JOURNAL OF APPLIED ECONOMETRICS
J. Appl. Econ. **18**: 315–336 (2003)
Published online 22 October 2002 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/jae.696

A FINITE-SAMPLE HIERARCHICAL ANALYSIS OF WAGE VARIATION ACROSS PUBLIC HIGH SCHOOLS: EVIDENCE FROM THE NLSY AND HIGH SCHOOL AND BEYOND

JUSTIN L. TOBIAS* AND MINGLIANG LI
Department of Economics, University of California-Irvine, 3151 Social Science Plaza, Irvine, CA 92697-5100, USA

SUMMARY

Using data from both the National Longitudinal Survey of Youth (NLSY) and High School and Beyond (HSB), we investigate if public high schools differ in the "production" of earnings and if rates of return to future education vary with public high school attended. Given evidence of such variation, we seek to explain why schools differ by proposing that standard measures of school "quality" as well as proxies for community characteristics can explain the observed parameter variation across high schools. Since analysis of widely-used data sets such as the NLSY and HSB necessarily involves observing only a few students per high school, we employ an exact finite sample estimation approach. We find evidence that schools differ and that most proxies for high school quality play modest roles in explaining the variation in outcomes across public high schools. We do find evidence that the education of the teachers in the high school as well as the average family income associated with students in the school play a small part in explaining variation at the school-level. Copyright © 2002 John Wiley & Sons, Ltd.

1. INTRODUCTION

In an important paper which seemed to initiate a resurgence in the 'school quality' literature, Card and Krueger (1992) [henceforth CK] argued that the quality of schooling positively affects the rate of return to future education. Using census data and a two-step estimation approach, they found: (1) returns to schooling varied by state of birth, and (2) aggregated state-level proxies for school quality could partly explain the observed variation in returns to schooling across states of birth. Their results suggested that school quality 'matters', and in particular, that the quality of schooling raises the rate of return to future education.

The work by Card and Krueger was followed by a host of interesting work on this topic (e.g. Betts, 1995, 1996; Grogger, 1996a,b; Heckman, Layne-Ferrar and Todd, 1996; Hanushek, Rivkin and Taylor, 1996; Eide and Showalter, 1998; Hanushek, Kain and Rivkin, 1998; Hoxby, 1998, 2000; Figlio, 1999; Olson and Ackerman, 1999, among others). One conclusion which emerged from several of the earlier studies was that state-level estimates, such as those obtained by Card and Krueger (1992), are likely to be upward biased due to their inability to simultaneously control for other omitted state-level variables. Further, the model of CK was also criticized for allowing the proxies for school quality to enter the wage equation only through an interaction with education, and not as independent linear variables.

In this paper, we continue to investigate this important topic and analyse several variants of the CK model using data taken from the National Longitudinal Survey of Youth (NLSY) and

* Correspondence to: Justin L. Tobias, University of California-Irvine, Department of Economics, 3151 Social Science Plaza A, Irvine, CA 92697-5100, USA. E-mail: jtobias@uci.edu

Received 29 January 2001
Revised 1 May 2002

High School and Beyond (HSB). Given previous criticism of estimates obtained using aggregated, state-level data, we employ these two widely used data sets which enable estimation of quality effects at the school level. Further, we also estimate models which allow quality effects to enter both through intercepts as well as through interactions with education.

Since studies using micro-level HSB and NLSY data are typically limited to a small number of observations per school, the school-level models employed in past work have often been simple linear regressions, where the estimated quality effects are defined as the coefficients associated with the quality variables. That is, school-level studies have not attempted to re-estimate the CK (1992) model, or close variants of that model in which returns to schooling would vary by high school attended (rather than state of birth), presumably because we observe only a few students per high school in these widely used data sets. Rather than abandon estimation of models which will enable us to address such issues using school-level data, we instead pursue a finite-sample Bayesian alternative.

In particular, we describe models that enable us to address the following questions: (1) Do public high schools differ in the 'production' of students' earnings? (2) Do returns to schooling vary across public high schools? and (3) Is the observed variation across public high schools attributable to observed differences in school 'quality' or proxies for community characteristics?

The models we employ will need to account for specific features of our data—that (in the NLSY) we observe individuals over time, and that outcomes are observed for individuals attending the same high school. As shown in Section 2, hierarchical models are ideally suited for such a situation. Further, estimation of such models is ideally handled in a Bayesian setting since we observe only a small number of students in each high school in these data sets.

Throughout this analysis our outcome variable of interest is the log of hourly wages. To permit heavy-tailed conditional log wage distributions as found by Lydall (1968) and Heckman and Sedlacek (1990), we add inverted-gamma mixing variables to extend our analysis to the family of Student-*t* distributions. In terms of key conclusions, we find that student outcomes vary according to public high school attended, and that standard measures of school 'quality' generally play a small role in explaining the observed parameter variation across high schools. We do, however, find some evidence that the education of teachers in the high school as well as the average family income of students in the high school play a part in explaining school-level parameter variation.

In the following section, we describe our most general Bayesian hierarchical model, and discuss its similarities (and differences) to previous specifications employed in the literature. Section 3 describes the different data sets used and the procedure for constructing the school indicators from the NLSY. Section 4 presents the empirical results and the paper concludes with a summary in Section 5.

2. THE MODEL

The most general model we employ in this paper is the following hierarchical model:[1]

[1] Here, we let $N(a, b)$ denote a normal density with mean a and variance b, and IG and W denote Inverted Gamma and Wishart densities, respectively. Throughout this paper, we parameterize the inverted gamma as follows: if $x \sim IG(a, b)$ then $p(x) \propto x^{-(a+1)} \exp[-1/(bx)]$ so that $E(x) = 1/[b(a-1)]$, $a > 1$, $b > 0$. We also parameterize the Wishart density in (10) so that R is approximately the prior mean of the covariance matrix Σ, while ρ is a 'degrees of freedom' parameter.

$$y_{ist}|x_{ist}, \zeta_{is}, \pi, \sigma_\varepsilon^2 \overset{ind}{\sim} N(\zeta_{is} + x_{ist}\pi, \sigma_\varepsilon^2) \tag{1}$$

$$\zeta_{is}|z_{is}, w_{is}, \theta_s, \beta, \lambda_{is}, \sigma_\zeta^2 \overset{ind}{\sim} N(z_{is}\theta_s + w_{is}\beta, \lambda_{is}\sigma_\zeta^2) \tag{2}$$

$$\theta_s|Q_s, \theta, \Sigma \overset{ind}{\sim} N(Q_s\theta, \Sigma) \tag{3}$$

$$\lambda_{is}|\nu \overset{iid}{\sim} IG(\nu/2, 2/\nu) \tag{4}$$

$$\pi|\pi_o, V_\pi \sim N(\pi_0, V_\pi) \tag{5}$$

$$\beta|\beta_0, V_\beta \sim N(\beta_0, V_\beta) \tag{6}$$

$$\theta|\theta_0, V_\theta \sim N(\theta_0, V_\theta) \tag{7}$$

$$\sigma_\varepsilon^2|\kappa_1, \kappa_2 \sim IG(\kappa_1, \kappa_2) \tag{8}$$

$$\sigma_\zeta^2|\tau_1, \tau_2 \sim IG(\tau_1, \tau_2) \tag{9}$$

$$\Sigma^{-1}|\rho, R \sim W(\rho, \rho R) \tag{10}$$

$$\nu \sim p(\nu) \tag{11}$$

The model above is deemed 'hierarchical' since equations (1)–(3) consecutively model relationships at 'smaller' levels of dependence, and some right-hand side parameters in the current stage are also used as dependent variables in the subsequent stage. At the first layer of the hierarchy (equation (1)), we simply present our longitudinal log wage equation where the *ist* subscript denotes variables and parameters for individual i at time t that was observed to previously attend high school s. We let y_{ist} denote the log hourly wage of individual i at time t who attended high school s, and x_{ist} denote a $1 \times k$ vector of time-varying characteristics affecting log wages, such as labour market experience. The parameter ζ_{is} represents an individual-specific effect, capturing differences in outcomes across individuals through differences in intercepts.[2] We carry along the s subscript in this individual effect to remind us that several different individuals i attended the same high school s.

In the second level of the hierarchy (equation (2)) we explain variation at the individual level. Specifically, we explain variation in the person-specific effects ζ_{is} by centering them over time-invariant characteristics of the individual (w_{is} and z_{is}). The characteristics in w_{is} include variables such as family characteristics, and measured cognitive ability.[3] That is, we suppose that an individual from the first stage might possess a large fixed effect because that individual has, say, high ability or was raised in a wealthy, well-educated family.

