

RETURNS TO SCHOOLING AND BAYESIAN MODEL AVERAGING: A UNION OF TWO LITERATURES

Justin L. Tobias

University of California – Irvine

Mingliang Li

University of California – Irvine

Abstract. In this paper, we review and unite the literatures on returns to schooling and Bayesian model averaging. We observe that most studies seeking to estimate the returns to education have done so using particular (and often different across researchers) model specifications. Given this, we review Bayesian methods which formally account for uncertainty in the specification of the model itself, and apply these techniques to estimate the economic return to a college education. The approach described in this paper enables us to determine those model specifications which are most favored by the given data, and also enables us to use the predictions obtained from *all* of the competing regression models to estimate the returns to schooling. The reported precision of such estimates also account for the uncertainty inherent in the model specification. Using U.S. data from the National Longitudinal Survey of Youth (NLSY), we also revisit several ‘stylized facts’ in the returns to education literature and examine if they continue to hold after formally accounting for model uncertainty.

Keywords. Returns to Schooling; Bayesian; Model Averaging; Variable Selection

‘It is my impression that rather generally, not just in econometrics, it is considered decent to use judgement in choosing a functional form, but indecent to use judgement in choosing a coefficient. If judgement about important things is quite all right, why should it not be used for less important ones as well?’ – J.W. Tukey (1978), taken from Poirier (1995, page 524).

1. Introduction

In the above quote Tukey points out that researchers (not just econometricians!) implicitly use rather strong prior information to arrive at the selection of a particular model or set of models, but then become leery of incorporating further prior beliefs regarding the coefficients *within that model*. Although one could appeal to Tukey's quote in defense of the practice of incorporating prior information about a set of parameters, one could also argue that researchers should not impose such strong prior information to select their model(s), which is, perhaps, the 'important' thing.

One topic which receives considerable attention in the economics literature, and seems to be characterized by strong (and heterogeneous across researchers) prior beliefs regarding the appropriate model specification, is the estimation of the economic returns to education. A quick search of 9 journals¹ in JSTOR over the period 1970–1995 produced 176 citations with 'Education' appearing in the title and 53 with 'Schooling' in the title.² These numbers are typically smaller than, though perhaps competitive with, the results obtained when searching for the appearance of other words in article titles, such as 'exchange rate' (131 articles), 'inflation' (333 articles), and 'unemployment' (361 articles). The amount of work done on this topic relative to the amount done on other key issues in economics clearly suggests the importance of identifying the true economic returns to schooling.

Given this vast amount of work, it is probably not surprising to learn that different researchers have used a variety of different specifications to arrive at estimates of the economic returns to education. In this literature, numerous studies have described the need to control for various explanatory variables such as measures of 'cognitive ability,' or 'sheepskin effects' (*i.e.* nonlinearities upon degree completion). As these variables are potentially correlated with education, or obviously affect the shape of the schooling-earnings relationship, their inclusion can have a potentially significant impact on estimated returns to education.

To illustrate the existence of specification uncertainty in this literature, we have reviewed 38 different articles, chosen with preference given to articles published after 1970 whose primary focus was on the 'returns to schooling' and appeared in general-interest economics journals, and recorded the specifications used in these studies. In our review of these 38 studies, we determined if any attempt was made to control for cognitive ability or nonlinearities upon grade completion (*i.e.* at 12 and 16 years of schooling).³ We additionally record if the study adds controls for family characteristics such as parental education or income, a marriage indicator, regional variation in the pricing of education, and variation in returns to education with measured cognitive ability and family characteristics. Presented in Table 1 below are the probabilities that the 38 reviewed studies attempted to control for each of these variables.

The results in Table 1 suggest that researchers are not sure of the 'correct' functional form of a log wage equation. Most studies (nearly 70 percent) made some attempt to control for cognitive ability, while fewer studies (nearly 60 percent) permitted possible nonlinearities in the schooling-earnings relationship.

Table 1. Probabilities of Including Variables among 38 Studies Surveyed.

Variable	Probability of Inclusion
Ability	.68
$I(Ed \geq 12)$.58
$I(Ed \geq 16)$.58
Family Characteristics	.47
Marriage	.29
Ed*Ability	.21
Ed*Region	.05
Ed*Family	.02

Approximately one-half of the studies added controls for family characteristics, while few studies tried to control for variation in returns to schooling across regions, or variation in returns with ability or family characteristics.

The standard approach in this literature in the presence of specification uncertainty involves ‘pre-testing,’ or estimating a variety of regression relationships and then reporting the estimates associated with a single preferred specification (perhaps by inspection of the associated t -statistics). As is well-known in the statistics and econometrics literatures, this approach is replete with deficiencies (see, *e.g.* Berger and Pericchi (2001)). First, it is important to recognize that the standard error associated with the model-selected point estimate is not simply the standard error obtained *within* the selected model. These are incorrect as they are obtained *conditioned* on the model, and do not reflect the uncertainty attributable to the selection of the model itself. Second, the process of model selection ignores the point predictions obtained from all of the other competing models, and simply places probability one on the point prediction obtained from the selected model. In general, we might think that the data do not unanimously favor a particular specification, but instead, spreads its affection over a variety of competitors. Hence, we might think of replacing *model selection* with *model averaging*. In the model averaging framework, we will obtain point predictions from all possible competing models, determine ‘weights’ associated with each specification, and report as our final estimate a weighted average of the model-specific predictions.⁴

In this paper, we describe and empirically investigate the impact of specification uncertainty on point estimates of the returns to education. We recognize from Table 1 that researchers have used a variety of specifications to estimate the returns to schooling, while the question of which specification should be used has not been thoroughly addressed in this literature. As described in the following section, many studies have described the empirical importance of a *subset* of the variables mentioned in Table 1 and argued for their inclusion in regression equations, while few studies have systematically addressed the empirical importance of all of these variables.

Given this seemingly unresolved issue, we describe a method for determining those specifications that are most favored by the given data and apply model averaging techniques to estimate the economic return to the receipt of a college

degree. We do this in a Bayesian approach, since in the Bayesian setting we are able to *assign posterior probabilities to the competing regression models*. Essentially, we think of different models as different ‘parameters,’ and describe methods for (1) determining the posterior probabilities associated with different models, (2) obtaining posterior distributions of the return to a college degree which are not dependent on a particular model, and (3) computing standard errors of parameters of interest which incorporate the uncertainty in the model itself.

We review some of the literature on Bayesian model selection and model averaging, and discuss its implementation in practice. Although our focus in this paper is primarily on implementation of these methods, it is important to note that Bayesian model averaging has a number of desirable properties, including (1) giving better ‘predictive performance’ than predictions obtained from any particular model in terms of the expected log predictive score (Madigan and Raftery (1994)), (2) adapting easily to the comparison of non-nested models, (3) building in penalties for model complexity, (4) formally accounting for the uncertainty of the model specification itself in the reported precision of parameters of interest, and (5) adapting to select the ‘true’ model (with sufficient data), if the true model is contained in the set of competing models. More complete discussions of the use of Bayesian model selection and model averaging can be found in Poirier (1985, 1991), Madigan and Raftery (1994), Gelfand and Dey (1994), Draper (1995), Kass and Raftery (1995), Kass and Wasserman (1995), Raftery, Madigan and Hoeting (1997), Hoeting *et al* (1999), Wasserman (2000), Berger and Pericchi (2001), and Fernandez, Ley and Steel (2001a, 2001b), among others.

The outline of the paper is as follows. In section 2, we review several studies in the returns to schooling literature, and describe procedures for Bayesian model averaging. The National Longitudinal Survey of Youth (NLSY) data used in this analysis is described in section 3. In section 4, we illustrate our methods using a restricted set of models while in section 5, we employ our methods in a much larger model space which consists of over 262,000 different specifications. We determine the specifications that are most favored by the NLSY data, provide clear evidence of model uncertainty, and revisit some ‘stylized facts’ in the education literature to see if they continue to hold after accounting for such uncertainty. The paper concludes with a summary in section 6.

2. A Review of Previous Studies and Bayesian Model Averaging

To estimate the return to schooling, the vast majority of studies employ a regression model of the following form:^{5,6}

$$y_i = Z_i\gamma + Ed_i\theta + \epsilon_i, \quad (1)$$

where y_i is typically the log wage or log earnings of individual i , Ed denotes years of education completed, Z denotes a set of other covariates affecting wages, and θ is our primary parameter of interest, which we interpret as measuring a private return to schooling.

2.1 *Previous Studies*

This *baseline* regression, similar to that described in pioneering work by Mincer (1958, 1970, 1974) and Becker and Chiswick (1966), has seen many generalizations. In fact, we can rationalize the contributions of many previous studies in this literature as simply describing the need to control for additional variables in Z in order to obtain accurate estimates of θ . This is true, for example, in studies concerned about self-selection into higher education, in which case Z might contain inverse-Mills ratio terms in order to obtain selection-corrected conditional mean functions given a joint normality assumption (Willis and Rosen (1979) and Heckman, Tobias and Vytlačil (2001, 2003)). More robust techniques have been developed in recent work that incorporate the use of the propensity score in this context (*e.g.* Ahn and Powell (1993) and Heckman, Ichimura, Smith and Todd (1998)).

Other studies in this literature have simply focused on a few key variables which should appear in Z , and have derived methods for estimating returns to schooling in the presence of these variables. In the following sections, we review several key variables whose inclusion has garnered considerable attention, and also cite key studies that argued for the need to control for these variables.

