MODEL UNCERTAINTY IN THE PREDICTION OF UNIVERSITY GRADUATION RATES

Cullen F. Goenner*,** and Sean M. Snaith*

Empirical analysis requires researchers to choose which variables to use as controls in their models. Theory should dictate this choice, yet often in social science there are several theories that may suggest the inclusion or exclusion of certain variables as controls. The result of this is that researchers may use different variables in their models and come to disparate conclusions with respect to predicted effects and their statistical significance. In such cases one is uncertain of which particular set of regressors forms the model that represents the data. The approach used below accounts for uncertainty in variable selection by using Bayesian model averaging (BMA). Accounting for uncertainty, we demonstrate that BMA provides better out-of-sample prediction for university graduation rates than results based on alternative variable selection methods.

KEY WORDS: graduation rates; prediction; Bayesian model averaging; variable selection.

INTRODUCTION

Graduation rates are an increasingly important measure of institutional success in an era in which students, media, legislators, and administrators expect greater accountability for educational outcomes. The problem is how to accurately assess institutional performance. For instance, simply comparing graduation rates of two institutions fails to account for the differences in mission, cost, and resources available to each institution. One would expect that institutions with more financial resources and better prepared students will have more positive student outcomes as demonstrated by higher student retention and graduation rates. Controlling for these outside factors is thus crucial to assessing performance across institutions.

A common method that is used by both *U.S. News* and Mortenson (1997) to evaluate an institution's contribution to student outcomes is to compare an

*Department of Economics, University of North Dakota, Grand Forks, ND.

**Address correspondence to: Cullen F. Goenner, Department of Economics, 290 Gamble Hall, University of North Dakota, Grand Forks, ND 58202. E-mail: cullen.goenner@und.nodak.edu

institution's actual graduation rate with that predicted by a model that controls for the characteristics of the student body and institution. Institutions that do better than their predicted rate are viewed as adding value to student outcomes; those which achieve rates less than predicted are viewed as inefficient. Several states, including Virginia, North Carolina, and New York, have examined linking this measure of performance to the funding provided to state institutions in order to reward efficiency. In 1999, the Virginia State Council of Higher Education adopted that funds be set aside to reward institutions with graduation rates that are better than what is statistically predicted based on entering student characteristics.

The difference between the actual and predicted graduation rates is a testament not only to the contribution of the institution to education but also to the accuracy of the model that is used to make the prediction. A poorly specified model that ignores important student and institutional characteristics will result in large errors that can be mistakenly associated with either poor or excellent performance by educational institutions. The accurate estimation of individual coefficients is also important to institutions in that factors that the institution controls (student to faculty ratio, expenditures per student, etc.) can be modified to improve educational outcomes. Thus, to properly evaluate institutional performance one must be able to make accurate predictions.

The problem, as Porter (2000) notes in his analysis of the robustness of the predictions made by *U.S. News*, is that predicted graduation rates vary based on the variables included in the model used to make the prediction. The model used by *U.S. News* to predict 6-year graduation rates for Doctoral I universities controls for each institution's average Scholastic Aptitude Test (SAT) score and the logarithm of average expenditures per student. Estimating this model specification, Porter finds that the coefficient and standard error for SAT score are .105 and .008, respectively, suggesting a positive and statistically significant relationship. The coefficient and standard error for spending were 3.556 and 2.029, respectively, suggesting a positive yet statistically insignificant relationship. By adding to *U.S. News*' model specification control variables that measure student quality and institutional constraints, Porter displays that variable selection influences predicted effects of coefficients as well as their standard errors. In column four of Porter's Table 1, he reports that the effect of SAT is reduced to .084, and the effect of spending is increased to 5.623. Further, the coefficient for spending is now statistically significant. As Porter points out, inclusion of these additional variables affects predicted graduation rates and creates questions over their reliability and usefulness in policymaking.

Researchers and policymakers are left uncertain of which model and its prediction are those that represent the factors that causally determine graduation rates, that is, the "true" model. This is particularly difficult given that research into retention and graduation rates suggests that several factors at the individual and institutional level influence student outcomes. With no unique theory to guide their choice, different researchers often choose different variables from among a theoretically interesting list of candidates to include in their model specification. The choice of this subset of variables may be based on the focus of the article or on the differing theoretical importance various researchers put on these variables.