In z_{is}, we include variables whose parameters are permitted to vary at the school level. Since both the NLSY and HSB data sets offer multiple observations on individuals within the same school s, we are able to obtain estimates of parameters that vary across schools. To investigate issues raised in previous work, equation (2) introduces school-specific intercepts and returns to education (through

[2] Note that marginalized over the fixed effects, outcomes for a given individual are correlated over time, though we have assumed no contemporaneous correlation across individuals for simplicity.

[3] In this analysis, education is time-invariant since we condition only on those individuals who have completed their schooling, and thus the values of education do not change throughout the sample period. As for ability, only one test (the ASVAB) is available in the NLSY, while for the sophomore cohort in the HSB, two sets of test scores are available. For our HSB analysis, we use only the 1982 test scores for the sophomores and treat this as our time-invariant ability measure.

θ_s), and we seek to learn about the extent of heterogeneity across schools in the 'production' of earnings and rates of return to future education. Formally, then, in equation (2) we specify

$$\theta_s \equiv \begin{bmatrix} \alpha_s \\ \delta_s \end{bmatrix}, \quad z'_{is} \equiv \begin{bmatrix} 1 \\ Ed_{is} \end{bmatrix}$$

where Ed denotes years of schooling completed. Thus, α_s denotes the intercept for those attending high school s, while δ_s represents the return to schooling for those attending high school s.

Equation (3) seeks to explain why intercepts and returns to schooling vary across high schools. In this equation, we include observed characteristics in Q_s in an attempt to explain why outcomes would vary across public high schools. We do this by centering the 2×1 vector of school-specific intercept and return to education parameters over measures of the 'quality' of the school, as well as proxies for the characteristics of the community comprising the school. On this latter point, one might be tempted to label a school as 'good' simply because the majority of students in the school come from wealthy, well-educated families and thus tend to be successful in the labour market. This would not represent a school quality effect, but picks up a neighbourhood or community effect that is capable of explaining why the students of some schools are observed to be more successful than students of other schools. To this end, we attempt to separate out the effects of inputs of the school (measures of school quality) from neighbourhood or community characteristics by including both as explanatory variables.

We do this by constructing average values of parental education or income within the school, as well as average ability values of the students within the school and use these as explanatory variables to control for 'community characteristics'. As discussed in Section 3, the HSB data typically enables us to obtain at least 60 observations per school from the 1980 base-year interviews, so that the constructed average characteristics should provide a reasonably accurate assessment of the community comprising each high school. Formally, we accomplish this by defining:

$$Q_s \equiv \begin{bmatrix} Q_{1s} & 0 \\ 0 & Q_{2s} \end{bmatrix}, \quad \theta \equiv \begin{bmatrix} \theta^{INT} \\ \theta^{RTS} \end{bmatrix}$$

Both Q_{1s} and Q_{2s} are row vectors containing quality variables and community characteristics that may explain variation in the school-specific intercepts and returns to education (respectively). We let θ^{INT} denote those parameters associated with the quality and community variables used to explain variation in the school-specific intercepts, and θ^{RTS} be the corresponding parameters associated with the high school-specific returns to schooling. For the proxies for high school quality, we will include measures such as pupil–teacher ratio, district-level expenditure per student, number of books per student in the school's library, and average teacher characteristics in the school. Because of differences in the variables available in the NLSY and HSB, the quality (and other) variables used will be slightly different across the two data sets. A complete description of the variables used is provided in Section 3.

2.1. Other Observations Regarding the Model

The reader might have noted that each of the subsequent stages in the hierarchy could be sequentially substituted into the previous stage, thus producing a single elaborated model at the first stage. So why not present and estimate only this 'reduced-form' equation at the first stage? Our representation of the model provided in (1)–(11) offers several improvements over the reduced-form representation. Specifically, we prefer to analyse the model in this form for the following three reasons:

(1) We would like to remain consistent with the spirit of Card and Krueger's original specification and the hierarchical efforts employed in previous work, (e.g. Rumberger and Thomas, 1993).
(2) The representation in equations (1)–(11) provides a clear way to visualize various levels of dependence in the data, and, perhaps most importantly,
(3) We are interested in obtaining finite-sample inference about the school-level parameters α_s and δ_s, and thus do not want to 'substitute them out' into the first stage.

Some other features and assumptions of the model are worth discussing. The specifications in (4) - (11) provide priors for the remaining parameters of the model, and with the exception of ν in (4), involve hyperparameters to be chosen by the researcher. We have also added the Inverted Gamma (IG) mixing variables (λ) to the disturbance term variance at the second stage of the hierarchy to allow for departures from normality.[4] Formally, letting $p(\lambda_{is}|\nu) \overset{iid}{\sim} IG(\nu/2, 2/\nu)$ as we have in equation (4), and letting u_{is} denote the error term implied from (2), we note that

$$p(u_{is}|\nu, \sigma_\zeta^2) = \int_0^\infty p_N(u_{is}|\lambda_{is}, \sigma_\zeta^2) p_{IG}(\lambda_{is}|\nu) \mathrm{d}\lambda_{is} \sim t_\nu(0, \sigma_\zeta^2)$$

a univariate Student-*t* distribution with ν degrees of freedom. Thus, marginalized over the mixing variables, the disturbances have a Student-*t* distribution. This modelling feature is added to allow for the possibility that the conditional log wage distribution may have heavier tails than those implied by the normal, as found empirically by Lydall (1968), and Heckman and Sedlacek (1990), among others. We have chosen to add these mixing variables at the second stage of the hierarchy as in Wakefield *et al*. (1994) to allow for possible outlying individuals or unequal variances across individuals.

We also treat the degrees of freedom parameter ν as a parameter to be determined within our model, and specify a prior p for v in equation (11). For this application, we discretize the set of possible values for the degrees of freedom parameter and draw ν values from the set {2, 3, 4, 5, 6, 7, 8, 10, 12, 16, 24, 32}. By choosing a reasonably fine grid over integer-valued ν, we allow many possibilities for the tail-behaviour of the error terms, and let the data determine the most appropriate set of distributional assumptions.[5] In practice, we choose our prior for the degrees of freedom parameter p to be uniform over these discrete elements of ν.

It is important to recognize that the model above is similar to that employed in Card and Krueger (1992) where s referred to *state* of birth rather than *school* attended. When extending such an analysis to the school level and making use of the widely used HSB and NLSY data, we necessarily obtain a small number of observations per school. We combat this 'problem' by simply employing an exact finite-sample Bayesian estimation approach, and describe our algorithm for fitting this highly parameterized model in Section 4 and in further detail in the Appendix. We also note that the addition of the subsequent stages of the hierarchy tend to 'shrink' coefficients towards a common mean, and thus the posterior distributions of our person-specific and school-specific

[4] See, for example, Andrews and Mallows (1974), Carlin and Polson (1991) and Geweke (1993) among others.
[5] Alternatively, we could relax the discretization assumption, and draw ν values using a Metropolis-within-Gibbs step. By choosing a fine grid over a wide range of ν values, our discretization assumption should not be very restrictive.

effects combine information obtained from the given individual or school with information obtained from all individuals or schools.[6]

It is also important to recognize that in applications of hierarchical modelling, some studies—including Card and Krueger (1992)—proceed to fit the model 'sequentially', and this sequential approach might not always be justified.[7] That is, in the sequential approach, we would obtain point estimates of the individual-level fixed effects ζ_{is} from a regression in (1), use these point estimates to obtain the school-level θ_s parameters in (2), and then use these estimated parameters to estimate θ in (3). When writing down the joint likelihood of the complete model, it becomes evident (see the Appendix) that estimation of, say, ζ_{is} incorporates the information in both the first and second stages, and thus, it may not be advisable to obtain estimates of these parameters using information only from the first stage. Specifically, and as noted before, the second stage tends to 'shrink' the individual estimates towards a common mean, and when such parameters are not estimated precisely from the first stage, the contribution of the second stage will be non-ignorable. Further, the sequential approach requires 'correcting' reported standard errors, since the dependent variable used in the current stage was estimated from the previous stage. Our Bayesian estimation approach briefly outlined in Section 4, and described completely in the Appendix, automatically and appropriately handles all of these issues.

3. THE DATA SETS

The data used to determine the extent and source of variation across public high schools in the 'production' of earnings are taken from the National Longitudinal Survey of Youth (NLSY) and High School and Beyond (HSB).

3.1. The NLSY

The National Longitudinal Survey of Youth (NLSY) is a panel study of young men and women ranging in age from 14 to 22 in 1979. Individuals in the NLSY are annually re-interviewed, providing a wealth of information on education, earnings, family characteristics, and other labour market outcomes. Importantly for our purposes, the NLSY contains a supplemental high school survey in which principals or supervisors of high schools attended by the 1979 NLSY participants respond to a series of questions regarding their high schools. The completed questionnaires provide information on school-level quality variables such as teacher education and pupil–teacher ratios. These key quality variables are then included as explanatory variables in equation (3) to attempt to explain why schools differ.

