

## HIGH SCHOOL COMPLETION AND FUTURE YOUTH UNEMPLOYMENT: NEW EVIDENCE FROM HIGH SCHOOL AND BEYOND

MINGLIANG LI\*

*Department of Economics, State University of New York at Buffalo, USA*

### SUMMARY

In this paper, I provide new evidence from High School and Beyond (HSB) on the effects of compulsory attendance on high school completion and future youth unemployment. I develop Bayesian estimation approaches to the simultaneous equation model with ordered probit and two-limit censored regression and the bivariate duration model, accounting for the heterogeneity in returns to education and the nonlinearity in the effects of compulsory attendance. I find substantial variability in returns to education across schools and evidence of diminishing marginal effects of compulsory attendance on high school completion. The simulation results suggest that increasing the compulsory attendance age raises the probability of completing high school and reduces the proportion of time the individuals are unemployed. These effects are much more pronounced for disadvantaged students but less pronounced for advantaged students, suggesting the potential effects of compulsory attendance on reducing the inequality in education and employment. Copyright © 2005 John Wiley & Sons, Ltd.

### 1. INTRODUCTION

Economists interested in the benefits of schooling have traditionally focused on the wages (or earnings) return to education. Yet, much less research has been undertaken that evaluates the importance of other types of benefits of education, such as its potential effects on unemployment. In this paper, I estimate the effects of high school completion on unemployment, accounting for the endogeneity of education and the heterogeneity in schooling returns. In the empirical literature, the exploitation of natural experiments in estimating schooling returns has become a popular approach. For example, Angrist and Krueger (1991) establish that the season of birth is related to educational attainment because of the school start age policy and compulsory attendance laws.<sup>1</sup> While the availability of exogenous change in economic environment is an advantage to the researcher, in the presence of heterogeneity, it will typically not obviate the need for modelling individual reactions to the experiment (Rosenzweig and Wolpin, 2000).

I analyse the effects of schooling on unemployment using data from High School and Beyond (HSB), employing compulsory attendance laws as an instrument for education, and accounting for the heterogeneity in schooling returns. Stern *et al.* (1989) use HSB, and associate high school completion with unemployment controlling for the endogeneity of schooling. However, their exclusion restrictions in identifying the causal effects of schooling on unemployment are arbitrary.

---

\* Correspondence to: Professor Mingliang Li, Department of Economics, 437 Fronczak Hall, State University of New York at Buffalo, Buffalo, NY 14260, USA. E-mail: mli3@buffalo.edu

<sup>1</sup> Other studies using state compulsory attendance laws as an instrument for schooling include: Lang and Kropp (1986), Harmon and Walker (1995), Acemoglu and Angrist (2000), Lleras-Muney (2002), Goldin and Katz (2003), Lochner and Moretti (2003) and Oreopoulos (2003).

I utilize the often-neglected information contained in HSB on compulsory attendance to estimate the employment returns to education. Because of the increasing popularity of the use of compulsory attendance as an instrument, the census data has been widely employed in the recent literature on schooling returns. HSB offers an abundance of individual-level variables with potential influence on schooling and unemployment, which are not available in the census data. Therefore, HSB has the advantage of checking the robustness of earlier results to the inclusion of additional control variables likely to be related to unemployment or schooling.

A recent literature on the structural estimation of dynamic discrete choice models grew substantially over the last two decades (Rust, 1987; Keane and Wolpin, 1994). Several important papers from this literature find smaller returns to education than the reduced-form models do. For example, Eckstein and Wolpin (1999) develop and structurally estimate a sequential model of high school attendance and work decisions. They find that if dropouts were forced to remain in school for 5 years after entry without working, then their graduation rate would increase only to 13%. Keane and Wolpin (1997) predict that a college tuition subsidy of \$2000 would increase high school graduation rates by 3.5 percentage points and increase college graduation rates by 8.4 percentage points, but have a negligible impact on the expected present value of lifetime utility. Belzil and Hansen (2002) find that the local wage returns to schooling are very low until grade 11 (1% per year or less), increase to 3.7% in grade 12, and exceed 10% between grades 14 and 16.<sup>2</sup>

This paper follows the traditional literature and estimates a reduced-form model of employment returns to education, allowing for nonlinearity in the effects of compulsory attendance on schooling attainment and heterogeneous returns to education across high schools. Specifically, I employ a fourth-order polynomial in the time eligible to drop out to capture the nonlinear effect of compulsory attendance, and find some evidence of diminishing marginal effects of compulsory attendance on high school completion.