An alternative is to use statistical methods to select from the candidate variables a subset to use as controls. Draper and Smith (1981), Weisberg (1985), and Miller (1990) discuss several statistical methods of variable selection in applied analysis. Although the selection method varies among researchers, frequently, candidate variables are screened by removing those with small t statistics. More formalized methods choose the variables from the list of candidates that optimize some prediction criterion, such as maximizing the adjusted R^2 or minimizing Mallow's C_p. Other methods rather than regressing over all subsets of models, use stepwise methods of variable selection. Efroymson's stepwise regression starts from an empty subset and adds at each iteration the variable that gives the largest reduction in the residual sum of squares, while accounting for partial correlations to see if any variables in the subset should be dropped. Olejnik, Mills, and Keselman (2000) review these three methods in the context of applied research in education. These methods and their variants are often computed in standard statistical packages, so it is quite easy for the researcher to use any of these methods to select variables. A problem is that none of these methods have any theoretical foundation. Therefore, researchers are left with an arbitrary choice between methods, which may lead to different methods being used, resulting in different variables being selected and hence different predictions.

With respect to predicting graduation rates, it is evident that model specification influences predictions and the estimated effects of coefficients. Without knowing which model specification is the true model that causally explains graduation rates, it is inherently risky to base inference on the predictions from any single model. To account for this uncertainty in variable selection we apply Bayesian model averaging (BMA) to the prediction of graduation rates at Doctoral I universities. Rather than basing predictions on a single model, as is the case with standard variable selection methods, BMA determines estimated effects by taking a weighted average of estimates over models whose specification is supported by the data. The advantage of this method is that it provides a neutral method of variable selection that incorporates the uncertainty in this selection in a manner that improves the accuracy of out-of-sample predictions.

BAYESIAN MODEL AVERAGING (BMA)

BMA is a relatively new method developed by Bayesian statisticians (Draper, 1995; Raftery, 1995) to account for uncertainty in model specification. This method has been applied in several fields including political science (Bartels, 1997), sociology (Raftery, 1995), finance (Avramov, 2002), environmental science (Lamon and Clyde, 2000), and medical science (Volinsky, Madigan, Raftery, and Kronmal, 1997). The basis for this method lies in the work of Leamer (1978) in which he points out that "ambiguity over the model should dilute the information about regression coefficients, since part of the evidence is spent to specify the model" (p. 91). The solution to this problem, Leamer believed, could be found by using a Bayesian perspective. Leamer writes that a "Bayesian approach is sufficiently flexible that, with suitable alterations, specification searches can be made legitimate, or at least understandable" (p. 2). With Leamer's framework and advances in computing, Raftery among others devised the alterations, discussed below, necessary to implement BMA.

A Bayesian perspective provides a natural way of dealing with uncertainty in that unknown parameters of interest, such as regression coefficients, are expressed in terms of probability. Rather than generating a single point estimate of the coefficients using classical means, Bayesian analysis generates the entire probability distribution of the coefficients. This allows one to examine the extent, in terms of probability, to which the data support one model relative to another versus simply viewing a model as either better or worse than an alternative. The basis of Bayesian methods is Bayes theorem:

$$P(\beta/D) = \frac{P(D/B)P(B)}{P(D)} \tag{1}$$

Bayesian analysis requires the researcher to specify their beliefs and uncertainty in the parameters of interest β prior to observing the data D. These beliefs are captured in the prior probability density $P(\beta)$. $P(D/\beta)$ is known as the likelihood and represents the probability of observing the data D given that the parameter estimates β are true. $P(D)$ is the unconditional probability of observing the data whether β is true or not. Observing data D, we update our beliefs about the distribution of the parameter estimates using Bayes theorem to obtain the posterior distribution $P(\beta/D)$. For a further introduction to Bayesian econometrics see Zelner (1971) and Leamer (1978).

In the situation in which several models $\{M_1 \ldots M_K\}$ are theoretically possible, it is risky to base inference on the point estimates from a single model M_k. Bayesian model averaging allows us to account for this type of uncertainty. Hoeting, Madigan, Raftery, and Volinsky (1999) provide an excellent tutorial of these methods. To estimate the effect of a parameter in the presence of model uncertainty, one calculates the posterior distribution of the parameter given the data as:

$$P(\beta/D) = \sum_{k=1}^{K} P(\beta/M_k, D)P(M_k/D) \tag{2}$$

The posterior distribution $P(\beta/D)$ is a weighted average of the posterior distribution under each of the K models, with weight equal to the posterior model probabilities $P(M_k/D)$. The posterior model probability (*PMP*) represents the probability that model M_k is the true model that causally explains the data when conditioning on the data and assuming that one of the K models is the true model. By Bayes' rule and the law of total probability, the posterior model probability is

$$P(M_k/D) = \frac{P(D/M_k)P(M_k)}{\sum_{l=1}^{K} P(D/M_l)P(M_l)} \tag{3}$$

where $P(D/M_k)$ is the likelihood and $P(M_k)$ is the prior probability that model M_k is the true model, given one of the K models is the true model. If a noninformative prior is assumed in which each of the K models are equally likely to be the true model ($P(M_1) = \ldots P(M_k) = 1/K$), then the posterior model probability becomes:

$$P(M_k/D) = \frac{P(D/M_k)}{\sum_{l=1}^{K} P(D/M_l)} \tag{4}$$

The integrated likelihood is given by

$$P(D/M_k) = \int P(D/\beta_k, M_k)P(\beta_k/M_k)d\beta_k \tag{5}$$

where β_k is a vector of parameters (coefficients and variance), $P(D/\beta_k, M_k)$ is the likelihood and $P(\beta_k/M_k)$ is the prior density of the parameters under model M_k. Using the Laplace method for integrals, Raftery (1995) shows that the integrated likelihood of model k is approximately equal to $\exp(-\frac{1}{2}\ \mathrm{BIC}_k)$ where BIC_k is the Bayesian information criterion (BIC) of model k. Schwarz (1978) shows that the BIC is

$$BIC_k = -2\log(\hat{L}) + d_k\log(N) \tag{6}$$

with $\hat{L}$ equal to the maximized likelihood under model k, d_k is the number of parameters in model k, and M is the sample size. The second term penalizes more complex models. Using the approximation of $P(D/M_k) = \exp(-\frac{1}{2}\ \mathrm{BIC}_k)$ and the prior assumption that models are equally likely, the posterior model probability (PMP) becomes:

$$P(M_k/D) \approx \frac{\exp\left(-\frac{1}{2} BIC_k\right)}{\sum_{l=1}^{K} \exp\left(-\frac{1}{2} BIC_l\right)} \tag{7}$$

Once the posterior distribution has been determined, one can summarize the effects of the parameters on the dependent variable by calculating the posterior mean, posterior variance, and posterior effect probabilities. Raftery (1995) reports the posterior mean and variance can be approximated by

$$E(\beta_1/D,\beta_1 \neq 0) \approx \sum_{A1} \hat{\beta}_1(k)^2 P(M_k/D)$$

$$Var(\beta_1/D,\beta_1 \neq 0) \approx \sum_{A1} [Var(k) + \beta_1(k)^2] P(M_k/D) - E(\beta_1/D,\beta_1 \neq 0)^2 \tag{8}$$

where $\hat{\beta}_1(k)$ and Var(k) are the maximum likelihood estimates and variance of β_1 under model k, and the summation is over models that include β_1 (set A_1). The posterior effect probability measures the probability that a particular parameter is part of the true model. It is the sum of the posterior model probabilities for models that include β_1.

$$P(\beta_1 \neq 0/D) = \sum_{A1} P(M_k/D) \tag{9}$$

To implement BMA one must specify the universe of models to average over, where a model refers to a particular set of regressors. Here it is assumed that we have n candidate variables to include in our regression, of which we are unsure of the combination that forms the true model. Thus, there are 2^n different models that are possible and make up the set of models to consider. With 24 regressors, the summation in Eq. (2) would be over more than 16 million models and involve calculating the integrals implicit to Eq. (2). Hoeting et al. (1999) outline two ways in which to manage the summation. The first, which is used in the analysis below, discards models that are not supported by the data. The second method, which is discussed by Madigan and York (1995), uses Markov chain Monte Carlo model composition to approximate Eq. (2).

Madigan and Raftery (1994) argued that models not supported by the data should not be included in Eq. (1) and appeal to what they refer to as Occam's window to discard models. The first restriction of Occam's window is to exclude models that predict the data sufficiently less than predictions of the best model, where predictions are based on the posterior model probability (PMP) of each model $P(M_k/D)$. As discussed earlier, the PMP is the posterior probability that model M_k is the true causal model given the data and is approximated using the BIC as specified in Eq. (7). Models in set A' are included to be averaged over

$$A' = \left\{ M_k : \frac{\max\ PMP_l}{PMP_k} \leq C \right\} \tag{10}$$

where C is a cutoff chosen by the researcher. The cutoff used in the analysis below is 20, which Raftery (1995) discusses as providing "strong" evidence in favor of one model over another. Set A' includes models with PMPs that are at least $1/20^{th}$ of the model that is highest. A second, optional, restriction may include the removal of complex models that receive less support than simpler models that are subsets. If a model within set A' is contained in another model and the simpler model has higher posterior model probability, then the more complex model is excluded. This method of excluding models Hoeting et al. (1999) report often reduces the number of models to average over to fewer than 10.

To apply BMA to the data below we use the S-plus function *bicreg* developed by Raftery and Volinsky (1996). Hoeting et al. (1999) provide a discussion of where *bicreg* and other S-plus programs to implement BMA are publicly available on the Internet. *Bicreg* calculates for linear regression models the posterior mean, variance, and effect probabilities as well as reports the PMPs of the models averaged over. The function uses the BIC of each model to approximate the PMP of each. It then applies the restrictions of Occam's window to specify a reduced set of models that are supported by the data. The results below use only the first restriction of Occam's window that models with PMPs significantly less than that of the best model be eliminated. The decision to use only the first restriction is based on Raftery's (1995) suggestion that this restriction alone provides better prediction, whereas both restrictions together are more useful in reporting uncertainty.