Due to oversampling within neighbourhoods we are able to obtain multiple observations per high school within the NLSY. Unfortunately, the NLSY does not contain a high school identifier variable. Such a variable can be constructed, however, by matching school-level characteristics provided in the school survey as well as another demographic variable.[8]

[6] For more on the properties of such shrinkage estimators, and examples of shrinkage predictions versus other predictions see, for example, Zellner and Hong (1989), Baltagi, Griffin and Xiong (2000) and Tobias (2001).

[7] In Card and Krueger's analysis, however, the information provided by the first stage should 'overwhelm' information in the second stage, and thus CK's resulting estimates should not be affected.

[8] To my knowledge, the only other study that has created these identifiers by 'reverse engineering' using NLSY data is Betts (1995).

We thus create our high school indicator variables by matching on three different sources of information provided in the NLSY. We first match on a set of 12 school characteristics which include school enrollment, number of teachers in the school, number of full-time counsellors in the school, number of books in the school's library, percentage of students in the school who are White, Asian, Black and Hispanic, the percentage of faculty who are White, Black or Hispanic, and the percentage of students in the high school who enter their sophomore year but drop out prior to graduation.

We also match on a second set of characteristics describing the curriculum offered by the high school. In particular, we assign individuals to the same high school provided their schools offered (or failed to offer) programmes in the following seven areas: agricultural occupations, business or office occupations, distributive education, health occupations, home economics occupations, trade or industrial occupations, and technical occupations. Finally, to further safeguard against combining students from different schools into the same high school, we also match by reported state of residence of the individual at age 14. Students are classified as belonging to the same high school only if they agree on all values of this extensive list of school-level variables as well as reported state of residence at age 14.

Given the constructed school indicators, we restrict the sample to white males reporting hourly wages between \$1 and \$100 in the given year, and analyse outcomes over the period 1988–1993.[9] We also impose the requirements that at least 2 observations per individual are available over this period, that education did not vary over time for any individual, and that at least 3 people are observed in each high school. This creates a total of 98 different public high schools attended by 481 different individuals for a total of 2493 person-year observations. Thus, on average, we observe each individual for 5.2 of the 6 years of the panel, while each high school contains an average of 4.9 students.

As our dependent variable, we use the log of hourly wages in real 1990 dollars. For our set of time-varying explanatory variables (X), we include potential labour market experience and its square. For the individual-level time invariant characteristics, we include highest grade completed by the respondent's mother,[10] an 'ability' measure, and high school school-specific intercepts and education variables. The ability (or test score) measure is constructed from the 10 component tests of the ASVAB battery which is given to the NLSY participants. Since performance on the 10 component tests is increasing in age, each of the 10 component tests is first residualized on age, and our ability index is then defined as the first principal component of those standardized residuals, which is then standardized.[11] Without such an ability measure, one could confound school effects with ability or family background effects if individuals are ability-sorting into high schools. Finally, for the school quality variables, we include pupil–teacher ratio (defined as the number of students enrolled in the school by the number of teachers in the school) teacher education (percentage of teachers in the high school with at least a Master's degree), and number of books per student in the school's library. To control for community factors, we also include the average education of the mothers within the school, as well as the average ability of students within the school. As noted in Section 2, these community characteristics are added

[9] In the early years of the NLSY survey, many individuals are still enrolled in school, and thus we focus on the 6-year period from 1988 to 1993 as our panel.

[10] Other family characteristics, such as family income and father's education further reduced the sample size, and thus we include only mother's education. In the HSB analysis, we are able to include a broader range of family characteristics.

[11] See Cawley *et al.* (1997), DiNardo and Tobias (2001) and Heckman, Tobias and Vytlacil (2001) for more on the construction and use of this ability measure.

to equation (3) in an attempt to separate the effects of school quality from neighbourhood or community effects.[12]

3.2. High School and Beyond

High School and Beyond (HSB) is a survey conducted on behalf of the National Center for Education Statistics, and was constructed with the intent of yielding a sample of students that are representative of the population of American high school students. In the 1980 base-year interview, nearly 30,000 high school sophomores and an equal number of seniors distributed among approximately 1000 different US high schools were interviewed. Approximately 15,000 sophomores and 12,000 seniors of the original sample were then selected for follow-up interviews. The selected senior and sophomore subsamples were re-interviewed biennially until 1986, and the sophomore subsample was also re-interviewed in 1992. The follow-up interviews provide important information on educational attainment, employment status, and earnings.

To gather as many high schools and observations per high school as possible, and to obtain information on individuals who are more likely to have completed their schooling, we focus on outcomes reported in the 1986 interview, and pool the sophomore and senior cohorts together.[13] We restrict the sample to white males attending public high schools who report to be working full time in 1986, and also report hourly wages between \$1 and \$100 per hour. We also exclude observations where other key covariates are missing. Further imposing the requirement that at least two students are observed in each high school produces a sample of 1599 observations from 371 different public high schools. If we require 3 students per high school, 1467 observations from 305 different schools are obtained.[14] The number of observations available per school range from 1–15, and 68% of the high schools analysed contain 3-8 (white male) students per high school. Thus, it is important to emphasize that the HSB data provides many more high schools to be analysed than the NLSY (and thus, effectively more observations to estimate the parameters of equation (3)), though both data sets contain only a small number of individuals within each school.[15]

The HSB data also provides an elaborated set of measures of school 'quality', and some of these quality variables coincide with those used in the NLSY. In our HSB analysis, we include pupil–teacher ratio, books per student in the school's library, percentage of teachers with at least a Master's degree, percentage of faculty who have been at the school for at least 10 years (Teacher Senior), and district-level expenditure per pupil. To control for 'community effects' associated with the school, we go back to the 1980 base-year interviews and compute the average education of parents within each school, average parental income, and average test scores of the students in the school. Since we are typically able to obtain at least 60 observations per school from the 1980 base-year interviews, the constructed average characteristics should provide a reasonably accurate assessment of the community comprising each high school.

[12] We do not take up the potential issue of aggregation bias here. See Rivkin (2001) for further discussion of this issue.

[13] Using 1992 interview data enables us to look at older individuals, but only provides information on outcomes of the sophomore cohort. Further, the earnings variables available in the 1992 interview are not current, but are 'historical', as the respondents provide information on previous annual earnings from 1987 to 1991, thus introducing an increased likelihood of measurement error in reported earnings.

[14] When allowing for school-specific intercepts and returns to education, we require variation in educational attainment within each school. This produces a total of 1096 observations from 226 public high schools, which remains significantly larger than the number of schools obtained with the NLSY data.

[15] We are grateful to a referee for suggesting the additional analysis of the HSB data.

It is also important to note that unlike the NLSY, the HSB data as we employ it here is not a panel data set, and thus the model given in (1)–(11) does not directly apply. This does not create any additional problems, but only simplifies the model presented in (1)–(11). Specifically, the individual-level parameters ζ_{is} at the first stage of the hierarchy must be removed when using the HSB data, since we are not analysing outcomes for a given individual over time. So, we can think of the HSB models as starting with equation (2), where z_{is} still contains an intercept and education whose associated parameters are permitted to vary across schools, w_{is} contains controls for 'ability' and family characteristics, and the dependent variable is the observed log wage (y_{is}). In other words, we observe a cross-section of log hourly wages for individuals attending different high schools, and thus begin our HSB analysis with equation (2). While the HSB data we employ does not provide a panel, we are able to obtain more high schools with the HSB data and to analyse more measures of school quality (including expenditure per student).[16]

4. EMPIRICAL RESULTS

In this section, we present estimation results using both the NLSY[17] and HSB data. To fit the model described in Section 2, we employ the Gibbs sampler—a simulation-based algorithm which involves iteratively sampling from the complete posterior conditional distributions. Under certain regularity conditions, the draws produced by successively sampling from these conditional distributions converges to drawing from the joint posterior distribution itself, which is obtained from (1)–(11) via Bayes' theorem and is provided in the Appendix. For all models, we run the Gibbs sampler for 50,000 iterations, and discard the first 10,000 as the burn-in period. In practice, we employ blocking or grouping steps to reduce the autocorrelation in the resulting parameter chains. For reference, the joint posterior and the complete posterior conditionals are provided in the Appendix.[18]

4.1. Model 1: Do Schools Differ? [$Q_{1s} = Q_{2s} = 1$]

Since it is perhaps premature to explain why schools differ before establishing that differences exist, we first estimate a model in which Q_{1s} and Q_{2s} contain only intercepts. In such a model, we obtain only the common intercept and education return means across schools, as well as the variability of these parameters across schools. In later models, we will attempt to find variables

[16] As pointed out by a referee, with the NLSY data, a first-year high school student would be assigned the same values of school quality as a senior student, though they may have faced different levels of quality during their 4-year education. This may introduce measurement error in school quality. To address this issue, we repeated the analysis using only those students within each school that were one and two years apart, and thus were more likely to face the same quality of schooling. None of the substantive results were affected. Such a problem could also exist in HSB as we pool the senior and sophomore cohorts together. We believe this problem to be minor, as it seems unlikely that schools would undergo drastic changes over a two-year period. Nonetheless, it is important to recognize this potential limitation of our approach.