To model the school-level heterogeneity in employment returns to education, I exploit within-school variation in the high school grades completed by the individuals. The model has a hierarchy of two levels, the individual level and the school level. At the individual level, I specify a vector of dummy variables indicating the high school grade completed by the individual and relate it to the outcome of unemployment, along with a linear specification in the other individual-level variables. The school-level heterogeneity is incorporated by permitting the intercept and slopes of the dummy variable specification to vary across schools. At the school level, a multivariate normal distribution is used to model this variation. Although this model captures the school-level heterogeneity in returns to education, the identification strategy relies on within-school variation in the high school grades completed by the individuals for some schools in the sample (Tobias and Li, 2003; Koop and Tobias, 2004).

I develop a Bayesian estimation approach to the simultaneous equation model with ordered probit and two-limit censored regression. I find that for average students, increasing the compulsory attendance age from 16 to 18 increases the probability of completing high school by 5.1 percentage points (from 89 to 94.1 percentage points, or 5.7%), and decreases the proportion of time the individuals are unemployed by 0.22 percentage points (or 2.5%). These effects are much more pronounced for disadvantaged students and are less pronounced for advantaged students,

---

<sup>2</sup> I thank an anonymous referee and the editor for bringing my attention to this literature. This paper does not account for forward-looking behaviour in the schooling decision. In the empirical sections, I compare my policy forecasts with the results from the literature of structural dynamic discrete choice models.

suggesting the potential effects of compulsory attendance on reducing the inequality in education and employment.

It is possible that unemployment behaviour and schooling behaviour are characterized by strong state dependence, so that the probabilities of leaving unemployment and leaving high school today depend on the amount of time an individual has been unemployed and has been in school in the past. Compulsory school attendance may have long-lasting effects on unemployment and schooling. To this end, I utilize the information contained in HSB on unemployment duration and high school duration and develop a Bayesian estimation approach to the bivariate duration model. I find that for average students, increasing the compulsory attendance age from 16 to 18 increases the high school duration by 0.9 months (from 24.5 to 25.4 months, or 3.7%), and increases the hazard of leaving unemployment in the first month of unemployment by 0.5 percentage points (or 1.3%).

The remainder of the paper is organized as follows. Section 2 examines the effects of compulsory attendance on high school completion and future unemployment by developing a Bayesian estimation approach to the simultaneous equation model with ordered probit and two-limit censored regression. Section 3 studies the impact of dropout eligibility on high school duration and future unemployment duration by employing a Bayesian approach to the bivariate duration model. Section 4 concludes.

## 2. HIGH SCHOOL COMPLETION AND FUTURE UNEMPLOYMENT

### 2.1. Data

High School and Beyond is a national survey of US high school students, and was conducted by the National Center for Education Statistics (NCES). The base year survey was in the spring of 1980, and up to 36 sophomores from each of about 1000 high schools in the United States participated in the base year survey. In the springs of 1982, 1984 and 1986, first, second and third follow-up surveys collected information on high school completion and post-schooling unemployment. The survey investigator used school administrative records to classify the dropout status of the 1980 sophomores. Students who dropped out of high school reported the high school grades they completed before dropping out. In Table I, I present the descriptive statistics for the data used in this section.