2.1.1 *Ability and Family Background*

Numerous studies in the education literature have emerged describing various ways to control for ability and family characteristics when estimating the returns to education. ‘Twins (or siblings) studies,’ for example, (*e.g.* Ashenfelter and Krueger (1994), Miller, Mulvey and Martin (1995), Altonji and Dunn (1996), Behrman, Rosenzweig and Taubman (1996), Ashenfelter and Rouse (1998), Arias, Hallock and Sosa-Escudero (2001)) provide a creative way for controlling for unobservable ability and family endowments. The approach underlying these studies is to find data on twins or siblings whose ability and family characteristics are arguably identical, thus allowing researchers to purge regression specifications of these variables by differencing across the twins or siblings. Analysis of the specification purged of these effects (under certain conditions) will provide consistent estimates of the return to schooling parameter, as the unobserved effects which are potentially correlated with schooling have been differenced out.

The marginal role of cognitive ability itself has been featured in numerous articles (*e.g.* Ashenfelter and Mooney (1968), Weisbrod and Karpoff (1968), Griliches and Mason (1972), Blackburn and Neumark (1993), Murnane, Levy and Willett (1995), Grogger and Eide (1995), and Heckman and Vytlačil (2001)), while family background variables have been paramount in others studies (Morgan and David (1963), Morganstern (1973), Behrman and Wolfe (1984), and Lam and Schoeni (1993)).

As further evidence of the vast literature dedicated to describing methods for controlling for these factors, consider the quote by Blackburn and Neumark (1995, page 217):

‘One of the longest-running debates in empirical labor economics regards bias in OLS estimates of the economic return to schooling. The overriding concern

pertains to individual-specific productivity components not reflected in the usual human-capital measures, as these ability components may be positively correlated with both wages and schooling.’

2.1.2 *Nonlinearities in the Schooling-Log Wage Relationship*

Another source of specification uncertainty arises when trying to describe the function mapping schooling to earnings. In particular, several studies have focused on the importance of controlling for nonlinearities in education upon degree completion or ‘sheepskin effects’ (e.g. Hungerford and Solon (1987), Belman and Heywood (1991), Heywood (1994) and Heckman Layne-Ferrari and Todd (1996), Jager and Page (1996)). These studies note that the marginal return to an extra year of schooling may not be constant over the schooling support, and in particular, we might see jumps or premiums paid upon the completion of degrees – such as a high school degree, Bachelor’s or Master’s degree.

Hungerford and Solon (1987, page 177) describe the empirical importance of these nonlinearities as they state:

‘All of our results point to the existence of sheepskin effects in the returns to education. This finding suggests, first, that treating the log wage as a smooth function of years of education, as is conventionally done in the earnings function literature, gives an inferior fit to the data.’

2.1.3 *Returns to Schooling Varying with Ability*

Rather than simply describing the need to control for measured cognitive ability, other studies argued for the additional need to include ability-education interactions. These studies note that the receipt of an additional year of schooling may have differential impacts across individuals, and this variation can potentially be explained through differences in individual ability. For example, Hause (1972), Blackburn and Neumark (1993), Murnane, Levy and Willett (1995), Grogger and Eide (1995), Heckman and Vytlačil (2001), and DiNardo and Tobias (2001) and Tobias (2003) investigate the question of whether returns to schooling vary with ability. Generally, these studies find that higher ability individuals benefit more (in terms of wages) from additional years of education than lower ability individuals.

For regression models failing to include an ability-education interaction, Hause (1972, page S111) states:

‘This [specification] does not seem plausible, since it implies that people of low ability have a greater incentive to invest in schooling. The increase in earnings from an increment of schooling is the same for all, regardless of ability, but the opportunity cost of foregone earnings while acquiring the schooling is lower for those with less ability. This argument implies the earnings function is misspecified unless ability increases the marginal product of schooling.’

Finally, other studies have described the need to control for still other explanatory variables when attempting to estimate the return to education. Korenman

and Neumark (1991) discuss the importance of including a marriage variable in earnings regressions, Heckman, Layne-Ferrar and Todd (1996) discuss the potential for regional variation in the pricing of education (and schooling quality), and Altonji and Dunn (1996) allow returns to schooling to vary with family characteristics by interacting education with parental education. Thus, all of the ‘optional’ variables mentioned in Table 1 have a precedent in the literature – with some studies taking a strong stance on the inclusion of particular variables among this set – yet few studies examining the combined effect of all of these variables.

What seems called for, then, is a study which considers the set of all possible regression models based on the ‘optional’ variables appearing in Table 1 (and others), and determines the specifications most favored by the given data. Since point estimates of the returns to education will surely be affected by including or failing to include these particular variables, it is important to seek out those specifications that are most favored by the given data. Moreover, we will not simply *choose* a particular model among this large set of competitors, but rather we will use the information provided by all of the models to obtain point estimates of the return to a college degree. Additionally, our approach will formally incorporate the uncertainty in the model itself when reporting standard errors of parameters of interest.

2.2 Bayesian Model Averaging

Given the discussion of the previous sections, the problem we will address in this paper is a standard problem of variable selection in the context of a typical log-wage equation:

$$y = W\alpha + X\beta + \epsilon, \quad \epsilon \sim N(0, \sigma^2 I_n). \quad (2)$$

In the above, y denotes the log of hourly wages, W denotes a set of fixed explanatory variables that will appear in all of our regressions,⁷ and X is a $n \times k$ matrix of k ‘optional’ covaraites. The inclusion or exclusion of the elements of X will define the set of possible regression models under consideration.

The approach in this paper is to note that *all* of these possible specifications will provide the researcher with a point prediction of some parameter of interest, such as the private return to an additional year of schooling. Since we have k variables that could be included in our regression, the set of all possible models (or regressions) has 2^k elements. In this paper we will employ an approach that uses the information provided by *all* of these regressions, determines those specifications that receive the most support from the data, and produces standard errors that formally account for uncertainty in the model specification itself.

Throughout our investigation, our primary parameters of interest are of the following forms:

$$\Delta_{16,12} = E(y|Ed = 16, X_{-Ed} = \bar{X}_{-Ed}) - E(y|Ed = 12, X_{-Ed} = \bar{X}_{-Ed}).$$

and

$$\begin{aligned} \Delta_{16,12|A=A^0} &= E(y|Ed = 16, A = A^0, X_{-Ed,A} = \bar{X}_{-Ed,A}) \\ &\quad - E(y|Ed = 12, A = A^0, X_{-Ed,A} = \bar{X}_{-Ed,A}), \end{aligned}$$

where we have used the notation X_{-x_j} to denote all variables other than x_j .

The first parameter (multiplied by 100) gives the percentage increase in hourly wages resulting from having a 4-year college degree relative to a high school degree, holding all other variables constant and equal to mean values. The second parameter is the same effect, but gives the return for those individuals with ability fixed at A^0 . For example, if $A^0 = 1$, then $\Delta_{16,12|A=1}$ will be interpreted as the return to a 4-year degree for individuals one standard deviation above the mean of the ability distribution.

2.2.1 Basic Elements of Discrete Model Averaging

We begin our brief discussion of the mechanics behind model averaging by noting the following equation which follows from the law of total probability:

$$p(\theta|y) = \sum_{i=1}^T p(\theta|M_i, y) \Pr(M_i|y). \quad (3)$$

In the above, M_i indexes the T different models, p generically refers to a density function and y denotes the observed data.

The above equation states that the posterior distribution of θ which is *not conditioned on a particular model* is simply obtained as a weighted average of the model-specific posterior distributions $p(\theta|M_i, y)$, where the terms $\Pr(M_i|y)$ serve as the weights. These weighting terms are called *model probabilities*, as they simply give the posterior probability in support of model M_i among the class of competing models.

Thus, to obtain information about some quantity of interest, such as the return to a 4 year degree, *without* conditioning on a particular model, (3) provides us with the answer. We simply compute the desired effect for all the models we can imagine, and weight the predictions from each model ‘strategically.’ The actual procedure for carrying out the weighting is determined by calculating the posterior probability associated with each competing model.

The calculation of the posterior model probabilities is a non-trivial exercise, and certainly can make the implementation of model averaging quite difficult in general. Fortunately, for the case of the standard linear regression model examined here, it is possible to calculate these model probabilities *analytically*⁸ and thus calculation of quantities as in (3) becomes possible when the number of optional explanatory variables is not excessively large.

To summarize in non-technical terms, one can carry out the process of (Bayesian) model averaging by proceeding as follows:

1. Calculate the model-specific posteriors $p(\theta|M_i, y)$ for all of the possible competing models.

2. Compute the posterior model probabilities $\Pr(M_i|y)$ for all the competing models. For the linear regression model examined here, closed-form solutions exist for the calculation of these quantities (Expressions provided in appendix).
3. Use the posterior model probabilities obtained from step 2 together with the model-specific posteriors from step 1 to obtain $p(\theta|y)$ which does *not* depend on a particular model, as in equation (3).

It is important to recognize that the procedure outlined above is feasible when the number of candidate models is reasonably small. However, in a regression problem with a large number of ‘optional’ variables (*e.g.* $k \geq 20$), explicit evaluation of every model becomes computationally prohibitive. In these situations, one possible solution is to use the Markov Chain Monte Carlo Model Composition (MC^3) method (*e.g.* Madigan and York (1995), Raftery, Madigan and Hoeting (1997) and Fernandez, Ley and Steel (2001a)) to additionally search throughout the model space. In this paper, however, we do not employ such methods, but instead, analytically compute the probabilities associated with more than 262,000 candidate models.