[17] Initial estimation using the NLSY data revealed a preference for normality, and thus we do not add the inverted-gamma mixing variables and degrees of freedom parameter to the NLSY model. We also find little evidence of correlation between α_s and δ_s and thus specify separate equations for these parameters in (3). For reference, we continue to permit such correlation with the HSB data, though again no evidence of strong correlation is obtained. Finally, a time trend or time dummies were also added to x_{ist} with the NLSY data, producing no change in the qualitative conclusions.

[18] The priors employed here are centred over values that seemed reasonable to us, and were consistent with results of previous studies, yet were 'non-informative' or 'flat' enough so that the data information dominates. Details are available upon request.

Table I(a). Coefficient posterior means, standard deviations and probabilities of being positive: Model 1, HSB data

Level of Hierarchy	Variable	Post. mean	Post. std	$\Pr(\cdot > 0 \mid D)$
Time-invariant characteristics (β)	Test score	−0.000619	0.0134	0.482
	Number of siblings	−0.00734	0.00535	0.0844
	Parent education	−0.0160	0.00491	0.000575
	Family income ($1000)	0.00486	0.00101	1.000
	Experience	0.165	0.0354	1.000
	Experience^2	−0.00828	0.00401	0.0186
Common school intercept (θ^{INT})	Intercept	−0.135	0.251	0.295
Common school return to education (θ^{RTS})	Education	0.126	0.0140	1.000
Variance parameters	$\Sigma_{1,1}$	0.00530	0.00292	1.000
	$\Sigma_{1,2}$	0.000116	0.000962	0.542
	$\Sigma_{2,2}$	0.00115	0.000630	1.000
	σ_ε^2	0.108	0.00880	1.000
Degrees of freedom	ν	4.750	0.701	1.000

Table I(b). Coefficient posterior means, standard deviations and probabilities of being positive: Model 1, NLSY data

Level of hierarchy	Parameter/ variable	Post. mean	Post. std	$\Pr(\cdot > 0 \mid D)$
Time-varying characteristics (π)	Experience	0.0840	0.0115	1.00
	Experience^2	−0.003	0.0005	0.000
Time-invariant characteristics (β)	Test score (ability)	0.141	0.023	1.00
	Mother education	0.0048	0.010	0.698
Common school intercept (θ^{INT})	Intercept	0.553	0.128	1.000
Common school return to education (θ^{RTS})	Education	0.089	0.010	1.000
Variance parameters	$\sigma_\alpha^2(\Sigma_{1,1})$	0.004	0.001	1.000
	$\sigma_\delta^2(\Sigma_{2,2})$	0.00013	0.00003	1.00
	σ_ζ^2	0.126	0.010	1.000
	σ_ε^2	0.101	0.003	1.000

which explain why the schools differ, but our preliminary goal here is to simply document that student outcomes do indeed tend to vary across high school attended.[19]

Presented in Tables I(a) and I(b) are estimation results from this model using both the High School and Beyond (1A) and NLSY (1B) data sets. To assess if schools do indeed differ, the most relevant parameters are the variance parameters of the covariance matrix Σ. If these parameters are 'small', then we would have posterior support for little variation in outcomes across public

[19] In previous work, Betts (1995) concluded that schools differ by adding a set of high school dummies, and rejecting the joint null hypothesis of equality of these dummy variable coefficients. Grogger (1996b) using HSB data allowed for school-specific random effects and also finds strong evidence of variation in outcomes at the school level.

high schools. Conversely, large values of these parameters concentrated in regions away from zero indicate that an important portion of the overall disturbance variance can be attributed to unexplained variation at the school level.

Starting with the HSB analysis in Table I(a),[20] we see evidence suggesting that schools do indeed differ, and in particular, that rates of return to future education vary with high school attended. The posterior means of θ^{RTS} and σ_δ[21] were 0.126 and 0.034, respectively. To interpret these results, the second stage of the hierarchy would imply that the return to an additional year of schooling for each high school is independently 'drawn from' a normal distribution with mean 0.126 and standard deviation of 0.034. Clearly, this suggests substantial variability in returns across public high schools. The NLSY data suggests a similar, though less variable result. The NLSY results suggest that returns are normally distributed around a mean of 0.089 with a standard deviation of 0.011. Again, this suggests that returns to schooling do indeed vary across public high school attended, as the draws from this second-stage are not concentrated (or nearly degenerate) around a common mean, but rather, seem to vary across high schools.

Another way to 'diagnose' if schools differ is to simply compute the posterior probability that, say, the return to schooling for each school exceeds the overall return, i.e. $\Pr(\delta_s > \overline{\delta}|\text{Data})\ \forall s$. For each data set, we define the overall (mean) return to be the average of all the posterior means of each school, (i.e. $\overline{\delta} \equiv [1/S]\sum_{s=1}^{S} E(\delta_s|\text{Data})$) and treat this as a known quantity. This calculation is of interest since it also addresses the potential limitations of inference obtained with only a small number of observations available within each school. If the small number of observations within each school did not enable precise estimation of the school-specific parameters, we might expect such probabilities to be close to 0.5 for all schools, as the school specific posteriors would be quite uninformative or 'flat'. If the above probabilities were all close to 0.5, we might have evidence that there is just not enough information in the data to determine if some schools are better than others (or alternatively, that schools do not appear to differ). When computing these probabilities using both sets of data, we found that $\Pr(\delta_s > \overline{\delta}|\text{Data})$ ranged from 0.2 to 0.75 using the HSB data, and 0.03 to 0.88 using the NLSY data. This suggests that at least some schools differ, as some of them are either very unlikely or reasonably likely to beat the average return to an additional year of education. As for variation in intercepts, we found similar evidence of heterogeneity using the HSB data. The probabilities that each school exceeded the overall intercept mean ranged from 0.1 to 0.86. The NLSY, however, revealed a slightly different result, as the probabilities of exceeding the overall mean ranged from 0.32 to 0.63.[22] Further evidence of school-level heterogeneity can be obtained when *individually* comparing the posterior distributions of the α_s and δ_s parameters across schools. As will be shown in Section 4.3 using our 'full model' the δ_s posteriors differ

[20] For the HSB, we use the constructed parental education and test score variables. Parental education is taken as the maximum of the mother's and father's education over the 1980 and 1986 interviews. We also tried different specifications by adding mother's education and father's education separately. These different models produced virtually identical results and no changes to the key questions addressed here. The test score variable is an average of a reading test score, a vocabulary test score, and the first part of a math test given to the sophomores and seniors. Again, results were not found to be sensitive to the construction of the test score. For example, when using only the first part of the math test as our ability measure (which has the largest proportion of questions which are common to both the sophomores and seniors), similar results were obtained.

[21] Note that σ_δ is the square root of the variance parameter, and thus is not the coefficient reported in Table I(a).

[22] Since the intercepts only capture baseline differences at zero years of education, heterogeneity in returns to education across schools will generate differences in the 'performances' of high schools. Outcomes for individuals with some schooling will then clearly depend on the public high school attended by that individual.

across public high schools, suggesting that rates of return to future education vary by public high school attended.

4.2. Model 2: Quality Through Intercepts Only: [$\delta_s = \delta$]

In this section, we abstract from the possibility of school-specific returns to education, and simply account for heterogeneity across schools by allowing intercepts to differ across the schools. This model is of interest since it is essentially the quality-through-intercepts model that has been estimated numerous times since Card and Krueger (1992) (e.g. Rumberger and Thomas, 1993; Grogger, 1996a,b and, in fact, was investigated prior to CK: e.g. Johnson and Stafford (1973) and Nechyba (1990)).[23] In this model, the quality variables make their appearance in the log wage equation only as separate linear variables after substituting everything back into the first stage. However, the analysis of this paper goes a step further than the basic quality-through-intercepts model, since it enables us to recover information regarding the school-specific intercepts or 'fixed effects' in addition to the roles of the quality variables themselves.