Once the dependent variable and the n candidate independent variables that form the 2^n models are chosen, the researcher must specify for each model a prior probability that the model considered is the true model. This prior should reflect the researcher's beliefs prior to examining the data. In some instances researchers may have strong *a priori* information that suggests the inclusion of one variable or another, in which case the prior probability should reflect these beliefs. Often though there is little information about the relative plausibility of models. Hoeting et al. (1999) suggest in this situation that noninformative priors, each model *a priori* is equally likely, is a neutral choice.

EMPIRICAL ANALYSIS

Previous research (Astin, 1997; Kroc, Howard, and Hull, 1995; Mortenson, 1997; Murtaugh, Burns, and Schuster, 1999; Porter, 2000; Smith, Edminster, and Sullivan, 2001) that examined the prediction of university graduation rates make a strong assumption: the variables they select to form their models are those that causally explain the data. As Porter's results make clear, the choice of independent variables significantly influences predictions of graduation rates. What these results fail to make clear is which model specification is the true model specification that causally explains the data. The reporting of confidence intervals and standard errors of the estimates does not account for the effects of uncertainty in model specification as they are explicitly based on the results from a single model.

The purpose of the present analysis is to account for uncertainty in variable selection when modeling the 6-year graduation rate at Doctoral I universities. Appropriate variables to consider as controls are those measures that have been theoretically (Astin, 1993; Pascarella and Terenzini, 1991; Tinto, 1987) linked to educational persistence and graduation. Our selection of candidate variables to control for is guided by Astin's (1991) Input-Environment-Output (IEO) model. Influencing outcomes such as graduation rates are the personal qualities of the student body (inputs) and the environment in which students interact. The relationship among these variables is often complex in that controlling for one effect may influence the effect of another. Including both types of effects is thus important to predicting student outcomes.

The characteristics of an institution's student body provide measures that we can use to control for the quality of an institution's inputs. One would expect that institutions with well-prepared students would have high graduation rates. Astin (1991) categorizes the types of student input measures as those that describe the demographic characteristics, cognitive functioning, aspirations and expectations, self-ratings, values and attitudes, behavioral patterns, and educational background of students. Also relevant to the attainment of graduates is the manner and method of production at each institution. For instance Tinto (1987) finds that environments that support student integration into the academic community encourage student retention and subsequent graduation. The amount of resources and type of institution (size, control, mission, religious affiliation, etc.) are other factors that determine the environment.

While theory suggests the importance of various input and environmental factors on graduation rates, it does not suggest which operational measures of these factors should be included as control variables in our model specification. There is no single measure of these effects; therefore, researchers are confronted with a variable selection problem. The choice is further complicated by theory that is unclear to a variable's effect when controlling for others. For instance, Porter (2000) stresses the necessity of including variables that control for race in the prediction of graduation rates, and Tinto (1987) discusses that conditioning on college preparation reduces the effects of racial variables. The result of this is that different researchers (Astin, 1997; Kroc et al., 1995; Mortenson, 1997; Murtaugh et al., 1999; Porter, 2000; Smith et al., 2001) select different variables in their model specification of graduation rates.

The variables used in our analysis of 6-year graduation rates are roughly categorized as measures of the background of the student body, institutional control and setting, and the quality of the institution. There are 24 control variables that we consider as candidates to form the model that generates the data. These variables were selected based on theory provided by Astin's (1991) IEO model, their past use (Astin, 1997; Kroc et al., 1995; Porter, 2000) to predict graduation rates, and data availability. Several variables are included to capture each of these effects, because theory does not suggest a unique set of variables to capture the influence of student inputs and institutional environment. We allow the data to speak to which variables should be included and will account for model uncertainty in reporting our results. The variables are drawn from the online version of *U.S. News Americas Best Colleges 2002* and the Integrated Postsecondary Education Data System (IPEDS). A summary of variable definitions and their source is provided in Table 1. Our sample, with no missing observations, consists of 184 Doctoral I universities.

Capturing the demographics of the student body are variables that include the percentage of the student body that are African American, Asian, Native American, and Hispanic, along with the percentage that are male, the average age, and the percentage of out-of-state students. Student preparation for college also influences graduation rates. Controlling for this are variables that measure the percentage of the student body in the top 10% of their high school class, the SAT score of the lowest quartile, and the SAT score of the highest quartile.