In the sample, 1.5% of the 1980 sophomores dropped out of high school after completing the ninth grade, 4.2% after completing the tenth grade, 3.8% after completing the eleventh grade, and 90.5% completed high school. The survey collected information on the post-schooling unemployment of the 1980 sophomores. The unemployment outcome is defined to be the proportion of time in the labour market for which an individual is unemployed. The average unemployment rate is 11% for the entire sample, 9.8% for high school graduates, 22.1% for dropouts, 28.2% for dropouts completing the ninth grade, 21.6% for dropouts completing the tenth grade, and 20.4% for dropouts completing the eleventh grade.<sup>3</sup>

These descriptive statistics suggest the positive effect of high school completion on future youth employment, although I have not yet controlled for the endogeneity of schooling and observable

---

<sup>3</sup> These descriptive statistics are close to the unemployment rates usually reported for this age group. For example, Stern *et al.* (1989) report from the National Longitudinal Survey of Youth (NLSY) unemployment rates of 22.5% for high school dropouts and 15.5% for graduates. They also report from HSB unemployment rates of 19.7% for dropouts and 13.8% for graduates. Their numbers are close to my descriptive statistics, although in their HSB sample, they restrict their attention to unemployment information of the 1980 sophomores before the spring of 1984.

Table I. Descriptive statistics for the data: high school completion and unemployment

	Sample mean	Standard error
Parental income (\$10 000)	2.13	1.1
Base year cognitive test	0	1
Father's education	12.3	4.66
Mother's education	12.2	3.7
Number of siblings	2.63	1.83
Female	0.523	0.5
Minority	0.267	0.443
County level employment growth rate 80–82 (%)	−0.311	5.46
Age <sup>a</sup>	−0.342	0.496
Time eligible to drop out <sup>b</sup>	0.654	0.839
Post-secondary education	0.71	0.454
High school grade completion <sup>c</sup>		
Completing the ninth grade	0.0145	0.12
Completing the tenth grade	0.0418	0.2
Completing the eleventh grade	0.0384	0.192
Proportion of time unemployed <sup>d</sup>	0.11	0.204

<sup>a</sup> Age is the age of the individual on January 1, 1980, minus 16.

<sup>b</sup> Time eligible to drop out is the time from which the individual is eligible to drop out: time eligible to drop out = (date eligible to drop out – January 1, 1980) ÷ 365 days.

<sup>c</sup> For high school grade completion, the excluded group are individuals who have completed high school.

<sup>d</sup> Proportion of time unemployed is the proportion of time in the labour market for which an individual is unemployed.

individual characteristics. HSB offers an abundance of individual-level variables with potential influence on schooling and unemployment, which are not available in the census data. In particular, HSB records the score of a standard cognitive test taken in the base year of the survey (1980) that can be employed as a measure of cognitive ability. The following variables are also available: parental income and education, number of siblings in the household, race, gender, age and county level employment growth rate between 1980 and 1982. The individuals reported whether they had any post-secondary education.

To control for the endogeneity of schooling, I employ compulsory attendance laws as an instrument. Most of the instrumental variable approaches use census data in which the states of residence of the individuals can be identified. One reason that few researchers have used HSB to study the effects of compulsory attendance on high school completion and post-schooling unemployment is that the geographical locations of the 1980 sophomores are not revealed to the public.

Following the attempts of Hanushek and Taylor (1990), Rivkin (1991), Ganderton (1992), Grogger (1996a,b), I identify the states in which the schools of the 1980 sophomores were located. Specifically, to identify the locations of the schools and their students, I utilize the local labour market conditions for HSB schools between 1980 and 1982 that were previously released by the NCES. This data set does not identify the actual locations of the schools and their students, but it contains demographic information at the state level for each school and its students in HSB. I use publicly available demographic data to match this data set, and thus identify the location of each school and its students at the state level.

After I identify the state of residence of each 1980 sophomore, I construct the state compulsory attendance age of each individual following Angrist and Krueger (1991), whose source of

information is the Digest of Education Statistics published by the NCES.<sup>4</sup> In the sample, 74.7% of the individuals have a compulsory attendance age of 16 or less, 16.5% have a compulsory attendance age of 17, and 8.8% have a compulsory attendance age of 18. Because the date of birth of each 1980 sophomore is available, I can construct a variable indicating the time an individual is eligible to drop out.<sup>5</sup> After excluding observations with missing variables, the final sample contains 5238 students from 871 schools.