Results of this analysis for both the HSB and NLSY are presented in Tables II(a) and II(b), respectively. At the third stage (equation (3)) using the HSB data, we include a common school intercept, indicators if the high school is located in a suburban or rural area,[24] pupil–teacher ratio,

Table II(a). Coefficient posterior means, standard deviations and probabilities of being positive: Model 2, HSB data

Level of hierarchy	Variable	Post. mean	Post. std	$\Pr(\cdot > 0 \mid D)$
Time-invariant characteristics (β)	Test score	0.00230	0.0138	0.566
	Number of siblings	−0.00444	0.00538	0.202
	Parent education	−0.0139	0.00513	0.00317
	Family income($1000)	0.00386	0.00106	1.000
	Education	0.121	0.0138	1.000
	Experience	0.159	0.0349	1.000
	Experience^2	−0.00800	0.00396	0.0211
Quality variables and community proxies affecting school intercepts (θ^{INT})	Intercept	0.270	0.399	0.751
	Suburban	0.0788	0.0464	0.954
	Rural	−0.0125	0.0479	0.398
	Pupil–teacher ratio	−0.000742	0.00341	0.414
	Book/student	−0.00155	0.000924	0.0474
	Teacher education	0.00133	0.000646	0.980
	% teacher senior	−0.000914	0.000629	0.0709
	District expenditure ($1000)	0.00735	0.0227	0.627
	Average father ed.	−0.0415	0.0311	0.0914
	Average mother ed.	−0.00214	0.0389	0.479
	Average family income ($1000)	0.00984	0.00586	0.954
	Average test score	−0.0189	0.0250	0.221
Variance parameters	σ_α^2 $(\Sigma_{1,1})$	0.00405	0.00234	1.000
	σ_ε^2	0.111	0.00836	1.000
Degrees of freedom	ν	4.916	0.717	1.000

[23] Eide and Showalter (1998) use quantile regression but do not permit an education-quality interaction within the quantiles. Other studies, such as Betts (1995, 1996), Hanushek, Rivkin and Taylor (1996) and Heckman, Layne-Ferrar and Todd (1996) have introduced the quality variables both as levels and as interactions with education.

[24] Urban is the excluded category.

Table II(b). Coefficient posterior means, standard deviations and probabilities of being positive: Model 2, NLSY data

Level of hierarchy	Variable	Post. mean	Post. std	$\Pr(\cdot > 0 \mid D)$
Time-varying characteristics (π)	Experience	0.079	0.0140	1.00
	Experience^2	−0.003	0.0006	0
Time-invariant characteristics (β)	Education	0.078	0.012	1.00
	Test score (ability)	0.149	0.024	1.00
	Mother education	0.0001	0.010	0.506
Quality variables and community proxies affecting school intercepts (θ^{INT})	Intercept	0.6700	0.2458	0.999
	Pupil–teacher ratio	−0.0011	0.0029	0.352
	Teacher education	0.0010	0.0005	0.968
	Books/student	−0.0002	0.0007	0.391
	Average mother ed.	0.0104	0.0125	0.798
	Average test score	0.0116	0.0258	0.671
Variance parameters	$\sigma_\alpha^2(\Sigma_{1,1})$	0.004	0.001	1.000
	σ_ζ^2	0.132	0.011	1.000
	σ_ε^2	0.101	0.003	1.000

books per student in the school's library, percentage of teachers with at least a Masters degree (Teacher Education), percentage of teachers who have been at the school for at least 10 years (Teacher Senior) and District Level Expenditure (in \$1000s of dollars) as our quality variables. As stated in Section 2, to attempt to account for variation in performance across schools that would be attributable to community effects, we also include average parent and student characteristics within the school in equation (3). Specifically, we construct and include the average education of the fathers and mothers in the school, average family income (in \$1000s of dollars) of students in the school as well as average student test scores within the school.

The results shown in Table II(a) (HSB) indicate that schools located in suburban areas with smaller average class sizes, with more educated teachers, spending more per student and with high average family incomes are the schools that perform best in the production of earnings. Somewhat surprisingly, the coefficients associated with the books per student and teacher senior variables were often perversely signed (and in some cases had very low posterior probabilities of being positive).[25] This finding is, however, consistent with previous work which finds little effect of proxies for school quality measured at the school level (e.g. Betts, 1995, 1996; Grogger, 1996a,b; Hanushek, Rivkin and Taylor, 1996). For the quality coefficients that possess the expected signs, the actual sizes of the implied impacts of the quality and community variables are typically quite small, and the associated posterior standard deviations are often large. For example, lowering average pupil–teacher ratio by 10 (a large reduction!) increases hourly wages by only 0.74%, while increasing average annual family income associated with the school by \$1000 increases hourly wages by only 1%.

[25] Our initial reaction was that more experienced teachers might be better at transmitting skills to students, and that schools with a large number of experienced teachers may also be those schools with a beneficial learning environment, since the teachers have chosen to remain at the given high school. In this sense, we would expect the coefficient on teacher senior to be positive. However, the expected sign of this coefficient is certainly less clear than those associated with the other school quality variables.

Some stronger and more 'significant' effects are found for teacher education, which is measured as the percentage of teachers in the high school with at least a Master's degree. A one-standard-deviation increase in teacher education (which corresponds to raising the percentage of teachers with a Masters degree by 23 in both HSB and the NLSY) results in a 3% expected increase in hourly wages. Some evidence is also provided that the addition of the quality and community variables helps to explain some of the parameter variation across public high schools—the posterior mean of σ_α^2 has reduced by 23%—dropping from 0.0053 in Table I(a) to 0.0041 in Table II(a).

Results from the NLSY tell a similar story. As with HSB, we find that the posterior means of the pupil–teacher ratio and teacher education variables are of the expected sign, while the books per student coefficient is nearly centred at zero, and the posterior mean of this coefficient is actually negative (the posterior probability that increasing the number of books per student in the school's library will have a positive effect on wages is 0.39, as shown in the table). Again, consistent with the HSB analysis, we see evidence of a positive teacher education effect. The posterior probability that an increase in teacher education results in an increase in hourly wages was found to be 0.98 in HSB and 0.97 in the NLSY. The size of the coefficient in the NLSY suggests that a one-standard-deviation (23 unit) increase in teacher education increases hourly wages by approximately 2.3%. The proxies for community characteristics in the NLSY, which include average education of the mothers in the school and average test scores of the students in the high school also have a positive (though not strongly positive!) effect in explaining the parameter variation across high schools. However, it is important to note that adding these controls for school quality and the proxies for community characteristics explained only two percent of the variance that was left unexplained in Model 1, which included only an intercept at the third stage.

4.3. Model 3: Full Model

In this section, we present results of our full model where we have allowed our measures of school quality as well as our proxies for community characteristics to affect both the school-specific intercepts and rates of return to future education. In CK's original model, quality was introduced only as an interaction with education, while here, we permit quality effects both through intercepts and returns to schooling.

Results of this analysis are presented in Tables III(a) (HSB) and III(b) (NLSY). While the results in Model 2 interpreted the quality variables only as levels, the quality effects must now be interpreted over the education support. This elaboration of the model introduces some very interesting results.

First, consider the contribution of community characteristics in explaining parameter variation across high schools using the HSB data. Clearly the most influential of these is average family income within the school. While the average family income coefficient has a high probability of being positive through the intercept (level), it also has a (relatively) high probability of being negative through the interaction with education. What this means is that the effect of community-level average family income is larger for those students with lower values of education. As shown in Table III(a), the overall effect of average family income remains positive over the majority of the education support, and declines to zero at approximately 19 years of education. At 12 years of schooling, and holding all other variables constant, a $ 1000 increase in the average family income of students in the high school leads to a 1.3% expected increase in post-schooling hourly wages. Clearly, then, the community proxies—particularly average family income—seem to play some role in explaining why schools are observed to differ.