2.3 *Why Model Averaging?*

The previous section described a method for obtaining model-averaged estimates in the context of a standard regression model. Our goal in this paper is to apply these techniques to estimate the returns to education, and to argue that use of these techniques offers several advantages over procedures conventionally implemented in the literature.

First, it is important to recognize that the model averaging framework is ‘efficient’ in the sense that it uses information provided by *all* of the regression models to obtain our final estimates of the return to schooling, as in equation (3). It is perhaps more commonplace to do a series of ‘pre-tests’ to arrive at some final preferred specification, and then to report the point estimates and standard errors from that selected specification. What equation (3) shows is that the predictions from *all* of the models should be used to calculate the final (posterior) distribution $p(\theta|y)$. Model *averaging* will reduce to more conventional model *selection* only if the posterior model probabilities unanimously favor one model and place zero probability on all other competing models. In general, this will not be the case, as the data will divide its loyalties over a variety of competing specifications and will not unambiguously select one particular specification.

Second, the model averaging framework produces standard errors associated with estimated returns to schooling that are ‘correct’ in the sense that they also formally account for model uncertainty. To further support this point, let θ be some parameter of interest. It is straight-forward to show:

$$\text{Var}(\theta|y) = E_M[\text{Var}(\theta|M, y)] + \text{Var}_M[E(\theta|M, y)].$$

Thus, the model averaged variance on the left hand side which is *not* conditioned on any particular model is equal to the average variance in our parameter of interest across the models (the first term) plus the variation in point predictions

across the models (second term). Often-reported model-selected variance estimates tend to approximate the first term, but clearly ignore the second. For this reason, model averaged variances can be much larger than model-selected variances, as they correctly account for the fact that the model selection process is itself uncertain.⁹

In terms of returns to schooling, Table 1 suggests that specification uncertainty is indeed a serious problem in this literature. Some studies have controlled for nonlinearities in the schooling-earnings relationship, others have described the need to control for ability and family background characteristics, and yet others have not attempted to control for these variables. Our approach is to use the Bayesian model averaging paradigm to embrace model uncertainty and obtain posterior distributions associated with the receipt of a college degree that are not conditioned on particular models.

3. The Data

The data employed in this analysis are taken from the National Longitudinal Survey of Youth (NLSY). The NLSY is a widely-used panel survey containing a wealth of demographic information as well as information on the labor market experiences of a sample of relatively young men and women in the U.S. The survey begins in 1979 (where the respondents range in age between 14 and 22), and we obtain earnings and other data until 1993. To focus on possible dynamic changes in the earnings relationship, we obtain results year-by-year for the period 1985–1993.

The dependent variable used is the log of hourly wages at the respondent's most recent job. Using test scores from the ASVAB (Armed Services Vocational Aptitude Battery) which was administered to the NLSY participants, we construct a univariate 'ability' measure from the 10 ASVAB component tests. This ability measure is constructed from the standardized residuals of all of the 10 test scores after purging the tests of an age trend, since the participants vary in age at the time of the test and age is a strong predictor of test scores. Our final univariate ability measure is then standardized to have mean zero and unit variance. We also limit our focus to white males reporting hourly wages between \$1 and \$ 100 in the given year.

To make things tractable and to focus on key ideas, we limit the number of 'optional' variables and thus choose to keep a subset of these variables as 'fixed.' These fixed variables include a constant, indicators for residence in the northeast northcentral and south, potential labor market experience and its square, an urban indicator, and a linear education term. Thus, our set of competing models will be indexed by all combinations of a set of 'optional' explanatory variables, and we will require that the set of core explanatory variables above appear in all of our regressions.

In our most general specification, we allow a variety of different variables to appear. Our optional variables consist of 4 nonlinearities in education ($I(Ed \geq 12)$, $I(Ed \geq 16)$, $Ed * I(Ed \geq 12)$, and $Ed * I(Ed \geq 16)$), which allow for jumps upon degree completion, and also permit slopes to vary across the education support. We control for measured cognitive ability (A), and allow for interactions of

ability with schooling (Ed^*A , $A^*I(Ed \geq 12)$, $A^*I(Ed \geq 16)$). The addition of these variables relax the requirement that the marginal return to a unit of education is linear in ability over the education support.¹⁰ We also add a marriage indicator (*Married*), highest grade completed by the respondent's mother and father (*Hgcmom* and *Hgcdad*), interactions of education with the parental education variables, ($Ed^*Hgcmom$ and $Ed^*Hgcdad$), education interacted with three region indicators: (Ed^*Ncent , Ed^*Neast , Ed^*South), education interacted with an urban indicator (Ed^*Urban) and the local unemployment rate (Ed^*Urate). In total, this creates 18 optional variables in our most general model, so that we will entertain $2^{18} = 262,144$ specifications!

4. A 'Small World' of 16 Models

To illustrate our methods we first analyze a restricted set of models with only 4 optional variables, and thus allow for a total of $2^4 = 16$ possible models. As stated previously, the 'fixed' variables include an intercept, indicators for residence in the northeast, northcentral and south, an urban indicator, potential labor market experience and its square, and a linear education term. Our four optional variables include ability, an ability-education interaction, and indicators for the completion of at least 12 and 16 years of schooling (*i.e.* $I(Ed \geq 12)$ and $I(Ed \geq 16)$), respectively. These are perhaps the key variables whose inclusion has attracted the most attention in this literature, and thus investigation of this restricted class of models provides a useful way to begin illustrating the application of our methods.

4.1 Empirical Results

Presented in Table 2 are the posterior probabilities of including the optional variables using a cross-section of 1990 data from the NLSY. In our model averaging framework, these probabilities are obtained by finding all of the competing models where the optional variable of interest appears, and summing over all of the posterior probabilities associated with those models.

From the table, we see strong evidence in favor of the inclusion of the ability variable, and modest evidence in favor of the inclusion of the ability-education interaction and nonlinearity at the completion of 4 years of college education. There is very little support for any premium paid for the completion of a high school degree, as the posterior probability that we should include this variable in our model is less than 2 percent.

4.2 Posterior Model Probabilities

In Table 3 below, we break our results down in more detail, and present the posterior probabilities associated with each of our 16 different competing models. In addition, we also present our point-predictions of the return to the receipt of a four-year degree for each of these competing models.

Table 2. Posterior Probabilities of Including Alternate Variables Using Different Approaches.

	Probability
Ability (A)	69.46
Ability*Ed	36.16
$I(Ed \geq 12)$	1.60
$I(Ed \geq 16)$	30.91
$E(\Delta_{16,12} \text{Data})$.339
Std ($\Delta_{16,12} \text{Data}$)	.067

It is important to note from Table 3 that there is considerable evidence of model uncertainty. No model of the 16 competitors is ever assigned more than 50 percent of the weight, and three models which include ability on its own, an ability education interaction on its own, and both ability and a nonlinearity at 16 years of schooling consistently receive most (87 percent) of the weight. We also note that the range of our point estimates of the return to a college education is quite large – they imply that the receipt of a 4-year college education (relative to 12 years of schooling) increases hourly wages between 30–50 percent, depending on the model employed. In particular, if we would have conditioned on the specification containing only a nonlinearity at 16 years of schooling, we would report that a four year college degree tends to increase wages by 50 percent. When using a specification containing only ability and no interactions or nonlinearities, we would report a 30 percent increase in hourly wages. Thus, we have preliminary evidence that (1) different models can certainly give different predictions about

Table 3. Posterior Model Probabilities and $E(\Delta_{16,12}|M_k)$ for 16 Candidate Models.

Variable Included? (Y/N)				Model Probability	$E(\Delta_{16,12} M,y)$
A	Ed*A	I_{12}	I_{16}	$\Pr(M_i y)$	
N	N	N	N	.00	.407
Y	N	N	N	40.7	.306
N	Y	N	N	24.5	.314
Y	Y	N	N	2.93	.306
N	N	Y	N	.00	.396
Y	N	Y	N	.58	.315
N	Y	Y	N	.33	.310
Y	Y	Y	N	.05	.316
N	N	N	Y	.00	.497
Y	N	N	Y	22.1	.405
N	Y	N	Y	5.59	.402
Y	Y	N	Y	2.60	.413
N	N	Y	Y	.00	.476
Y	N	Y	Y	.44	.396
N	Y	Y	Y	.16	.387
Y	Y	Y	Y	.05	.406

key parameters of interest, and (2) there is non-ignorable posterior uncertainty regarding the appropriate specification of the earnings equation.

5. Empirical Results

Given the procedure just illustrated, we now turn these methods loose on an elaborated design that includes 18 optional explanatory variables for a total of $2^{18} = 262,144$ models.¹¹ We employ these techniques to revisit the following questions which have received attention in the schooling literature: (1) What are the *unconditional estimates of the return to a college education*? (2) Which specifications are most favored by the NLSY data? (3) Does there appear to be evidence of model uncertainty, or does the data tend to select one model with very high probability? (4) Is there evidence of ‘sheepskin effects,’ or nonlinearities upon degree completion? (5) Is measured cognitive ability an important determinant of earnings, and do returns to schooling vary with this measure of ability? (6) Are the returns to schooling increasing over time, and if so, is the increase experienced primarily by those of highest ability?

5.1 Summary of Model Exploration

For each year from 1985–1993, we obtain posterior probabilities associated with each of our 262,144 candidate models.¹² To summarize the results of this model exploration, we present in Table 4 the posterior probabilities that the coefficients on the optional variables are not equal to zero. In the limiting case where there is no model uncertainty, the data will favor one particular specification, and those variables appearing in that specification will be non-zero with probability one, while those absent from the selected specification will be zero with probability one.