Environmental factors we consider reflect the effects of organizational structure and geographic setting. Indicator variables for public versus private schools, religious versus nonreligious affiliation, and urban versus nonurban are included in the analysis along with the total enrollment. In addition, we include several variables to capture the quality of the institution and the level of student integration into the academic community. Measures of the quality of the institution are the percentage of alumni who give, per-student expenditures, acceptance rate, and weighted price. The weighted price is the weighted sum of tuition and fees for in-state and out-of-state students weighted by the percentage of in-state and out-of-state students respectively. Faculty quality is measured by the percentages of full-time faculty and faculty with a PhD. The role of integration is measured by the student-to-faculty ratio, the percentage of classes with less than 20 students, the percentage of classes with more than 50 students, and the percentage of the student body that lives in residence halls.

TABLE 1. Description and Source of Candidate Variables

Predictor	Description	Mean	Std. Dev.	Source[a]
AfricaAm	% African American	8.049	9.6611	IPEDS FA2000.DAT
NativeAm	% Native American	0.598	1.036	IPEDS FA2000.DAT
Asian	% Asian	7.223	7.425	IPEDS FA2000.DAT
Hispanic	% Hispanic	4.946	5.932	IPEDS FA2000.DAT
Enroll	Undergraduate Enrollment	17512.7	10434.4	IPEDS IC98_SRV.DAT
OutState	% Out of State	29.272	25.853	USNWR
Age	Average Age of Undergraduates	20.891	1.168	USNWR
FacPhD	% of Faculty with PhD	89.136	8.010	USNWR
Stud/Fac	Student to Faculty Ratio	14.853	3.997	USNWR
Class20	% of Classes under 20 Students	46.071	12.427	USNWR
Class50	% of Classes above 50 Students	10.980	6.075	USNWR
FtFaculty	% of Faculty Full-Time	88.473	8.574	USNWR
TopTen	% in Top Ten Percent of H.S. Class	38.549	25.378	USNWR
AccptRt	Acceptance Rate	66.245	20.447	USNWR
Alumni	Alumni Giving Rate	18.511	10.085	USNWR
StExp	Per Student Expenditures	21398.8	16469.6	IPEDS F9596-B.DAT
LowSat	SAT Score of lowest quartile	1050.89	132.859	USNWR
HighSat	SAT Score of highest quartile	1258.70	119.809	USNWR
WtPrice	Tuition and fees	10758.9	8130.38	IPEDS IC2000.DAT
ResHall	% of Undergraduates in Residence Halls	46.1218	24.2025	IPEDS IC2000.DAT
Male	% Male	49.4431	8.833	IPEDS EF98_ANR.DAT
Public	1 if Public School, 0 otherwise	0.630	0.484	IPEDS FA2000.DAT
Religion	1 if Religiously Affiliated, 0 otherwise	0.136	0.344	IPEDS FA2000.DAT
Urban	1 if in Large City or in Fringe, 0 otherwise	0.435	0.497	IPEDS FA2000.DAT

[a]USNWR: *U.S. News and World Report America's Best Colleges 2002* rankings. Web Address: http://www.usnews.com/usnews/edu/college/rankings/ranknatudoc.htm. IPEDS: Integrated Postsecondary Education Data System. Web Address: http://nces.ed.gov/ipeds.

To account for uncertainty in the prediction of 6-year graduation rates at Doctoral I universities, we apply *bicreg* to our data, which consists of graduation rates and 24 candidate-independent variables for 184 institutions. This determines which of the more than 16 million models are supported by the data, which is defined as models whose PMP are within 1/20th of that with the highest PMP. Using the estimates from these models, weighted by each model's PMP, the program produces predictions that account for uncertainty in model specification. To compare predictions made by averaging over several models with those that are based on a single model, we estimate three models that are specified using three variable selection techniques, which include minimizing Mallow's C_p, maximizing adjusted R^2, and Efroymson's stepwise method.

To assess the out-of-sample predictive performance of BMA estimates against those of standard variable selection methods, which select a single model, we randomly split the data into two subsets containing one half of the data. The first set was used to build each model and generate the corresponding coefficient estimates, which were then used to make predictions for data in the second set.

Predictive performance was evaluated by comparing the predictive mean squared error of the various methods. Predictive mean squared error (MSE) is

$$\frac{1}{N}\sum_{i=1}^{N}(Y_i - \hat{Y}_i)^2$$

where Y_i is each of the N graduation rates in the prediction subset, and $\hat{Y}_i$ is the predicted mean of graduation rates from each method. The predicted values are found by multiplying the matrix of covariates in the prediction subset with the estimates derived from using the building set of data.

RESULTS

Our results display evidence of model uncertainty in the prediction of graduation rates. Fifty-six models were selected by BMA that were within Occam's window. The 10 models with highest PMP appear in Table 2. The model with the highest PMP only accounted for 7 percent of the total posterior model probability, which is to say that the data support several models as possibly being the true model. Choosing any single model to base predictions on in this case is inherently risky. Our estimates generated by BMA account for the uncertainty in model specification by averaging over the estimates of each of the 56 models, with the weight of each estimate given by its PMP. Therefore, estimates from models with a higher PMP receive more weight.