Table III(a). Posterior means, standard deviations and probabilities of being positive: Model 3, HSB data

Level of hierarchy	Variable	Post. mean	Post. std	$\Pr(\cdot > 0 \mid D)$
Time-invariant characteristics (β)	Test score	0.000575	0.0137	0.518
	Number of siblings	−0.00429	0.00536	0.211
	Parent education	−0.0142	0.00509	0.00235
	Family income ($1000)	0.00380	0.00107	1.000
	Experience	0.153	0.0354	1.000
	Experience^2	−0.00725	0.00399	0.0339
Quality variables and community proxies affecting school intercepts (θ^{INT})	Intercept	−1.445	1.978	0.234
	Suburban	0.408	0.323	0.899
	Rural	0.160	0.334	0.685
	Pupil–teacher ratio	−0.0254	0.0253	0.159
	Book/student	−0.00274	0.00696	0.344
	Teacher education	−0.00623	0.00471	0.0914
	% teacher senior	0.00434	0.00486	0.815
	District expenditure ($1000)	0.0229	0.171	0.555
	Average father ed.	−0.000938	0.168	0.498
	Average mother ed.	0.0716	0.205	0.636
	Average family income ($1000)	0.0360	0.0381	0.828
	Average test score	−0.0484	0.185	0.396
Quality variables and community proxies affecting school-specific returns to education (θ^{RTS})	Intercept	0.249	0.148	0.955
	Suburban	−0.0246	0.0240	0.151
	Rural	−0.0120	0.0249	0.315
	Pupil–teacher ratio	0.00186	0.00190	0.835
	Book/student	0.0000893	0.000523	0.569
	Teacher education	0.000583	0.000357	0.948
	% teacher senior	−0.000408	0.000370	0.137
	District expenditure ($1000)	−0.00115	0.0128	0.463
	Average father ed.	−0.00292	0.0127	0.409
	Average mother ed.	−0.00572	0.0155	0.356
	Average family income ($1000)	−0.00189	0.00284	0.252
	Average test score	0.00235	0.0140	0.567
Variance parameters	$\Sigma_{1,1}$	0.00440	0.00252	1.000
	$\Sigma_{1,2}$	0.000163	0.000874	0.571
	$\Sigma_{2,2}$	0.00125	0.000693	1.000
	σ_ε^2	0.106	0.00854	1.000
Degrees of freedom	ν	4.720	0.675	1.000

Intuitively, the fact that the 'return' to average family income within the school declines as one acquires more education also seems sensible-highly educated labour is very mobile, while those individuals with less education are more likely to remain attached to the area in which they were educated. Since average family income in the community is likely to be positively correlated with the overall local labour market conditions of the community, and those with low education are also more likely to be tied to that community—it seems natural that average family income would have the largest effect for those with low education. Indeed, this is the result suggested by Table III(a).

The results are reversed when interpreting the teacher education variable in both HSB and the NLSY. Both data sets suggest rather strongly that teacher education effects are increasing over the education support (the posterior probability of a positive teacher education—student education interaction is 0.95 in the HSB data and 0.96 in the NLSY). Again, we can construct a reasonable story to rationalize this result. Highly educated teachers may be more successful at transmitting

Table III(b). Posterior means, standard deviations and probabilities of being positive: NLSY data, full model

Level of hierarchy	Variable	Post. mean	Post. std	Prob. positive
Time-varying characteristics (π)	Experience	0.081	0.0119	1.00
	Experience^2	−0.0028	0.0005	0.000
Time-invariant characteristics (β)	Test score (ability)	0.151	0.024	1.00
	Mother education	−0.0041	0.010	0.351
Quality variables and community proxies affecting school intercepts (θ^{INT})	Intercept	0.867	0.246	1.00
	Pupil–teacher ratio	−0.0007	0.0033	0.420
	Teacher education	0.0001	0.0007	0.546
	Books/student	−0.0001	0.0007	0.475
	Average mother ed.	0.0012	0.0154	0.538
	Average test score	0.0036	0.0334	0.537
Quality variables and community Proxies affecting school-specific returns to education (θ^{RTS})	Intercept	0.0621	0.0255	0.991
	Pupil–teacher ratio	−0.0002	0.0006	0.375
	Teacher education	0.0002	0.0001	0.961
	Books/student	−0.0001	0.0003	0.361
	Average mother ed.	0.0004	0.0019	0.691
	Average test score	0.0009	0.0049	0.576
Variance parameters	$\sigma_\alpha^2(\Sigma_{1,1})$	0.004	0.001	1.000
	$\sigma_\delta^2(\Sigma_{2,2})$	0.00013	0.00003	1.000
	σ_ζ^2	0.125	0.010	1.000
	σ_ε^2	0.101	0.003	1.000

skills to students, and these skills may raise the rate of return to future education. This would imply a positive coefficient on the teacher education—student education interaction, as documented in Tables III(a) and III(b) using both the NLSY and HSB. At 12 years of schooling, the HSB results suggest that a one-standard-deviation (23-unit) increase in teacher education results in a 1.7% increase in hourly wages. A similar calculation in the NLSY suggests that a one-standard-deviation increase in teacher education results in an expected increase in hourly wages equal to approximately 3.2%.[26] Thus, both the NLSY and HSB data sets provide evidence of a teacher education effect that increases with educational attainment. In other words, a highly educated high school faculty may indeed raise the rate of return to the future education of its students. The remaining quality variables do not appear to play a significant role in explaining school-level parameter variation in the NLSY data. As for the HSB, suburban high schools continue to perform better than urban high schools over the majority of the education support, while a one-standard-deviation decrease in pupil-teacher ratio has a positive effect on wages for those of 14 years of schooling or less, but then becomes negative at additional years of education.

Finally, we again provide evidence that high schools differ using our full model and thus unite the results of this section with those obtained in Section 4.1.[27] In Figure 1, we randomly select 25 different high schools from the NLSY and present boxplots of the school-specific

[26] Though this number is larger than that obtained with the HSB data, it is important to recognize that its standard errors are also larger, perhaps owing to the fact that we have fewer high schools to analyse in the NLSY.

[27] One might question whether our key results, such as the existence of heterogeneity across schools, are sensitive to the prior employed. These key conclusions were robust to the specification of the prior. For example, using the HSB data and Model 2, and imposing a different prior which sets the mean and standard deviation of $\sigma_\alpha^2 = 0.0008$ produces a posterior mean and standard deviation of σ_α^2 equal to 0.002 and 0.0016, respectively. Thus, after imposing a prior that favours

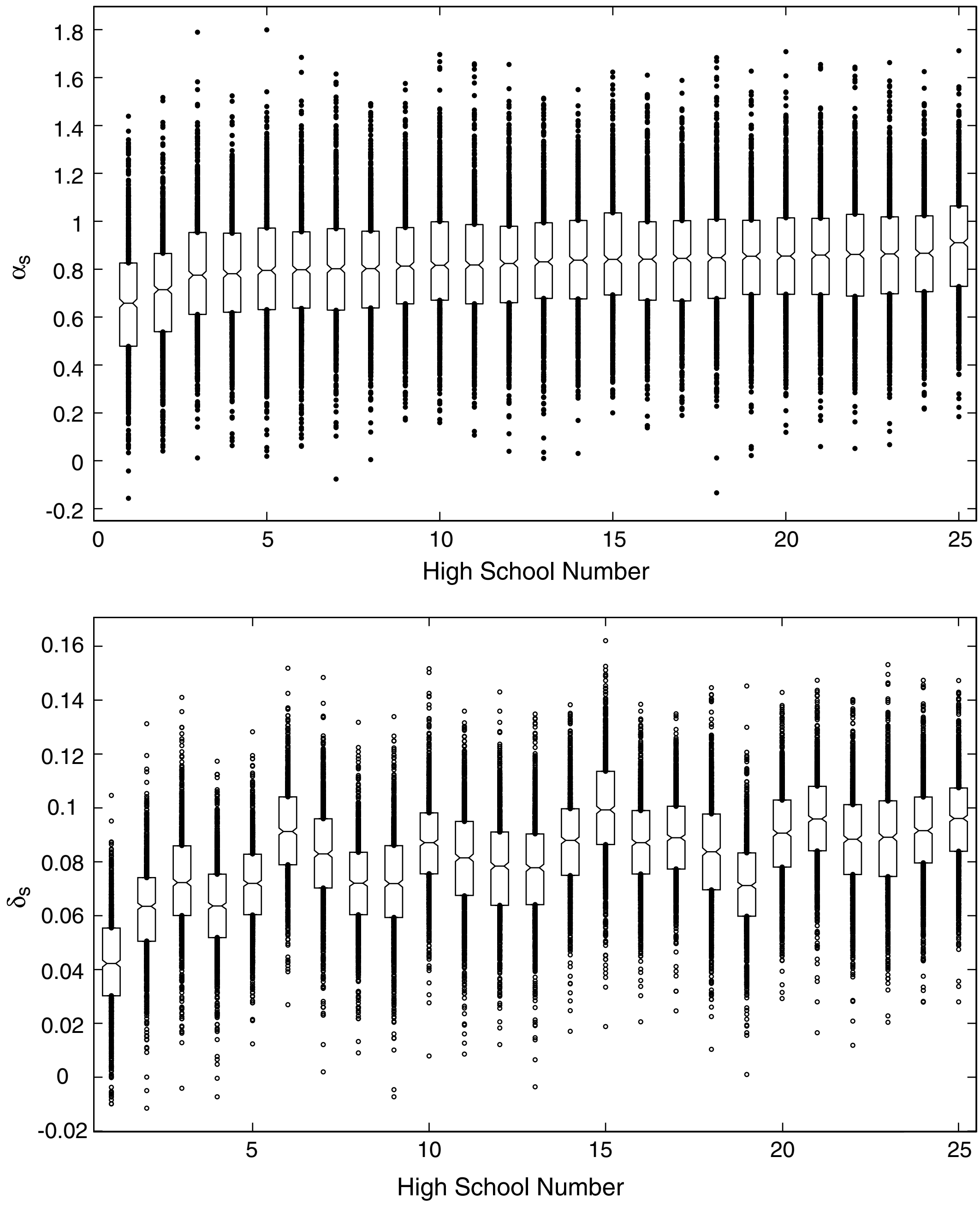

Figure 1. Boxplot of school-specific intercepts and returns to schooling (marginal posteriors of $\{\alpha_s\}$ and $\{\delta_s\}$ parameters): 25 randomly selected high schools from the NLSY

homogeneity across schools, the data pulls us toward a revised belief which reveals more heterogeneity at the school level, and the posterior also becomes tighter around the larger mean.