From Table 4 we see several important points which may help guide the functional forms used in future research. First, we see strong and consistent evidence for the inclusion of measured ‘ability’ across virtually all years studied (fifth row), and that the returns to schooling vary with measured ability (sixth row), since the Education-Ability interaction tends to receive large weight across the years studied. We also note that the importance of the ability-education interaction may call into question the structural assumptions used in twins studies, where ‘ability’ or family endowments enter as separate additive terms, and thus can be differenced out using data on twins or siblings. In order to identify an education effect, such studies need education to differ across the twins or siblings. However, if the true model also contains an ability-education interaction, and if education varies across the twins, then the differencing technique does not purge the model of the ability-education interaction. Table 4 is suggestive of the empirical importance of this variable.

We also see that the evidence supporting the existence of nonlinearities upon degree completion is strongest for those with at least 16 years of schooling, and this effect also tends to be strongest at the later years of the sample. This finding

Table 4. $\Pr(\beta_j \neq 0 | \text{Data})$ as a Percentage: 1985–1993.

Variable	Year								
	1985	1986	1987	1988	1989	1990	1991	1992	1993
$I(Ed \geq 12)$	8.18	9.63	16.65	1.97	4.14	2.45	2.80	21.95	4.60
$I(Ed \geq 16)$	3.01	4.97	1.60	1.64	4.07	35.58	75.88	63.41	66.16
$Ed^*I(Ed \geq 12)$	7.68	9.27	18.07	2.38	4.37	3.10	3.38	36.46	6.51
$Ed^*I(Ed \geq 16)$	—	—	1.61	1.63	4.43	23.38	61.16	45.77	47.68
A	59.56	81.38	42.20	18.66	58.52	69.17	74.66	80.18	71.51
A^*Ed	41.46	37.21	31.87	43.31	44.09	33.82	41.73	28.48	37.26
$A^*I(Ed \geq 12)$	6.27	11.94	9.02	45.41	5.36	7.32	8.93	8.64	9.06
$A^*I(Ed \geq 16)$	3.80	5.17	1.35	1.56	2.02	4.44	21.85	4.93	14.47
<i>Married</i>	100.00	98.29	99.99	99.98	91.84	98.80	99.99	100.00	100.00
<i>Hgemom</i>	1.41	3.20	1.52	1.47	2.16	1.30	1.45	1.46	2.87
<i>Hgcdad</i>	1.48	4.05	1.40	1.63	3.18	5.31	4.05	11.91	38.63
$Ed^*Hgdmom$	1.88	4.24	2.23	2.27	2.82	1.96	2.15	2.26	3.51
$Ed^*Hgcdad$	1.84	4.13	1.81	2.10	5.10	4.84	4.73	10.89	27.30
Ed^*Neast	16.83	11.09	8.03	9.26	12.79	6.52	6.73	7.86	10.32
Ed^*NCent	9.78	48.52	29.16	15.91	7.94	6.38	8.71	19.25	6.17
Ed^*West	23.24	10.63	10.74	8.33	8.67	8.81	11.47	7.81	6.35
Ed^*Urban	36.14	14.73	2.02	5.09	3.24	7.12	2.80	1.15	2.56
Ed^*Urate	11.55	9.76	22.18	9.72	12.95	7.28	7.00	7.97	9.21

may be consistent with the notion of ‘skill-biased technological change.’ For the later years in the sample, the data reveals its preference for models with the nonlinearity at 16 years of education, which is suggestive of a evolving U.S. labor market which has changed to place a growing premium on those of the highest skill, where skill is measured by years of schooling. Regardless of the interpretation, the data suggest evidence of nonlinearities in the schooling-log wage relationship, and that the attainment of a 4-year degree matters, while simply completing your high school degree has much less of an effect.

What is clear from Table 4 is the data reveal a preference for models which include measured ability, a marriage indicator, an ability-education interaction, and a nonlinearity upon receipt of a 4-year college degree (at least for the later years of the sample). The data seem to be less decisive on the inclusion of the remaining variables. That is, there *appears to be a considerable amount of model uncertainty*. In the absence of such uncertainty, the data would tend to select a particular model with probability one, and thus the entries in Table 4 would either be 100 or zero, and the ‘true’ specification would include all of the variables whose probability of being non-zero were 1. Here, this is clearly not the case, and we are not certain what is the ‘correct’ specification of the log wage equation.

In Tables 5a and 5b, we compare the performances of individual models and determine the particular specifications that are most favored by the data. Specifically, we find the top 10 of the 262,144 models using data from 1990, report their posterior probabilities, and also determine how these 10 models ranked in the remaining years of the sample. To save notation, we define the top ten models as

models $M_1 \rightarrow M_{10}$, (using 1990 data). Table 5A shows the variables included in these top 10 models, while Table 5B compares their performance for the remaining years of the sample.

As suggested by Table 4 and further illustrated in Table 5A, the top models in 1990 always contain the marriage variable, and typically contain ability, an ability-education interaction, and nonlinearities at 16 years of education. Of these top 10 models, *the data tend to reveal its preference for parsimonious*

Table 5A. Top 10 Models in 1990 and Their Included Variables.

M_1	A, Married
M_2	A, $Ed \geq 16$, Married
M_3	($Ed * A$), Married
M_4	A, $I(Ed \geq 16) * Ed$, Married
M_5	A, $I(Ed \geq 16)$, $I(Ed \geq 16) * Ed$, Married
M_6	($Ed * A$), $I(Ed \geq 16)$, Married
M_7	($Ed * A$), $I(Ed \geq 16) * Ed$, Married
M_8	($Ed * A$), $I(Ed \geq 16)$, $I(Ed \geq 16) * Ed$, Married
M_9	$I(Ed \geq 12) * A$, Married
M_{10}	A, West * Ed, Married

Table 5B. Model Rank and $Pr(M_j|Data)$. Top 10 Models in 1990.

Model (M_j) [Pr ($M_j Data$)]	Year								
	1985	1986	1987	1988	1989	1990	1991	1992	1993
Rank M_1	1	1	1	3	1	1	70	119	31
[(Pr($M_1 Data$)]	[12.40]	[12.53]	[10.22]	[6.81]	[23.40]	[13.05]	[.17]	[.12]	[.53]
Rank M_2	36	21	99	77	26	2	3	1	1
[(Pr($M_2 Data$)]	[.42]	[.56]	[.13]	[.09]	[.58]	[9.72]	[8.68]	[7.92]	[5.71]
Rank M_3	2	5	3	2	2	3	82	214	43
[(Pr($M_3 Data$)]	[9.28]	[3.18]	[6.51]	[19.10]	[15.39]	[8.50]	[.13]	[.06]	[.35]
Rank M_4	36	21	98	78	24	4	5	3	6
[(Pr($M_4 Data$)]	[.42]	[.56]	[.13]	[.09]	[.63]	[6.98]	[2.92]	[4.50]	[2.55]
Rank M_5	–	–	272	208	50	5	1	6	2
[(Pr($M_5 Data$)]	[.00]	[.00]	[.03]	[.02]	[.22]	[3.46]	[15.64]	[2.16]	[4.75]
Rank M_6	50	95	132	45	27	6	7	18	12
[(Pr($M_6 Data$)]	[.24]	[.11]	[.09]	[.25]	[.56]	[2.60]	[1.72]	[.90]	[1.33]
Rank M_7	50	95	130	46	24	7	28	36	27
[(Pr($M_7 Data$)]	[.24]	[.11]	[.09]	[.25]	[.63]	[1.81]	[.54]	[.44]	[.55]
Rank M_8	–	–	357	101	46	8	2	35	5
[(Pr($M_8 Data$)]	[.00]	[.00]	[.02]	[.06]	[.26]	[1.46]	[8.81]	[.45]	[2.57]
Rank M_9	16	227	18	1	25	9	208	417	170
[(Pr($M_9 Data$)]	[1.02]	[.05]	[1.04]	[21.81]	[.60]	[1.25]	[.04]	[.02]	[.08]
Rank M_{10}	5	9	16	19	7	10	342	701	308
[(Pr($M_{10} Data$)]	[3.75]	[1.67]	[1.12]	[.67]	[2.24]	[1.22]	[.02]	[.01]	[.04]

specifications. None of these ten models contain more than four optional variables, and the most favored model contains only two optional variables. It is also important to observe that the most favored model in 1990 is only preferred over the model ranking 2nd by a factor of 1.34 to 1, the model ranking 4th by a factor of 2 to 1, and the model ranking 10th by a factor of approximately 10.7 to 1. Given this, it seems unreasonable – and in fact misleading – to condition on one such specification and report estimates of the returns to a college degree using only that specification. The data have not revealed its preference for one particular model, and our point estimates of parameters of interest should incorporate this fact.

5.2 *Are there Nonlinearities in the Education-Log Wage Relationship?*

The results presented in Table 4 suggested that nonlinearities at 16 years of education were often present in the data, and also suggested a smaller role for nonlinearities upon completion of a high school degree. In Figure 1 below, we further investigate the existence of such ‘sheepskin effects’ in our model averaging framework. We provide OLS estimation results from two particular models in Figure 1A and 1B, and those obtained from our model averaging approach in Figure 1C.