Table 3 provides the posterior mean, standard deviation, and effect probabilities for each of our estimated coefficients. The first two values are similar in interpretation to the coefficient and standard error reported in standard analyses. The latter value, the PMP, represents the posterior probability that the coefficient is not equal to zero. Raftery (1995) provides a rough guide to the significance of posterior effect probabilities in citing 50–75%, 75–95%, 95–99%, and 100% as weak, positive, strong, and very strong evidence of a variable having an effect. From our results, we see that six variables—Native American, age, top 10, low SAT, male, and urban—each have high posterior effect probabilities, which indicate positive evidence for each having an effect. Three other variables—religion, faculty PhD, and alumni giving—received weak to somewhat less than weak support as having an effect. Nine variables had a posterior effect probability equal to zero, which means they were not selected in any of the 56 models averaged over. For these variables there is evidence against them having an effect on graduation rates.

The estimated effects for the nine variables that received support for predicting graduation rates are consistent with what theory suggests. Similar to the empirical findings of other researchers (Astin, 1997; Kroc et al., 1995; Morten-

TABLE 2. The Coefficients for the 10 Models with the Highest Posterior Model Probability (PMP)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10
Intercept	84.114	79.473	89.084	91.719	89.943	102.176	97.916	95.478	79.766	95.541
NativeAm	–1.466	–1.531	–1.484	–1.517	–1.477	–1.460	–1.446	–1.517	–1.311	–1.402
OutState					0.076	0.070		0.056		
Age	–4.224	–4.019	–4.283	–4.107	–4.326	–4.41	–4.339	–4.209	–3.949	–4.554
FacPhD	0.206	0.190			0.185					
TopTen	0.136	0.143	0.159	0.158	0.168	0.179	0.151	0.180	0.173	0.152
Alumni	0.222	0.204		0.170			0.186			
LowSat	0.059	0.059	0.073	0.064	0.056	0.061	0.065	0.062	0.063	0.075
WtPrice									0.001	
Male	–0.454	–0.410	–0.385	–0.397	–0.399	–0.396	–0.444	–0.362	–0.369	–0.434
Public									8.079	
Religion		3.653	4.292	3.979				3.579	5.832	
Urban	–4.548	–5.103	–5.589	–5.062	–5.191	–5.008	–4.450	–5.511	–5.685	–4.979
PMP	0.07	0.07	0.06	0.05	0.04	0.04	0.04	0.03	0.03	0.03

TABLE 3. Results of BMA Applied to the Prediction of 6-Year Graduation Rates

	Bayesian Model Averaging		
Predictor	Mean β/D	Std Error β/D	Pr(β ≠ 0/D)
Constant	86.79	17.82	100
AfricaAm	0	0	0
NativeAm	–1.256	0.6971	86
Asian	0	0	0
Hispanic	0	0	0
Enroll	0.00001	0.00004	12
OutState	0.0191	0.0346	29
Age	–4.161	0.5665	100
FacPhD	0.0848	0.1083	46
Stud/Fac	0	0	0
Class20	0	0	0
Class50	0.0134	0.0568	7
FtFaculty	0	0	0
TopTen	0.1568	0.0443	100
AccptRt	0	0	0
Alumni	0.0854	0.1108	45
StExp	–0.000003	0.00002	4
LowSat	0.0619	0.012	100
HighSat	0	0	0
WtPrice	0.0001	0.0002	24
ResHall	0	0	0
Male	–0.4016	0.076	100
Public	1.064	2.901	14
Religion	2.48	2.641	56
Urban	–5.151	1.197	100

son, 1997; Murtaugh et al., 1999; Porter, 2000; Smith et al., 2001), we find that better prepared student bodies, as measured by high school grade point average (GPA) and SAT scores, improve graduation rates. The coefficients for the percentage of the institution in the top 10% of their high school class and the SAT score of the bottom quartile were positive and significant with $Pr(\beta \neq 0/D) =$ 100%. Our results with respect to the racial composition of the institution are quite interesting, in that the percentage of African American, Asian, and Hispanic students had no influence on the prediction of graduation rates. Only the percentage of Native American students received positive support, $Pr(\beta \neq 0/D) =$ 86%, for having an effect on graduation rates. The coefficient for the percentage of Native American students is negative. Further research needs to examine whether this effect is due to the small number of Native Americans at the institutions in our sample, or if it is indicative of the special needs of Native Americans relative to other racial groups.

In terms of the other variables we found very strong evidence, $Pr(\beta \neq 0/D) =$ 100%, that average age, percentage of male students, and urban environments had a negative effect on graduation rates. There was weak evidence, $Pr(\beta \neq 0/D) = 56\%$, for religious institutions having a positive effect on graduation rates and marginal evidence of a positive effect for the percentage of faculty with a PhD and alumni giving.