intercepts and returns to education.[28] In the top portion of Figure 1, the intercept estimates are presented, and we have arranged the posteriors in ascending order by sorting them according to their posterior medians. The lower portion of Figure 1 presents the corresponding school-specific return to education parameters. It is important to keep in mind that the boxplots can be compared vertically—for example, school 13 refers to the same high school in both the top and bottom graphs.

We first see that the intercept parameters do not appear to be overwhelmingly different from one another, as suggested previously in the analysis of Model 1 using the NLSY data.[29] The school-specific intercept posteriors often overlap considerably, and there does not appear to be strong evidence of heterogeneity in intercepts across schools. More interesting, however are the return to schooling parameters, which appear to be quite different across schools. As shown in the bottom half of Figure 1, the interquartile ranges of the 'best' and 'worst' schools often have no common intersection. This suggests that rates of return to future education clearly vary by public high school attended. It is also interesting to note that the returns to schooling in the lower half of the figure are not 'ordered' as they are in the top half, and with the exception of the worst-performing schools, appear to vary somewhat randomly around the overall mean. This is consistent with our finding of little correlation between the α_s and δ_s parameters, as reported in Tables I(a)–III(a).

5. CONCLUSION

In a seminal paper, Card and Krueger (1992) argued that returns to schooling varied by state of birth and that state-level variation in returns to schooling can be partly explained by state-level measures of school 'quality'. In this paper, we re-examined variants of the important work by Card and Krueger using micro-level quality data from the NLSY and HSB. Although the transition to micro-level data of this type necessarily limits us to small sample sizes, we escape the problem of aggregation bias (Hanushek *et al.*, 1996), and simply adopt a Bayesian approach to provide exact finite-sample inference.

We show that the model estimated by Card and Krueger can be regarded as a hierarchical model, and describe a Bayesian method for estimating such a model given that we observe only a few students per public high school. Using both data sets, we find that schools do indeed vary in the production of earnings, and that rates of return to future education do appear to vary across public high school attended.

A natural question which is of considerable policy interest is to try to find characteristics which help us to explain why schools differ. We did this by proposing that the parameter variation across public high schools could be partly explained by standard measures of school quality as well as proxies for 'community characteristics' associated with the high school. Generally, these school-level characteristics were found to be 'insignificant', suggesting that most of the variation across high schools is attributable to unobserved variation not captured through our measures of school quality or proxies for community characteristics. However, in both the HSB and NLSY data we found evidence that teacher education has an effect on student outcomes and has the largest effect

[28] For simplicity we take 2000 randomly selected post-convergence draws from the school-specific intercept and return to schooling posteriors, and obtain the boxplots using these randomly selected draws.

[29] Recall, however, that the HSB data suggested evidence of heterogeneity in both the intercepts and returns to education.

on those students completing the most schooling. We also found in the HSB that the average family income of students within each school was strongly correlated with estimated school-level parameters, and that the community-level average family income effect was strongest for those with the least amount of education. This result suggests that some of the parameter variation across schools can be attributed to 'neighbourhood effects', although most of the variation across high schools remains unobserved.

APPENDIX

Let $\Gamma = [\pi\ \beta\ \{\theta_s\}\ \theta\ \sigma_\varepsilon^2\ \sigma_\zeta^2\ \Sigma\ \{\lambda_{is}\}\ \{\zeta_{is}\}\ \nu]$ denote all the parameters in the model, and Γ_{-x} denote all parameters other than x. Further, let t_i denote the number of observations obtained for individual i, n denote the total number of individuals in the sample, $\overline{S}$ the total number of high schools in the sample, r_s denote the number of students in school s, and $\overline{N} \equiv \sum_{i=1}^{n} t_i$ denote the total number of person-year observations. Let y_{is} denote the $t_i \times 1$ vector of log wages for individual i and X_{is} the $t_i \times k_\pi$ matrix of $x's$ for individual i, where k_r denotes the length of the vector r. Similarly, let Z_s, W_s and ζ_s denote the $r_s \times 2$, $r_s \times k_\beta$ and $r_s \times 1$ (respectively) set of explanatory variables and fixed effects for individuals observed in school s. Let $\Lambda \equiv \mathrm{diag}\{\lambda_{is}\}$, and Λ_s be the $r_s \times r_s$ diagonal matrix with the mixing variables for school s placed on the diagonal. Let X and y be the $\overline{N} \times k_\pi$ and $\overline{N} \times 1$ vector of time-varying explanatory variables and log wages (respectively) for the full sample, $\overline{\zeta}$ denote the associated $\overline{N} \times 1$ vector of fixed effects, arranged according to y, and W and ζ be the stacked $n \times k_\beta$ and $n \times 1$ set of time-invariant explanatory variables and individual fixed effects, respectively. Finally, let $\overline{Z\theta}$ be the $n \times 1$ vector which multiplies each z_{is} by the corresponding θ_s, $\overline{\theta} \equiv [\theta_1'\ \theta_2' \cdots \theta_{\overline{S}}']'$ and $Q = [Q_1' Q_2' \cdots Q_{\overline{S}}']'$.

The Joint Posterior

Let $\phi(x; \mu, \Sigma)$ denote the multivariate normal density for x with mean μ and covariance matrix Σ. Further, let $p_{IG}(w|a, b)$ denote that w has an inverted gamma density with parameters a and b, and $p_W(\Omega^{-1}|\rho, \rho R)$ denote the Wishart density for Ω^{-1} with parameters ρ and ρR. For simplicity, let $p_{IG}(\cdot)$ and $p_W(\cdot)$ denote the employed inverted gamma and Wishart prior densities, respectively, and $p_N(\cdot)$ the normal density without explicitly denoting the prior hyperparameters. Finally, let Γ denote all the parameters in the model. Given this notation and the model in (1)–(11), the joint posterior is obtained as follows:

$$p(\Gamma|D) \propto \left[\prod_{i=1}^{n} \phi(y_{is}; \zeta_{is} + X_{is}\pi, \sigma_\varepsilon^2 I_{t_i}) \phi(\zeta_{is}; z_{is}\theta_s + w_{is}\beta, \sigma_\zeta^2 \lambda_{is}) p_{IG}(\lambda_{is}|\nu/2, 2/\nu)\right]$$
$$\times \left[\prod_{s=1}^{\overline{S}} \phi(\theta_s; Q_s\theta, \Sigma)\right] p_N(\pi) p_N(\beta) p_N(\theta) p_{IG}(\sigma_\varepsilon^2) p_{IG}(\sigma_\zeta^2) p_W(\Sigma^{-1}) p(\nu)$$