The first graph (Figure 1A) plots the posterior mean of the expected log wage over different values of education using the baseline regression model that includes no ‘optional’ variables. Figure 1B adds two indicators capturing nonlinearities upon the completion of high school ($Ed \geq 12$) and college ($Ed \geq 16$), and Figure 1C contains the model averaged results. What we see from Figure 1B is that the model which includes nonlinearities in the regression specification clearly suggests ‘kinks’ at 12 and 16 years of schooling. These kinks are sufficiently important to clearly differentiate the shape of the education-log wage profile from the baseline regression estimate in Figure 1A. The point estimates themselves suggest rather strong effects, since the jump from 11 to 12 years of education results in an expected wage increase of approximately 8 percent. Given this, one might decide to label both of the nonlinearities as empirically important¹³ and thus report Figure 1B as an estimate of the education-log wage surface.

Of course, the results in Figures 1A and 1B are conditioned on particular models, and we could imagine many possible models – both nested and non-nested – which would suggest different profiles for the schooling – log wage relationship. To account for this, we plot our model averaged results in Figure 1C, which considers the set of all possible regressions with the 18 included variables.

We see at least two important points when comparing the figures. First, the model averaged results incorporate the data’s preference for a nonlinearity at 16 years of schooling. The nonlinearity at 12 years is assigned small enough weight so that education-log wage relationship is virtually linear over the interval to the left of 16 years of schooling. Second, the model averaged results have standard errors which are typically twice as large as those obtained from a particular

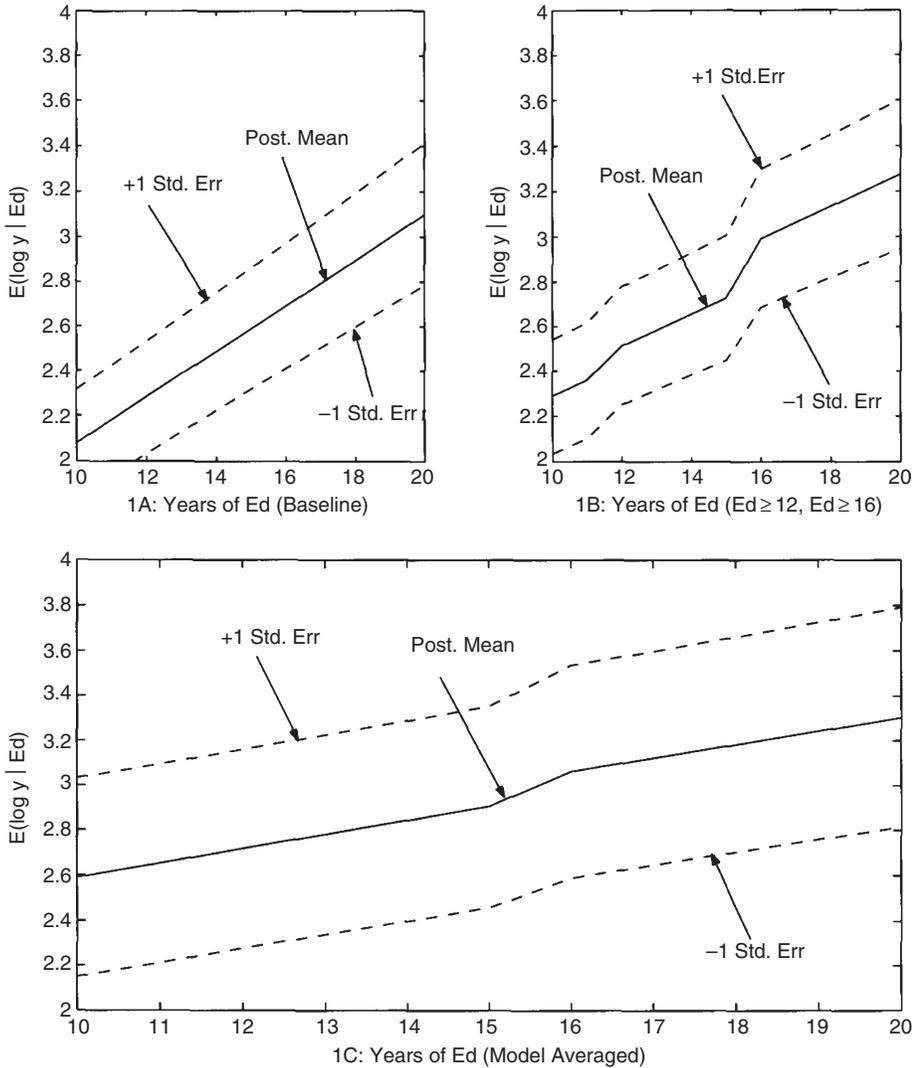


Figure 1. Posterior Means of Expected Log Wage Across Different Years of Schooling Using 1990 Data. Baseline Regression (1A), Nonlinearities at 12 and 16 (1B), and Model Averaged (1C) Results.

model. This source of uncertainty is automatically incorporated into the precision with which we are estimating our parameters of interest. Though not shown in the figure for the sake of brevity, results for the remaining years are quite similar, and continue to suggest a much larger role for the inclusion of a nonlinearity upon receipt of a four year degree.

5.3 *Have Returns to Schooling Increased Over Time?*

It has become a stylized fact in this literature that returns to schooling in the U.S. increased over the period of the 1980's (e.g. Blackburn and Neumark (1993), Murnane, Levy and Willett (1995), Grogger and Eide (1995), Heckman and Vytlačil (2001), Taber (2001)). To suggest a similar conclusion using our NLSY data, we present in Figures 2A and 2B posterior means and standard deviations of $\Delta_{16,12}$ over the period 1985–1989 (which overlaps with the years of study in the articles above). The results in Figure (2A) are obtained from the baseline regression model and the results in Figure 2B also contain an ability term (2B). In general, estimated returns to schooling are increasing over the time period, though the increase is very modest. If we were to regress the point estimates in Figure 2B on an intercept and a time trend, then we would predict that the receipt of a college degree increased wages by about 36 percent in 1985, and increased by .65 percent each year over the period.¹⁴ Further, we note that the estimate of the return to education is much smaller after including the ability variable, as the estimates obtained from Figure 2B are much smaller than those obtained in Figure 2A.

So again, the question arises: which model is the appropriate one to use? In Table 2 we provided evidence which suggested not only model uncertainty, but also some evidence that the data favored different models across the different years of study. These facts should be incorporated into our estimates of the return to education and the reported precision with which we are estimating these effects.

In Figure 2C, we provide year-by-year model averaged estimates of the return to education for 1985–1993. These point estimates suggest that returns to schooling were indeed increasing over the period until 1991, but then tend to fall over the last years of the sample. It is important to note that these estimates are not conditioned on particular specifications, but are obtained by letting the data decide how to weight the predictions obtained from a family of possible regressions. Overall, the results suggest that the return to a college degree increased wages from roughly 30–40 percent, and the highest returns were found in the early 90's. We also note that the posterior means are bounded away from zero, even after formally accounting for uncertainty in the model specification.

5.4 *Are returns to schooling concentrated among the most able?*

Some studies, such as Blackburn and Neumark (1993), Murnane Levy and Willett (1995), Grogger and Eide (1995) and Heckman and Vytlačil (2001) have investigated whether the increase in return to schooling over this period is attributable to a reward for additional schooling, or a changing premium for measured ability.

The typical approach in these studies is to allow for ability-education and ability-education-time interactions to determine if the growing return to education has been experienced primarily by those of highest ability. The general conclusion has been that after controlling for ability, the growth in the college wage premium is more modest, and Heckman and Vytlačil (2001) report that the college wage gap has been growing only for those in the highest ability quartile.

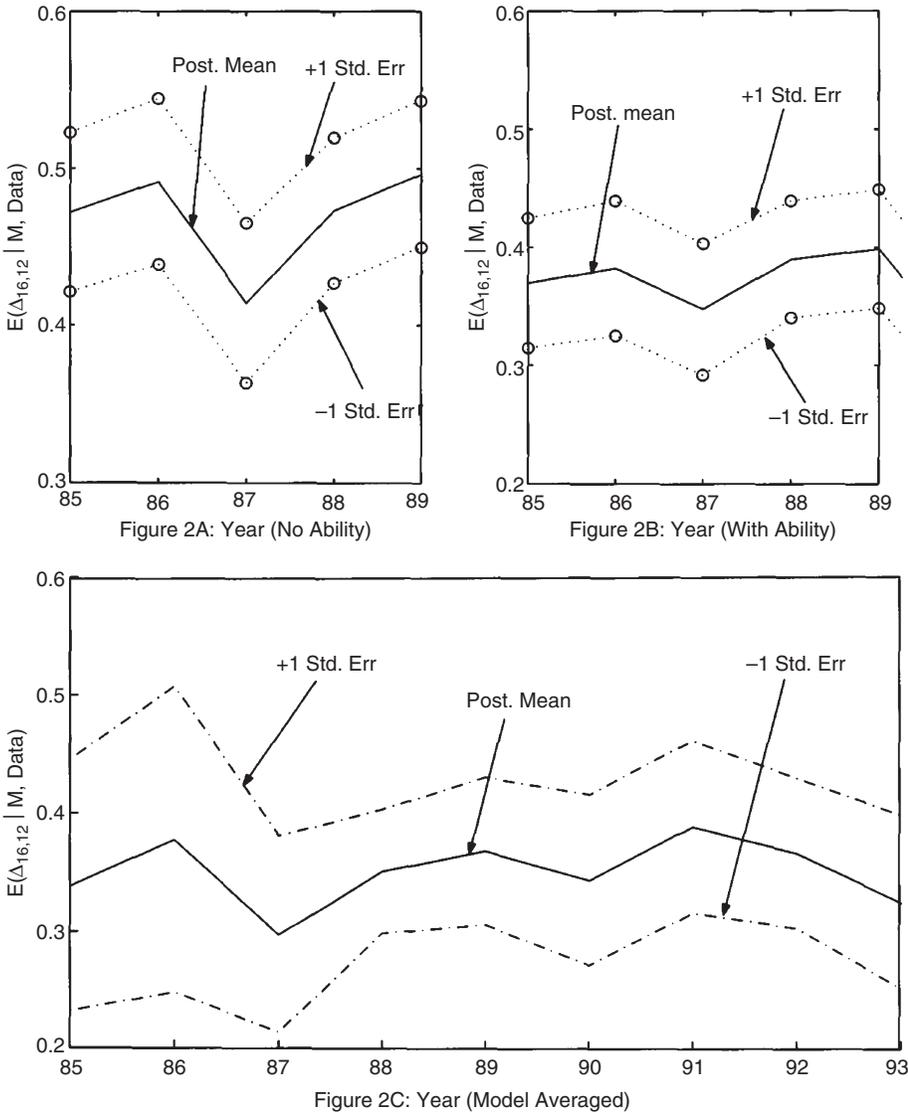


Figure 2. Posterior Means of $\Delta_{16,12}$ from 1985–1993. Baseline Regression (2A), Regression With Ability (2B), Model Averaged Results (2C).