## 2.2. Model

High school grade completion is by definition an ordered integer, and the discrete nature of the choice set implies that standard estimators of the simultaneous equation model are not consistent due to the nature of the disturbance terms.<sup>6</sup> Let  $y_{hi}$  denote the high school grade completed by individual  $i$ , and  $z_{hi}$  denote the latent outcome corresponding to  $y_{hi}$ , where  $h$  labels the schooling outcome,  $y_{hi} = 1$  if individual  $i$  dropped out of high school after completing the ninth grade,  $y_{hi} = 2$  if he dropped out after completing the tenth grade,  $y_{hi} = 3$  if he dropped out after completing the eleventh grade, and  $y_{hi} = 4$  if he completed high school. Consider an ordered probit:

$$\begin{aligned} z_{hi} &= x_{hi}\beta_h + \varepsilon_{hi}, \quad \varepsilon_{hi} \sim N(0, \Sigma_{hh}), \quad \gamma_{y_{hi}} < z_{hi} \leq \gamma_{y_{hi}+1} \\ \gamma_1 &= -\infty, \quad \gamma_2 = 0, \quad \gamma_2 < \gamma_3 < \gamma_4, \quad \gamma_4 = 1, \quad \gamma_5 = \infty \end{aligned} \quad (1)$$

for  $i = 1, 2, 3, \dots, n$ , where  $x_{hi}$  denotes a  $1 \times k_h$  vector of individual-level variables (base year cognitive test score, parental income, parental education, number of siblings, gender, race, county level employment growth rate between 1980 and 1982, a fourth-order polynomial in age and a fourth-order polynomial in the time eligible to drop out),  $\varepsilon_{hi}$  is the individual-level unobservable,  $N(\mu, \sigma^2)$  indicates the normal distribution with mean  $\mu$  and variance  $\sigma^2$ ,  $\Sigma_{hh}$  is the variance of the unobservables,  $\{\gamma_j\}_{j=1}^5$  are the cutoff points, and  $n$  is the total number of individuals.

Let  $w_{ui}$  denote the proportion of time individual  $i$  is unemployed, and  $z_{ui}$  denote the latent outcome corresponding to  $w_{ui}$ , where  $u$  labels the unemployment outcome,  $z_{ui} \leq 0$  if individual  $i$  was never unemployed ( $w_{ui} = 0$ ),  $z_{ui} = w_{ui}$  if he was unemployed for part of the time period ( $0 < w_{ui} < 1$ ), and  $z_{ui} \geq 1$  if he was always unemployed ( $w_{ui} = 1$ ). Consider a two-limit censored regression:

$$\begin{aligned} z_{ui} &= x_{ui}\beta_u + q_i\theta_s + \varepsilon_{ui}, \quad \theta_s \sim N(\theta, \Sigma_\theta), \quad \varepsilon_{ui} \sim N(0, \Sigma_{uu}) \\ z_{ui} &\begin{cases} \leq 0 & \text{if } w_{ui} = 0 \\ = w_{ui} & \text{if } 0 < w_{ui} < 1 \\ \geq 1 & \text{if } w_{ui} = 1 \end{cases} \end{aligned} \quad (2)$$

for  $i = 1, 2, 3, \dots, n$  and  $s = 1, 2, 3, \dots, S$ , where  $x_{ui}$  denotes a  $1 \times k_u$  vector of individual-level variables (base year cognitive test score, parental income, parental education, number of siblings, gender, race, age and a dummy variable indicating any post-secondary education),  $\varepsilon_{ui}$  is the

<sup>4</sup> A summary of the compulsory schooling requirement in effect in each state in 1980 is provided in Appendix 2 of Angrist and Krueger (1991). Some states allow students to drop out of school upon attaining their sixteenth birthday, while others compel students to attend school until their seventeenth or eighteenth birthday. Angrist and Krueger (1991) also discuss three exceptions in footnote 9: Mississippi, South Carolina and Maine.

<sup>5</sup> Time eligible to drop out is the time from which the individual is eligible to drop out. Specifically, time eligible to drop out = (date eligible to drop out – January 1, 1980) ÷ 365 days. On average, the 1980 sophomores were 15.7 years old on January 1, 1980, and eligible to drop out from the fourth week of August 1980.