Standard Gibbs Algorithm

The following complete conditional posterior distributions are obtained:

$$\zeta_{is}|\Gamma_{-\zeta_{is}}, D \overset{ind}{\sim} N(D_\zeta d_\zeta, D_\zeta), \quad i = 1, 2, \ldots, n \tag{A1}$$

where

$$D_\zeta = \left(t_i/\sigma_\varepsilon^2 + 1/[\sigma_\zeta^2\lambda_{is}]\right)^{-1}, \quad d_\zeta = \left(\sum_{t=1}^{t_i}(y_{ist} - x_{ist}\pi)/\sigma_\varepsilon^2 + (z_{is}\theta_s + w_{is}\beta)/[\sigma_\zeta^2\lambda_{is}]\right)$$

$$\pi|\Gamma_{-\pi}, D \sim N(D_\pi d_\pi, D_\pi) \tag{A2}$$

where

$$D_\pi = \left(X'X/\sigma_\varepsilon^2 + V_\pi^{-1}\right)^{-1}, \quad d_\pi = \left(X'(y - \overline{\zeta})/\sigma_\varepsilon^2 + V_\pi^{-1}\pi_0\right)$$

$$\theta_s|\Gamma_{-\theta_s}, D \overset{ind}{\sim} N(D_{\theta_s} d_{\theta_s}, D_{\theta_s}), \quad s = 1, 2, \ldots \overline{S} \tag{A3}$$

where

$$D_{\theta_s} = \left(Z_s'\Lambda_s^{-1}Z_s/\sigma_\zeta^2 + \Sigma^{-1}\right)^{-1}, \quad d_{\theta_s} = \left(Z_s'\Lambda_s^{-1}(\zeta_s - W_s\beta)/\sigma_\zeta^2 + \Sigma^{-1}Q_s\theta\right)$$

$$\beta|\Gamma_{-\beta}, D \sim N(D_\beta d_\beta, D_\beta) \tag{A4}$$

where

$$D_\beta = \left(W'\Lambda^{-1}W/\sigma_\zeta^2 + V_\beta^{-1}\right)^{-1}, d_\beta = \left(W'\Lambda^{-1}(\zeta - \overline{Z\theta})/\sigma_\zeta^2 + V_\beta^{-1}\beta_0\right)$$

$$\theta|\Gamma_{-\theta}, D \sim N(D_\theta d_\theta, D_\theta) \tag{A5}$$

where

$$D_\theta = (Q'(I_{\overline{S}} \otimes \Sigma^{-1})Q + V_\theta^{-1})^{-1}, d_\theta = (Q'(I_{\overline{S}} \otimes \Sigma^{-1})\overline{\theta} + V_\theta^{-1}\theta_0)$$

$$\Sigma^{-1}|\Gamma_{-\Sigma^{-1}}, D \sim W\left(\overline{S} + \rho, \left[\sum_{s=1}^{\overline{S}}(\theta_s - Q_s\theta)(\theta_s - Q_s\theta)' + \rho R\right]\right) \tag{A6}$$

$$\sigma_\varepsilon^2|\Gamma_{-\sigma_\varepsilon^2}, D \sim IG[\overline{N}/2 + \kappa_1, (1/2(y - \overline{\zeta} - X\pi)'(y - \overline{\zeta} - X\pi) + \kappa_2^{-1})^{-1}] \tag{A7}$$

$$\sigma_\zeta^2|\Gamma_{-\sigma_\zeta^2}, D \sim IG[n/2 + \tau_1, (1/2(\zeta - \overline{Z\theta} - W\beta)'\Lambda^{-1}(\zeta - \overline{Z\theta} - W\beta) + \tau_2^{-1})^{-1}] \tag{A8}$$

$$\lambda_{is}|\Gamma_{-\lambda_{is}}, D \overset{ind}{\sim} IG\left(\frac{\nu+1}{2}, \left(\frac{\nu}{2} + \frac{1}{2\sigma_\zeta^2}(\zeta_{is} - z_{is}\theta_s - w_{is}\beta)^2\right)^{-1}\right), \quad i = 1, 2, \ldots n \tag{A9}$$

$$p(\nu|\Gamma_{-\nu}, D) \propto p(\nu)\prod_{i=1}^{N}[\Gamma(\nu/2)(2/\nu)^{(\nu/2)}]^{-1}\lambda_{is}^{-[(\nu/2)+1]}\exp[-\nu/(2\lambda_{is})] \tag{A10}$$

All these distributions are easily sampled from, except for the conditional distribution of ν. We discretize the set of possible values for the degrees of freedom parameter ν and draw ν values from the set {2, 3, 4, 5, 6, 7, 8, 10, 12, 16, 24, 32}. Draws from this discrete distribution can be easily obtained.

ACKNOWLEDGEMENTS

We would like to thank Dale Poirier, Jeremy Verlinda, three anonymous referees and the co-editor for their helpful comments and suggestions. All errors are, of course, our own.

REFERENCES

Andrews DF, Mallows CL. 1974. Scale mixtures of normality. *Journal of the Royal Statistical Society Ser. B* **36**: 99–102.

Baltagi BH, Griffin JM, Xiong W. 2000. To pool or not to pool: homogeneous versus heterogeneous estimators applied to cigarette demand. *Review of Economics and Statistics* **82**: 117–126.

Betts JR. 1995. Does school quality matter? Evidence From the National Survey of Youth. *Review of Economics and Statistics* **77**(2): 231–250.

Betts JR. 1996. Do school resources matter only for older workers? *Review of Economics and Statistics* **78**(4): 638–652.

Card D, Krueger AB. 1992. Does school quality matter? Returns to education and the characteristics of public schools in the United States. *Journal of Political Economy* **100**(1): 1–40.

Carlin BP, Polson NG. 1991. Inference for nonconjugate Bayesian models using the Gibbs sampler. *The Canadian Journal of Statistics* **19**(4): 399–405.

Cawley J, Conneely K, Heckman J, Vytlacil E. 1997. Cognitive ability, wages, and meritocracy. In *Intelligence, Genes and Success: Scientists Respond to the Bell Curve*, Devlin B, Feinberg SE, Resnick DP, Roeder K (eds). Springer: New York; 178–192.

Chib S, Carlin BP. 1999. On MCMC sampling in hierarchical longitudinal models. *Statistics and Computing* **9**: 17–26.

DiNardo J, Tobias J. 2001. Nonparametric density and regression estimation. *Journal of Economic Perspectives* **15**: 11–28.

Eide E, Showalter MH. 1998. The effect of school quality on student performance: a quantile regression approach. *Economics Letters* **58**: 345–350.

Figlio D. 1999. Functional form and the estimated effects of school resources. *Economics of Education Review* **18**: 241–252.

Geweke J. 1993. Bayesian treatment of the independent Student *t* linear model. *Journal of Applied Econometrics* **8**: 19–40.

Grogger J. 1996a. Does school quality explain the recent black/white wage trend? *Journal of Labor Economics* **14**(2): 231–253.

Grogger J. 1996b. School expenditures and post-schooling earnings: evidence from High School and Beyond. *Review of Economics and Statistics* **78**(4): 628–637.

Hanushek EA. 1986. The economics of schooling: production and efficiency in public schools. *Journal of Economic Literature* **24**: 1141–1177.

Hanushek EA, Kain J, Rivkin S. 1998. Teachers, schools and academic achievement. NBER working paper #6691.

Hanushek EA, Rivkin SG, Taylor LL. 1996. Aggregation and the estimated effects of school resources. *Review of Economics and Statistics* **78**(4): 611–626.

Heckman JJ, Layne-Ferrar A, 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 JJ, Sedlacek GL. 1990. Self-selection and the distribution of hourly wages. *Journal of Labor Economics* **8**(1): S329–S363.

Heckman JJ, Tobias JL, Vytlacil E. 2001. Four parameters of interest in the evaluation of social programs. *Southern Economic Journal* **68**(2): 210–223.

Hoxby C. 1988. The effects of class size and composition on student achievement: new evidence from natural population variation. NBER working paper #6869.

Hoxby C. 2000. Peer effects in the classroom: learning from gender and race variation. NBER working paper #7867.

Johnson GE, Stafford FP. 1973. Social returns to quantity and quality of schooling. *Journal of Human Resources* **8**(2): 139–155.

Laird NM, Ware JH. 1982. Random-effects models for longitudinal data. *Biometrics* **38**: 963–974.

Lange N, Carlin BP, Gelfand AE. 1992. Hierarchical Bayes models for the progression of HIV infection using longitudinal CD4 T-Cell numbers (with discussion). *Journal of the American Statistical Association* **87**: 615–632.

Lydall H. 1968. *The Structure of Earnings*. Clarendon Press: Oxford.

Nechyba TJ. 1990. The southern wage gap, human capital and the quality of education. *Southern Economic Journal* **57**(2): 308–322.

Olson C, Ackerman D. 1999. High school inputs and labor market outcomes for male workers in their mid-thirties: new data and new estimates from Wisconsin. University of Wisconsin-Madison, mimeo.
Rivkin S. 2001. Tiebout sorting, aggregation and the estimation of peer group effects. *Economics of Education Review* **20**(3): 201–209.
Rumberger RW, Thomas SL. 1993. The economic returns to college major, quality and performance: a multilevel analysis of recent graduates. *Economics of Education Review* **12**(1): 1–19.
Tobias JL. 2001. Forecasting output growth rates and median output growth rates: a hierarchical Bayesian approach. *Journal of Forecasting* **20**: 297–314.
Wakefield JC, Smith AFM, Racine-Poon A, Gelfand AE. 1994. Bayesian analysis of linear and nonlinear population models using the Gibbs sampler. *Journal of the Royal Statistical Society, Ser. C* **43**: 201–221.
Zellner A, Hong C. 1989. Forecasting international growth rates using Bayesian shrinkage and other procedures. *Journal of Econometrics* **40**: 183–202.