In Figure 3 we plot posterior means and standard errors of $\Delta_{16,12}$ for those with ability equal to mean values, and those with ability values both one standard deviation below and one standard deviation above the mean. As evident from the figure, the posterior means can be ordered from year to year according to the ability values. That is, the highest ability individuals always have the highest

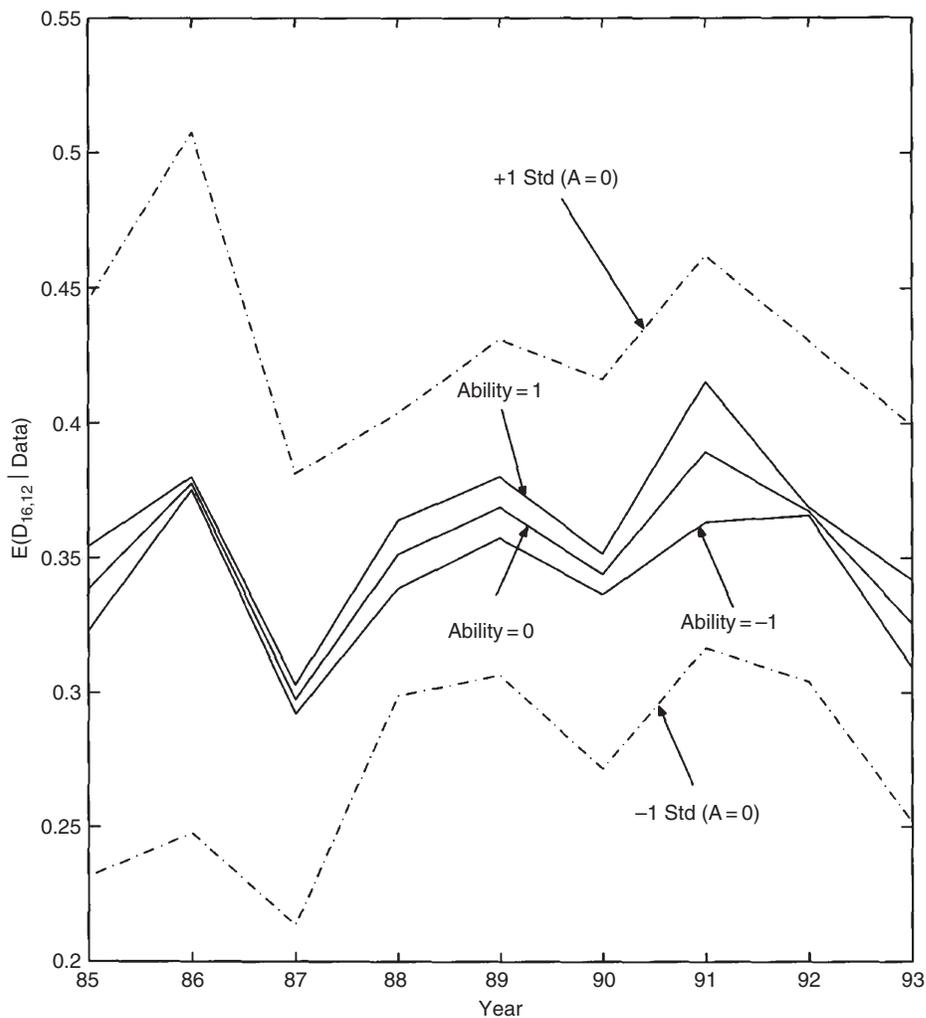


Figure 3. Model Averaged Posterior Means of $\Delta_{16,12}$ from 1985–1993 Evaluated at Ability = -1, 0, 1.

values of the expected return to a college education for each year studied. Further, we do not see convincing evidence that the gap is widening over time, but rather, it contracts and expands somewhat randomly across the different years.

Using our methods, we are also able to compute *year-by-year posterior probabilities that returns to schooling are concentrated among the more able*. These probabilities are reported in Table 6 below.

Table 6. $\Pr(\Delta_{16,12|A=1} > \Delta_{16,12} | \text{Data})$.

1985	1986	1987	1988	1989	1990	1991	1992	1993
.717	.601	.629	.739	.672	.635	.743	.561	.674

It is important to note that the probabilities reported in Table 6 incorporate the uncertainty in the model specification as well as parameter uncertainty within each model. When we incorporate this source of uncertainty, the probabilities exceed .5, though the largest in any year is .743. Thus, we have evidence that higher-ability individuals benefit more from a four year degree than those of average ability, though the posterior probability that this statement is true is clearly not overwhelming. When conditioning on one particular specification – as been done frequently in previous work – the probabilities could certainly be quite different than those in Table 6, and may suggest much stronger evidence that returns to schooling are concentrated among the most able. After accounting for uncertainty in the model itself, we also find evidence supporting the claim that returns to schooling are concentrated among the more able, though we can not make such a statement with strong confidence.

6. Conclusion

In this paper we investigated the effect of model uncertainty on estimates of the economic return to a college education. In so doing, we reviewed some of the previous literature on both estimation of the economic returns to schooling as well as Bayesian model selection. Since researchers have employed different specifications in the returns to schooling literature, and these different specifications yield different predictions about parameters of interest, our idea was to unite and review both of these literatures to obtain new and robust estimates of the economic return to a college degree.

Using these methods, we found that the NLSY data from the U.S. does not unanimously favor a particular model, and thus it is important to account for model uncertainty when estimating the returns to education. Specifically, we found that the data tend to favor the inclusion of a marriage indicator, a measure of cognitive ability, an ability-education interaction and a nonlinearity in education at 16 years of schooling. This nonlinearity at 16 years of education is most pronounced at the latter years of the sample, which is perhaps consistent with the notion of skill-biased technological change. The data are much less clear regarding the inclusion of the remaining variables which have occasionally appeared in previous work. Our approach, then, was to take the predictions from all models, weight each prediction by the posterior probability of that model, and average over the results to ‘integrate out’ the uncertainty in the model.

We found that over the period 1985–1993, the receipt of a Bachelor’s degree relative to a high school degree tended to increase hourly wages between 30–40

percent. We also found that point predictions of the return to a college education were highest for those of highest ability for all years studied.

Notes

1. These include the *American Economic Review*, *Econometrica*, *Economic Journal*, *Journal of Economic Literature*, *Journal of Economic Perspectives*, *Journal of Political Economy*, *Quarterly Journal of Economics*, *Review of Economic Studies* and *Review of Economics and Statistics*.
2. No titles were found to contain both words, and thus there are 229 citations to articles containing either 'education' or 'schooling.'
3. When studies included, say a quadratic or polynomial in education, we recorded these as controlling for the nonlinearities.
4. This approach is vaguely similar to 'meta-analysis' which is usually employed to average point estimates across studies by weighting the study-specific estimates by a function of their standard error (see Harmon *et al.* (2003), section 3.7 for a nice illustration). In this paper, we address a different question and examine the impact of specification uncertainty *within* a given study rather than *across* studies.
5. In this paper, we do not take up the issues of measurement error in schooling or the endogeneity of education and focus our attention on the issue of specification uncertainty in the context of the model above. Using the same NLSY data, Blackburn and Neumark (1995, p. 228) have addressed these issues and state, '...Once test scores are included in the regression, specifications tests provide little evidence that schooling is either endogenous or measured with error, or that ability is measured with error by the test scores.'
6. See section 2 of Harmon *et al.* (2003) for a nice description and derivation of this Mincer specification.
7. The reason for requiring that the variables in W must always appear in our analysis is to (1) limit the dimension of the problem to enable analytic calculation, and (2) use the theory of Mincer (1974) as a starting point which specifies the log wage is a linear function of education and a quadratic function labor market experience. Since all empirical studies use this Mincerian framework as a starting point (and the set of core variables in W would be included with probability approaching one anyway), we simply impose that the variables in W must be contained in all of our competing models.
8. See, for example Raftery, Madigan and Hoeting (1997) for a full description of the details. These details are also reviewed in the appendix of this paper.
9. Leamer (1978) first pointed out the important observation that model averaging also formally accounts for the uncertainty in the model specification.
10. Using quantile regression methods, Mwabu and Schultz (1996) document a similar finding, and report that 'ability' effects appear to vary over the education support, where 'ability' was not measured directly, but was captured in their residual.
11. This methodology of section 2 imposes the assumptions of normality and homoscedasticity, though some might question the appropriateness of such assumptions, particularly given the existing evidence on heavy-tailed wage distributions (*e.g.* Lydall (1968) and Heckman and Sedlacek (1990)). Though we do not discuss the details here, it is important to recognize that these assumptions can be relaxed using methods described in Carlin and Polson (1991), Geweke (1993), Albert and Chib (1993),

Wakefield *et al.* (1994), Tobias (2002) and Tobias and Li (2003), among others. When permitting departures from normality, no substantive differences were produced. Finally, we also examined how our results were affected by specifying priors that did not allow for the inclusion of higher-order interaction terms if the linear variables themselves did not appear in the model. Again, the substantive conclusions of our analysis were not substantially changed.