<sup>6</sup> I thank an anonymous referee and the editor for their suggestions of modelling the effects of high school grade completion on the proportion of time an individual is unemployed.

individual-level unobservable,  $\Sigma_{uu}$  is the variance of the unobservables, and  $S$  indicates the total number of schools.

To account for the school-level heterogeneity in returns to education, I exploit within-school variation in the high school grades completed by the individuals. In equation (2),  $q_i$  denotes a  $1 \times 4$  vector of dummy variables indicating the high school grade completed by individual  $i$ , where  $q_{i1} = 1$  if individual  $i$  dropped out of high school after completing the ninth grade ( $y_{hi} = 1$ ),  $q_{i2} = 1$  if he dropped out after completing the tenth grade ( $y_{hi} = 2$ ),  $q_{i3} = 1$  if he dropped out after completing the eleventh grade ( $y_{hi} = 3$ ),  $q_{i4} = 1$  (the intercept), and the excluded group are the individuals who completed high school ( $y_{hi} = 4$ ),  $\{\theta_s\}_{s=1}^S$  indicates the  $4 \times 1$  vector of school-specific coefficients of  $q_i$ ,  $\theta$  is the  $4 \times 1$  vector of common means of  $\{\theta_s\}_{s=1}^S$ , and  $\Sigma_\theta$  is the  $4 \times 4$  common covariance matrix of  $\{\theta_s\}_{s=1}^S$ .

Because of the endogeneity of high school completion in the unemployment equation, the model is a simultaneous equation model with ordered probit and two-limit censored regression.<sup>7</sup> Failure to recognize the endogeneity when it exists produces inconsistent estimation results. To control for the endogeneity of schooling, I allow for the correlation between  $\varepsilon_{hi}$  and  $\varepsilon_{ui}$ :  $\begin{pmatrix} \varepsilon_{hi} \\ \varepsilon_{ui} \end{pmatrix} \sim N(0_{2 \times 1}, \Sigma)$ , where  $\Sigma = \begin{pmatrix} \Sigma_{hh} & \Sigma_{hu} \\ \Sigma_{uh} & \Sigma_{uu} \end{pmatrix}$  is the covariance matrix. Given the latent outcomes  $z_{hi}$  and  $z_{ui}$ , this model is a standard simultaneous equation model. I stack  $\{z_{hi}\}_{i=1}^n$ ,  $\{x_{hi}\}_{i=1}^n$ ,  $\{z_{ui}\}_{i=1}^n$ ,  $\{x_{ui}\}_{i=1}^n$  and  $\{q_i \theta_s\}_{i=1}^n$  into  $z_h$ ,  $X_h$ ,  $z_u$ ,  $X_u$  and  $\overline{Q\theta}$  of sizes  $n \times 1$ ,  $n \times k_h$ ,  $n \times 1$ ,  $n \times k_u$  and  $n \times 1$ , respectively, and the data augmented likelihood function of this simultaneous equation model is proportional to:

$$\begin{aligned}
 p(Data|\Xi) &\propto |2\pi\Sigma \otimes I_n|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \begin{pmatrix} z_h - X_h\beta_h \\ z_u - X_u\beta_u - \overline{Q\theta} \end{pmatrix}' (\Sigma \otimes I_n)^{-1} \right. \\
 &\quad \times \left. \begin{pmatrix} z_h - X_h\beta_h \\ z_u - X_u\beta_u - \overline{Q\theta} \end{pmatrix} \right] \left\{ \prod_{s=1}^S |2\pi\Sigma_\theta|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (\theta_s - \theta)' \Sigma_\theta^{-1} (\theta_s - \theta) \right] \right\} \\
 &\quad \times \left[ \prod_{i=1}^n 1(\gamma_{y_{hi}} \leq z_{hi} < \gamma_{y_{hi}+1}) \right] \left\{ \prod_{i=1}^n [1(z_{ui} \leq 0)1(w_{ui} = 0) \right. \\
 &\quad \left. + 1(z_{ui} = w_{ui})1(0 < w_{ui} < 1) + 1(z_{ui} \geq 1)1(w_{ui} = 1)] \right\} \quad (3)
 \end{aligned}$$

where  $p$  denotes the probability,  $\Xi$  the parameters,  $I$  the identity matrix and  $1(\cdot)$  the indicator function.