Not only are we interested in the estimated effects of each of the variables on graduation rates, but also in their combined ability to predict graduation rates. We randomly split the data into subsets containing one half of the data. The results (not shown) generated by Bayesian model averaging applied to one half of the data are quite similar to those found in the complete data set. Variables that were found to have a strong effect on predicting graduation rates in the full data set are also supported in the sample. The only difference is in the reported effect of SAT scores. In the full sample the posterior effect probability for SAT scores of the lowest quartile was 100%, while that for the highest quartile was 0%. Thus in the full data set each model included the SAT score of the lowest quartile and excluded that of the highest. The data in the sample though included either the high SAT score, $Pr(\beta \neq 0/D) = 78\%$ or the low SAT score $Pr(\beta \neq 0/D) = 22\%$ in each of the models averaged over. Thus, the data clearly suggest that SAT scores influence graduation rates and that only one measure is needed, but one is not able to determine which one should be used with certainty. We then use these estimated coefficients to make predictions of graduation rates from the data not used to build the model. The predicted mean squared error of our BMA estimates is 60.69.

The variable selection methods used here for comparison purposes are minimizing Mallow's C_p, and Efroymson's stepwise method, and maximizing adjusted R^2. In our random sample, Mallow's C_p and Efroymson's methods both selected the same model, which had a predicted mean squared error of 68.12. Results appear in Table 4. Maximizing adjusted R^2 resulted in the same variables in the model specification, along with the inclusion of the percentage of the student body that are Asian. The mean squared error of this model was 72.95. As we can see, standard variable selection methods include several variables that BMA finds have little to no effect on the prediction of graduation rates when accounting for uncertainty in model specification. Further, BMA provides improved predictive performance as is evident in its lower mean squared error. This analysis was repeated on 20 different random samples from the data. The results find that BMA has the lowest MSE (64.17) on average relative to that of Mallow's C_p (68.22), adjusted R^2 (66.30), and Efroymson's method (67.25).

TABLE 4. Regression Results using Stepwise, Mallow's C_p, and Adjusted R^2 on a Random Sample

	Stepwise & C_p		Adjusted R^2	
Predictor	Coefficient	Std. Error	Coefficient	Std. Error
Intercept	98.64	31.97	105.3	32.39
NativeAm	−1.851	0.6017	−1.821	0.6008
Asian			0.1713	0.1461
Hispanic	−0.3536	0.1974	−0.4595	0.2167
Enroll	0.0002	0.0001	0.0001	0.0001
Age	−3.12	0.8435	−3.265	0.8506
Stud/Fac	−0.6275	0.3241	−0.6493	0.3238
Class20	−0.2337	0.0858	−0.2352	0.0856
TopTen	0.2224	0.0665	0.1958	0.0701
Alumni	0.2139	0.1221	0.2453	0.1247
HighSat	0.0337	0.0183	0.0333	0.0182
WtPrice	0.0008	0.0003	0.0007	0.0003
Male	−0.3797	0.1007	−0.404	0.1025
Public	7.433	4.203	6.95	4.213
Religion	7.8	2.856	8.093	2.86
Urban	−5.886	1.731	−6.474	1.799
N	92		92	
R^2	.8693		.8716	
SEE	6.55		6.53	
PMSE	68.125		72.951	

CONCLUSION

Variable selection is an important part of empirical research. The choice of variables to include as controls should be guided by theory, yet often there is no single theory that explains complex social relations. Measurement of theoretical effects further complicates a researcher's choice of model specification. Statistical methods of variable subset selection may aid in this choice, but are based on no theoretical foundation. As a result of these problems, it is common to find researchers who examine the same phenomena using several different model specifications. The use of different models may then lead to contrary findings. Our analysis of graduation rates at Doctoral I universities has shown that variable selection influences the effects of estimated coefficients, their statistical significance, and the overall prediction of the model. This creates uncertainty to the true relationship among the data.

As an empirical researcher it would be nice to account for uncertainty in variable selection in our results. Our article has discussed the use of BMA as a solution to this problem. In averaging over models that are supported by the data, we are able to account explicitly for uncertainty in our predictions. Here we have demonstrated that BMA provides better out-of-sample predictions of graduation rates than do the standard statistical methods of variable selection that make prediction based on a single model. On average, our out-of-sample forecast error for BMA, as measured by predictive mean squared error, was 6.3%, 3.3%, and 4.8% lower than that found by minimizing Mallow's C_p, maximizing adjusted R^2, and Efroymson's stepwise method, respectively.