12. In 1985 and 1986, no observations contained in our sample completed *more* than 16 years of schooling, and thus $Ed * I(Ed \geq 16)$ is omitted from the set of optional variables.
13. The *t*-statistics for the nonlinearities at 12 and 16 were 1.40 and 2.62, respectively.
14. The *t*-statistic associated with the trend was 1.26.
15. Loosely, you can think of point estimates of these quantities as being similar to the OLS estimates obtained from the given model.

References

- Ahn, H. and Powell, J. (1993) Semiparametric Estimation of Censored Selection Models with a Nonparametric Selection Mechanism, *Journal of Econometrics* 58: 3–29.
- Albert, J. H. and Chib, S. (1993) Bayesian Analysis of Binary and Polychotomous Response Data, *Journal of the American Statistical Association* 88: 669–679.
- Altonji, J. G. and Dunn, T. A. (1996) Using Siblings to Estimate the Effect of School Quality on Wages, *Review of Economics and Statistics* 78(4): 665–671.
- Angrist, J. (1995) The Economic Returns to Schooling in the West Bank and Gaza Strip, *American Economic Review* 85(5), 1065–1087.
- Angrist, J. and Krueger, A. B. (1991) Does Compulsory School Attendance Affect Schooling and Earnings?, *Quarterly Journal of Economics* 106: 979–1014.
- Angrist, J. and Krueger, A. B. (1992) Estimating the Payoff to Schooling Using the Vietnam-Era Draft Lottery, *NBER Working Paper No. 4067*.
- Arias, O., Hallock K. F. and Sosa-Escudero, W. (2001) Individual Heterogeneity in the Returns to Schooling: Instrumental Variables Quantile Regression Using Twins Data, *Empirical Economics* 26: 7–40.
- Ashenfelter, O. and Krueger, A. (1994) Estimates of the Economic Return to Schooling from a New Sample of Twins, *American Economic Review* 84(5), 1157–1173.
- Ashenfelter, O. and Mooney, J. D. (1968) Graduate Education, Ability and Earnings, *Review of Economics and Statistics* 50(1), 78–86.
- Ashenfelter, O. and Rouse, C. (1998) Income, Schooling and Ability: Evidence from a New Sample of Identical Twins, *Quarterly Journal of Economics* 113(1), 253–284.
- Becker, G. and Chiswick, B. (1966) Education and the Distribution of Earnings, *American Economic Review* 56: 358–369.
- Behrman, J. R., Rosenzweig, M. R. and Taubman, P. (1996) College Choice and Wages: Estimates Using Data on Female Twins, *Review of Economics and Statistics* 78(4): 672–685.
- Behrman, J. R. and Wolfe, B. (1984) The Socioeconomic Impact of Schooling in a Developing Country, *Review of Economics and Statistics* 66(2): 296–303.
- Belman, D. and Heywood, J. S. (1991) Sheepskin Effects in the Returns to Education: An Examination of Women and Minorities, *Review of Economics and Statistics* 73(4): 720–724.
- Berger, J. O. and Pericchi, L. R. (2001) Objective Bayesian Methods for Model Selection: Introduction and Comparison, working paper, Duke University.
- Blackburn, M., and Neumark, D. (1993) Omitted Ability Bias and the Increase in the Return to Schooling, *Journal of Labor Economics* 11(3): 521–544.

- Blackburn, M. L. and Neumark, D. (1995) Are OLS Estimates of the Return to Schooling Biased Downward? Another Look, *Review of Economics and Statistics* 77(2): 217–230.
- Carlin, B. P. and Louis, T. A. (2000) *Bayes and Empirical Bayes Methods for Data Analysis* (Second Ed.) Boca Raton: Chapman & Hall.
- Carlin, B. P. and Polson, N. G. (1991) Inference for Nonconjugate Bayesian Models Using the Gibbs Sampler, *Canadian Journal of Statistics* 19: 399–405.
- DiNardo, J. and Tobias, J. L. (2001) Nonparametric Density and Regression Estimation, *Journal of Economic Perspectives* 15(4): 11–28.
- Dorfman, J. H. (1997) *Bayesian Economics Through Numerical Methods: A Guide to Econometrics and Decision-Making with Prior Information*, New York: Springer.
- Draper, D. (1995) Assessment and Propagation of Model Uncertainty, *Journal of the Royal Statistical Society, Ser. B* 57: 45–98.
- Fernandez, C., Ley, E. and Steel, M. F. J. (2001a) Model Uncertainty in Cross-Country Growth Regressions, *Journal of Applied Econometrics* 16(5): 563–576.
- Fernandez, C., Ley, E. and Steel, M. F. J. (2001b) Benchmark Priors for Bayesian Model Averaging, *Journal of Econometrics* (100)2: 381–427.
- Gelfand, A. E. and Dey, D. K. (1994) Bayesian Model Choice: Asymptotics and Exact Calculations, *Journal of the Royal Statistical Society, Ser. B* 56: 501–514.
- Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (1995) *Bayesian Data Analysis*, London: Chapman & Hall.
- Geweke J. (1993) Bayesian Treatment of the Independent Student t Linear Model, *Journal of Applied Econometrics* 8: 19–40.
- George, E. I. and McCulloch, R. E. (1993) Variable Selection Via Gibbs Sampling, *Journal of the American Statistical Association* 88: 881–889.
- George, E. I. and McCulloch, R. E. (1997) Approaches for Bayesian Variable Selection, *Statistica Sinica* 7: 339–373.
- Griliches, Z. (1977) Estimating the Returns to Schooling: Some Econometric Problems, *Econometrica* 45(1), 1–22.
- Griliches, Z. (1979) Sibling Models and Data in Economics: Beginnings of a Survey, *Journal of Political Economy* 87: S37–S64.
- Griliches, Z. and Mason, M. (1972) Education, Income and Ability, *Journal of Political Economy* 80(3): S74–S103.
- Grogger, J. and Eide, E. (1995) Changes in College Skills and the Rise in the College Wage Premium, *Journal of Human Resources* 30(2): 280–310.
- Harmon, C., Oosterbeek, H. and Walker, I. (2003) The Returns to Education: A Review of Evidence, Issues and Deficiencies in the Literature, *Journal of Economic Surveys* 17(2): 115–155.
- Hause, J. C (1972) Earnings Profile: Ability and Schooling, *Journal of Political Economy* 80(3): S108–S138.
- Heckman, J. J., Ichimura, H., Smith, J. and Todd, P. (1998) Characterizing Selection Bias Using Experimental Data, *Econometrica* 66(5): 1017–1098.
- Heckman, J. J., Layne-Farrar, A. and Todd, P. (1996) Human Capital Pricing Equations with an Application to Estimating the Effect of Schooling Quality on Earnings, *Review of Economics and Statistics* 78(4): 562–610.
- Heckman, J. J., and Sedlacek, G. L. (1990) Self-Selection and the Distribution of Hourly Wages, *Journal of Labor Economics* 8(1): S329–S363.
- Heckman, J. J., Tobias, J. L. and Vytlačil, E. (2001) Four Parameters of Interest in the Evaluation of Social Programs, *Southern Economic Journal* 68(2): 210–233.
- Heckman, J. J., Tobias, J. L. and Vytlačil, E. (2003) Simple Estimators for Treatment Parameters in a Latent Variable Framework, *Review of Economics and Statistics* 85(3): 748–755.
- Heckman, J. and Vytlačil, 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): 1–12.