I use Bayesian estimation methods, which have several advantages. First of all, without data augmentation, the likelihood function associated with a bivariate limited dependent variables model is nontrivial. Due to the occurrence of multiple integrals in the likelihood function and the discrete and censored nature of the outcome variables, the full information maximum likelihood estimator is generally avoided in favour of less efficient, but computationally simpler, estimation

<sup>7</sup> The standard instrumental variable approaches assume the existence of an instrument for schooling. The exclusion restriction requires an element of the control variables in the schooling equation not to be contained in the control variables in the unemployment equation, such as a fourth-order polynomial in the time eligible to drop out. The practical importance of such an instrument has been widely documented. Further, the instrument itself will serve to define the treatment effects.

procedures, such as various two-step algorithms. Because I sample the posterior distribution using a Metropolis–Hastings within Gibbs algorithm with data augmentation, and do not rely on asymptotic theory or any approximation of the true posterior distribution, I can easily obtain exact finite sample results with cheaper computational costs.

Second, even if the number of students within each school is limited, the Bayesian approach will provide exact finite sample estimates of the school-specific returns to education. Importantly, the posterior estimates of the school-level heterogeneity incorporate not only information from the given school, but also information obtained from the returns to education estimates of all schools (Tobias and Li, 2003; Koop and Tobias, 2004).

Furthermore, with the Markov chain Monte Carlo estimation method, I can easily integrate out parameters because each parameter is assigned a probability distribution. Thus, I can formally take parameter uncertainty into account in computing predictive moments and quantiles of any function of interest. The usual list of predictive functions of interest includes posterior means  $E(\beta|Data)$ , standard deviations  $Std(\beta|Data)$  and the probabilities of being positive  $P(\beta > 0|Data)$  of the parameters. For a bivariate limited dependent variables model, the list of predictive functions of interest could also include the marginal effects of the control variables on high school grade completion  $\{\Delta P(y_h = g|\Delta x_h, Data)\}_{g=1}^4$  and their effects on unemployment  $\Delta E(w_u|\Delta x_u, Data)$ .

Although the sign of the posterior mean of a coefficient in an ordered probit (or a two-limit censored regression) does provide some information on whether a covariate is associated with an increase or decrease in the outcome variables, the magnitude of the posterior mean is not easy to interpret. The availability of Monte Carlo estimates of the above-mentioned posterior functions of interest makes the interpretation and analysis of the results much more transparent (Campolieti, 2000).

Using the Bayes' rule, the joint posterior distribution of the parameters  $p(\Xi|Data)$  is proportional to the product of the data augmented likelihood function  $p(Data|\Xi)$  in equation (3) and the joint prior distribution of the parameters  $p(\Xi)$ :  $p(\Xi|Data) \propto p(Data|\Xi)p(\Xi)$ . To complete the Bayesian analysis and obtain the posterior, it is necessary to discuss the priors  $p(\Xi)$ . I assume prior independence across parameters:  $p(\Xi) = p(\beta)p(\Sigma)p(\theta)p(\Sigma_\theta)p(\gamma_3)$ . Prior distributions of all the parameters  $p(\beta)$ ,  $p(\Sigma)$ ,  $p(\theta)$ ,  $p(\Sigma_\theta)$  and  $p(\gamma_3)$  are specified as follows:  $\beta = (\beta_h \ \beta_u)' \sim N(\beta_0, V_\beta)$ ,  $\Sigma \sim IW(\rho, \rho R)$ ,  $\theta \sim N(\theta_0, V_\theta)$ ,  $\Sigma_\theta \sim IW(\rho_\theta, \rho_\theta R_\theta)$  and  $\gamma_3 \sim Beta(u_1, u_2)$ , where  $\beta_0 = 0_{k \times 1}$ ,  $V_\beta = 1000I_k$ ,  $k$  denotes the length of vector  $\beta$ ,  $IW(\rho, \rho R)$  denotes the inverted Wishart distribution with degrees of freedom parameter  $\rho$  and scale parameter  $R$  such that  $E(\Sigma^{-1}) = R^{-1}$  and  $E(\Sigma) = \frac{\rho}{\rho-3}R$ ,<sup>8</sup>  $\rho = 6$ ,  $R = I_2$ ,  $\theta_0 = 0_{4 \times 1}$ ,  $V_\theta = 1000I_4$ ,  $\rho_\theta = 9$ ,  $R_\theta = 0.001I_4$ ,  $Beta(\alpha, \delta)$  denotes the Beta distribution with mean  $\frac{\alpha}{\alpha+\delta}$  and variance  $\frac{\alpha\delta}{(\alpha+\delta+1)(\alpha+\delta)^2}$ ,  $u_1 = 1$  and  $u_2 = 1$ .<sup>9</sup>