In addition to improving the prediction of graduation rates, BMA also helps institutions to determine more accurately which institutional factors are important to educational outcomes. The results of this article indicate that increasing the percentage of the student body that are Native American, percentage of the student body that are male, and average age each decrease an institution's graduation rate. Evidence also supports that increasing the percentage of the student body in the top 10% of their high school class and the SAT score of the lowest quartile both increase graduation rates. We also found that institutions in urban environments had lower graduation rates. These findings indicate characteristics of individuals with greater need and to which more institutional resources need to be targeted.

The primary limitation of implementing BMA is that it is not part of the standard statistical packages familiar to most in education. To implement BMA requires access to S-Plus statistical software, publicly available programs, and limited knowledge of computer programming. Despite this limitation the results are worth the effort. BMA has improved our ability to explain variation in graduation rates and has also improved our ability to predict graduation rates. Both of these factors are important to understand fully the relationship between the inputs and environment that generate positive student outcomes, such as graduation rates. It is a better understanding of these factors that will lead to policy that best promotes such an outcome.

REFERENCES

Astin, A. W. (1991). *Assessment for Excellence*, Macmillan, New York.

Astin, A. W. (1993). *What Matters in College,* Jossey-Bass, San Francisco.

Astin, A. W. (1997). How "good" is your institution's retention rate? *Research in Higher Education* **38:** 647–658.

Avramov, D. (2002). Stock return predictability and model uncertainty. *Journal of Financial Economics* **64:** 423–458.

Bartels, L. M. (1997). Specification uncertainty and model averaging. *American Journal of Political Science* **41:** 641–674.

Draper, D. (1995). Assessment and propagation of uncertainty, with discussion. *Journal of the Royal Statistical Society, Series B* 57: 45–97.

Draper, N. R., and Smith, H. (1981). *Applied Regression Analysis*, Wiley, New York.

Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. (1999). Bayesian Model Averaging: A tutorial. *Statistical Science* **14:** 382–417.
Kroc, R. D., Howard, R. H., and Hull, P. (1995). Predicting graduation rates: A study of land grant, Research I, and AAU universities. Paper presented at the Association for Institutional Research meeting, Boston, 1995.
Lamon, E. C., and Clyde, M. (2000). Accounting for model uncertainty in prediction of chlorophyll a in Lake Okeechobee. *Journal of Agricultural, Biological and Environmental Statistics* **5(3):** 297–322.
Leamer, E. E. (1978). *Specification Searches: Ad Hoc Inference with Non-Experimental Data,* Wiley, New York.
Madigan, D., and Raftery, A. E. (1994). Model selection and accounting for model uncertainty in graphical models using Occam's Window. *Journal of American Statistical Association* **89:** 1535–1546.
Madigan, D., and York, J. (1995). Bayesian graphical models for discrete data. *International Statistical Review* **63:** 215–232.
Miller, A. J. (1990). *Subset Selection in Regression*, Chapman and Hall, London.
Mortenson, T. (1997). Actual versus predicted institutional graduation rates for 1100 colleges and universities. *Postsecondary Education Opportunity* **58**.
Murtaugh, P. A., Burns, L. D., and Schuster, J. (1999). Predicting the retention of university students. *Research in Higher Education* **40:** 355–371.
Olejnik S., Mills, J., and Keselman, H. (2000). Using Wherry's adjusted R^2 and Mallow's C_p for model selection from all possible regressions. *The Journal of Experimental Education* **68(4):** 365–380.
Pascarella, E. T., and Terenzini, P.T. (1991). *How College Affects Students*, Jossey-Bass, San Francisco.
Porter, S. R. (2000). The robustness of the graduation rate performance indicator used in the *U.S. News & World Report* college rankings. *International Journal of Educational Advancement* **1(2):** 145–163.
Raftery, A. E. (1995). Bayesian Model Selection in social research. In: Marsden, P. V. (ed.), *Sociological Methodology 1995*, Blackwells Publishers, Cambridge, MA, pp. 111–195.
Raftery, A. E., and Volinsky, C. T. (1996). Splus function Biclogit, version 2.0.
Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics* **6:** 461–464.
Smith, W. R., Edminster, J. H., and Sullivan, K. M. (2001). Factors influencing graduation rates at Mississippi's public universities. *College and University* **76(3):** 11–16.
State Council of Higher Education of Virginia (1999). *Virginia Higher Education Performance Funding Model*, June 14, 1999.
Tinto, V. (1987). *Leaving College,* University of Chicago Press, Chicago.
Volinsky, C. T., Madigan, D., Raftery, A. E., and Kronmal, R. A. (1997) Bayesian model averaging in proportional hazard models: Assessing the risk of a stroke. *Applied Statistics* **46(3):** 433–438.
Weisberg, S. (1985). *Applied Linear Regression*, Wiley, New York.
Zelner, A. (1971). *Introduction to Bayesian Analysis in Econometrics,* Wiley, New York.

Received September 19, 2002.