- Heywood, J. S. (1994) How Widespread are Sheepskin Returns to Education in the U.S.? *Economics of Education Review* 13(3): 227–234.
- Hoeting, J. A., Madigan, D., Raftery, A. E. and Volinsky, C. T. (1999) Bayesian Model Averaging: A Tutorial, *Statistical Science* 14(4): 382–417.
- Hungerford, T. and Solon, G. (1987) Sheepskin Effects in Returns to Education, *Review of Economics and Statistics* 69(1): 175–177.
- Jaeger, D. and Page, M. (1996) Degrees Matter: New Evidence on Sheepskin Effects in Returns to Education, *Review of Economics and Statistics* 78(4): 733–740.
- Kass, R. E. and Raftery, A. E. (1995) Bayes Factors, *Journal of the American Statistical Association* 90: 773–795.
- Kass, R. E. and Wasserman, L. (1995) A Reference Bayesian Test for Nested Hypotheses and its Relationship to the Schwarz Criterion, *Journal of the American Statistical Association* 90: 928–934.
- Koop, G. (1994) Recent Progress in Applied Bayesian Econometrics, *Journal of Economic Surveys* 8: 1–34.
- Korenman, Sanders D. and David Neumark (1991) Does Marriage Really Make Men More Productive?, *Journal of Human Resources* 26: 282–307.
- Lam, D. and Schoeni, R. F. (1993) Effects of Family Background on Earnings and Returns to Schooling: Evidence from Brazil, *Journal of Political Economy* 101(4): 710–740.
- Leamer, E. (1978) *Specification Searches: Ad Hoc Inference with Nonexperimental Data*, New York: John Wiley & Sons.
- Lydall, H. (1968) *The Structure of Earnings*, Oxford: Clarendon Press.
- Madigan, D. and Raftery, A. E. (1994) Model Selection and Accounting for Model Uncertainty in Graphical Models Using Occam's Window, *Journal of the American Statistical Association* 89(428): 1535–1546.
- Madigan, D. and York, J. (1994) Bayesian Graphical Models for Discrete Data, *International Statistical Review* 63(2): 215–232.
- McLachlan and Peel (2000) *Finite Mixture Models*, New York: Wiley.
- Miller, P., Mulvey, C. and Martin, N. (1995) What do Twins Studies Reveal About the Economic Returns to Education?, *American Economic Review* 85(3): 586–599.
- Mincer, J. (1958) Investment in Human Capital and Personal Income Distribution, *Journal of Political Economy* 66: 281–302.
- Mincer, J. (1970) The Distribution of Labor Incomes: A Survey With Respect to the Human Capital Approach, *Journal of Economic Literature* 8(1): 1–26.
- Mincer, J. (1974) *Schooling, Experience and Earnings*, New York: Columbia University Press.
- Morgan, J. and David, M. (1963) Education and Income, *Quarterly Journal of Economics* 77(3): 423–437.
- Morgenstern, R. D. (1973) Direct and Indirect Effects on Earnings of Schooling and Socio-Economic Background, *Review of Economics and Statistics* 55(2): 225–233.
- Mwabu, G. and Schultz, T. P. (1996) Education Returns Across Quantiles of the Wage Function: Alternative Explanations for Returns to Education by Race in South Africa, *American Economic Review* 86(2): 335–339.
- Murnane, R., Willett, J. and Levy, F. (1995) The Growing Importance of Cognitive Skills in Wage Determination, *Review of Economics and Statistics* 77(2): 251–266.
- Poirier, D. J. (1985) Bayesian Hypothesis Testing in Linear Models with Continuously Induced Conjugate Priors Across Hypotheses: In *Bayesian Statistics 2*, J.M Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, eds. Amsterdam: North-Holland, 711–722.
- Poirier, D. J. (1991) A Bayesian View of Nominal Money and Real Output Through a New Classical Macroeconomic Window, (with discussion) *Journal of Business and Economic Statistics* 9: 125–161.
- Poirier, D. J. (1995) *Intermediate Statistics and Econometrics: A Comparative Approach*, Cambridge: MIT Press.

- Raftery, A. E., Madigan D. and Hoeting J. A. (1997) Bayesian Model Averaging for Linear Regression Models, *Journal of the American Statistical Association* 92(437): 179–191.
- Taber, C. (2001) The Rising College Premium in the Eighties: Return to College or Return to Unobserved Ability?, *Review of Economic Studies* 68(3): 665–691.
- Tobias, J. L. (2002) Model Uncertainty and Race and Gender Heterogeneity in the College Entry Decision, *Economics of Education Review* 21(3): 211–219.
- Tobias, J. L. (2003) Are Returns to Schooling Concentrated Among the Most Able? A Semiparametric Analysis of the Ability-Earnings Relationships, *Oxford Bulletin of Economics and Statistics* 61(1): 1–29.
- Tobias, J. L. and Li, M. (2003) A Finite Sample Hierarchical Analysis of Returns to Schooling Across Public High Schools, *Journal of Applied Econometrics* 18: 315–336.
- Wakefield, J. C., Smith, A. F. M., Racine-Poon, A. and Gelfand, A. E. (1994) Bayesian Analysis of Linear and Nonlinear Population Models Using the Gibbs Sampler, *Journal of the Royal Statistical Society, Ser. C* 43: 201–221.
- Wasserman, L. (2000) Bayesian Model Selection and Model Averaging, *Journal of Mathematical Psychology* 44: 92–107.
- Weisbrod, B. A., and Karpoff, P. (1968) Monetary Returns to College Education, Student Ability and College Quality, *Review of Economics and Statistics* 50: 491–497.
- Willis, R. J. and Rosen, S. (1979) Education and Self-Selection, *Journal of Political Economy* 87(2): S7–S36.
- Zellner, A. (1986) On Assessing Prior Distributions and Bayesian Regression Analysis with g Prior Distributions. In *Bayesian Inference and Decision Techniques – Essays in Honor of Bruno de Finetti*, Goel, P. K. and Zellner, A. eds, Amsterdam: North-Holland, 233–243.

Appendix: Model Averaging Details

From (2), the problem faced in this paper is a standard problem of variable selection in the context of a typical log-wage equation:

$$y = W\alpha + X\beta + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2 I_N), \quad (4)$$

with W denoting a fixed set of explanatory variables and X denoting an $n \times k$ matrix of k ‘optional’ covaraites.

Let us entertain T different competing models, and let $\mathcal{M} = \{M_1, M_2, \dots, M_T\}$ be the set of all of these competing models. In our particular application, we think of the set of all possible models as simply the set of all possible regressions we could run with k potential explanatory variables. Finally, let θ be some parameter of interest whose interpretation is constant across our models. From the law of total probability,

$$p(\theta|y) = \sum_{i=1}^T p(\theta|M_i, y)\Pr(M_i|y), \quad (5)$$

where M_i indexes the different models, p generically refers to a density function and y denotes the observed data.

The above equation states that the posterior distribution of θ which is *not conditioned on a particular model* is simply obtained as a weighted average of the

model-specific posterior distributions $p(\theta|M_i, y)$, where the terms $\Pr(M_i|y)$ serve as the weights. These weighting terms are called *model probabilities*, as they simply give the posterior probability in support of model M_i among the class of competing models in \mathcal{M} .

The model-specific posterior distributions $p(\theta|M_i, y)$ are easily obtained from standard Bayesian procedures (see, e.g. Poirier (1995)).¹⁵ To obtain these posteriors it is necessary to discuss the priors employed. We first standardize all variables to have mean zero and unit variance and then choose conjugate prior specifications of the form:

$$\beta \sim N(\mu_\beta, \sigma^2 V_\beta), \quad \frac{v\lambda}{\sigma^2} \sim \chi_v^2,$$

with $N(a,b)$ denoting a normal density with mean a and variance b , and χ_v^2 denoting a Chi-square density with v degrees of freedom. The parameters μ_β , V_β , v and λ are hyperparameters to be chosen by the researcher.

Values for the hyperparameters are chosen in a manner similar to the recommendation of Raftery, Madigan and Hoeting (1997) (denoted RMH). Once the variables are standardized, these authors list a set of desired results that the prior should satisfy, and consequently we arrive at the use of the following values:

$$V_\beta = 2.85^2 I_k, \quad \lambda = .28, \quad v = 2.58, \quad \mu_\beta = 0.$$

In generated data experiments, RMH show that this prior produced model averaged results that were quite successful at identifying the true model when it was contained as a competing model (pg. 188), and also had better predictive performance than any individual model that could have reasonably been chosen (section 5.2). Alternate benchmark prior specifications, based on the use of g -priors (Zellner (1986)) have been recently proposed by Fernandez, Ley and Steel (2001b). We found that results obtained from the RMH prior were robust in the sense that we arrived at similar conclusions using a variety of different reasonable prior selection rules, including a g -prior based on a hold-out sample, as well as specifying prior hyperparameters using estimation results obtained after fitting the same model on a sample of black males.

Getting back to the posterior model probabilities on the right-hand side of (5), we note that these can be obtained as follows:

$$\Pr(M_i|y) = \frac{p(y|M_i)p(M_i)}{p(y)} = \frac{p(y|M_i)p(M_i)}{\sum_{k=1}^K \Pr(y|M_k)p(M_k)}. \tag{6}$$

The terms $p(y|M_i)$ on the right-hand side of (6) are called *marginal likelihoods*, as they can be interpreted as the marginal density for the data that does not depend on any of the regression parameters. The terms $p(M_i)$ are simply prior model probabilities. A default choice (which is the one employed throughout this paper) is to simply assume that all of the competing regression models are equally probable *a priori* and thus $p(M_i) = p(M_j) = 1/T$.

The marginal likelihoods can be calculated using a variety of different methods. Fortunately for this particular regression application, it turns out that the marginal likelihoods have a simple closed-form solution. The explicit expression for the marginal likelihoods in this context is (*e.g.* RMH (1997, page 178))

$$p(y|M) = c(v, \lambda, V_\beta, X) \left(\lambda v + (y - X\mu_\beta)'(I_n + XV_\beta X')^{-1}(y - X\mu_\beta) \right)^{-(v+n)/2}, \quad (7)$$

where

$$c(v, \lambda, V_\beta, X) = \frac{\Gamma[(v+n)/2](v\lambda)^{v/2}}{\pi^{n/2}\Gamma[v/2]|I_n + XV_\beta X'|^{1/2}}.$$

It is important to recognize that the design matrix X , the covariance matrix V_β and the prior mean vector μ_β will change across the different models as different optional variables are included or excluded.

Thus, to summarize, the marginal likelihoods as in (7) are calculated for each model, thus enabling us to determine the posterior model probabilities as in (6). These probabilities are then used to weight the model-specific predictions to obtain our model averaged results.