### 2.3. Gibbs Sampler

I use a Metropolis–Hastings within Gibbs algorithm, a simulation-based algorithm, to implement the Bayesian analysis. This algorithm involves iteratively sampling from the complete conditional posterior distributions of the parameters. Under certain regularity conditions, the posterior draws

<sup>8</sup> For example, Poirier (1995), p. 136.

<sup>9</sup> These priors are very diffuse and essentially noninformative. With alternative assignments of hyperparameters such as  $V_\beta = 100I_k$  or  $V_\beta = 10000I_k$ ,  $\rho = 4$  or  $\rho = 12$ ,  $R = 0.5I_2$  or  $R = 2I_2$ ,  $V_\theta = 100I_4$  or  $V_\theta = 10000I_4$ ,  $\rho_\theta = 6$  or  $\rho_\theta = 12$ ,  $R_\theta = 0.0005I_4$  or  $R_\theta = 0.002I_4$ ,  $u_1 = 0.5$  or  $u_1 = 2$ , and  $u_2 = 0.5$  or  $u_2 = 2$ , estimation results virtually do not change.

from the successive sampling of the complete conditional posterior distributions converge to the drawing from the joint posterior distribution (Chib and Greenberg, 1995; Gelman *et al.*, 1995; Gilks *et al.*, 1996; Gamerman, 1997).

Li (1998) proposed a Gibbs sampler with data augmentation for a simultaneous equation model with probit in the first equation and tobit in the second equation. I extend Li's Gibbs method to a simultaneous equation model with ordered probit in the first equation and a two-limit censored regression in the second equation. I combine Li's method with Nandram and Chen's (1996) method for ordered probit models. Following Nandram and Chen, I fix  $\gamma_4 = 1$  and estimate  $\gamma_3$  and  $\Sigma_{hh}$ , instead of the more traditional way which fixes variance  $\Sigma_{hh} = 1$  and estimates two cutoff points  $\gamma_3$  and  $\gamma_4$  (Albert and Chib, 1993; Cowles, 1996).

There are several advantages of using this alternative method. First, Nandram and Chen's empirical results show that their algorithm substantially improves the convergence of the Gibbs sampler compared with the Albert–Chib and Cowles algorithms. Second, as I apply Nandram and Chen's method to Li's simultaneous equation model with limited dependent variables, I avoid the normalization of  $\Sigma_{hh} = 1$ . Otherwise, the simulation from the inverted Wishart distributions conditional on one of the elements on the diagonal is more complicated (Li, 1998; Nobile, 2000). Formally, I develop a Gibbs sampler of the simultaneous equation model with ordered probit and two-limit censored regression and iterate through sampling from the following complete conditional posterior distributions derived from the joint posterior distribution  $p(\Xi|Data)$ .

1. Sample the coefficients  $\beta$ :

$$\beta|\Xi_{-\beta}, Data \sim N(D_{\beta}d_{\beta}, D_{\beta})$$

where  $\Xi_{-\theta}$  denotes all the parameters other than  $\theta$ ,  $D_{\beta} = [X'(\Sigma \otimes I_n)^{-1}X + V_{\beta}^{-1}]^{-1}$ ,  $d_{\beta} = X'(\Sigma \otimes I_n)^{-1}z + V_{\beta}^{-1}\beta_0$ ,  $X = \begin{pmatrix} X_h & 0_{n \times k_u} \\ 0_{n \times k_h} & X_u \end{pmatrix}$  and  $z = \begin{pmatrix} z_h \\ z_u - Q\theta \end{pmatrix}$ .

2. Sample the latent outcome in the schooling equation  $z_{hi}$ :

$$z_{hi}|\Xi_{-z_{hi}}, Data \sim TN_{[\gamma_{y_{hi}}, \gamma_{y_{hi}+1}]}(\mu_{h|u}, \Sigma_{h|u})$$

for  $i = 1, 2, 3, \dots, n$ , where  $TN_{[a,b]}$  denotes the normal distribution truncated between  $a$  and  $b$ ,  $\mu_{h|u} = x_{hi}\beta_h + \Sigma_{hu}\Sigma_{uu}^{-1}(z_{ui} - x_{ui}\beta_u - q_i\theta_s)$  and  $\Sigma_{h|u} = \Sigma_{hh} - \Sigma_{hu}\Sigma_{uu}^{-1}\Sigma_{uh}$ .

3. Sample the cutoff point  $\gamma_3$ :

$$p(\gamma_3|\Xi_{-\gamma_3}, Data) \propto \left\{ \prod_{i=1}^n [\Phi(\gamma_{y_{hi}+1}|\mu_{h|u}, \Sigma_{h|u}) - \Phi(\gamma_{y_{hi}}|\mu_{h|u}, \Sigma_{h|u})] \right\} \gamma_3^{u_1-1} (1 - \gamma_3)^{u_2-1}$$

where  $\Phi$  denotes the cumulative distribution function of a normal density. This conditional posterior distribution cannot be sampled directly, so I follow Nandram and Chen (1996) to use the proposal density:  $\gamma_3^* \sim Beta(\alpha_1 n_2 + u_1, \alpha_2 n_3 + u_2)$ , where  $*$  denotes the candidate draw,  $\alpha_1 = 0.1$  and  $\alpha_2 = 0.1$  are tuning parameters,  $n_2 = \sum_{i=1}^n 1(y_{hi} = 2)$  and  $n_3 = \sum_{i=1}^n 1(y_{hi} = 3)$ .<sup>10</sup>

<sup>10</sup> I thank Justin Tobias for a helpful discussion on the algorithm.



The probability of accepting the candidate draw is  $\min(R, 1)$ , where

$$R = \left[ \prod_{i=1}^n \frac{\Phi(\gamma_{y_{hi}+1}^* | \mu_{h|u}, \Sigma_{h|u}) - \Phi(\gamma_{y_{hi}}^* | \mu_{h|u}, \Sigma_{h|u})}{\Phi(\gamma_{y_{hi}+1, j-1} | \mu_{h|u}, \Sigma_{h|u}) - \Phi(\gamma_{y_{hi}, j-1} | \mu_{h|u}, \Sigma_{h|u})} \right] \left( \frac{\gamma_{3, j-1}}{\gamma_3^*} \right)^{\alpha_1 n_2} \left( \frac{1 - \gamma_{3, j-1}}{1 - \gamma_3^*} \right)^{\alpha_2 n_3}$$

and subscript  $j - 1$  denotes the last accepted draw.

4. Sample the latent outcome in the unemployment equation  $z_{ui}$ :

$$z_{ui} | \Xi_{-z_{ui}}, Data \begin{cases} \sim TN_{[-\infty, 0]}(\mu_{u|h}, \Sigma_{u|h}) & \text{if } w_{ui} = 0 \\ = w_{ui} & \text{if } 0 < w_{ui} < 1 \\ \sim TN_{[1, \infty]}(\mu_{u|h}, \Sigma_{u|h}) & \text{if } w_{ui} = 1 \end{cases}$$

for  $i = 1, 2, 3, \dots, n$ , where  $\mu_{u|h} = x_{ui}\beta_u + q_i\theta_s + \Sigma_{uh}\Sigma_{hh}^{-1}(z_{hi} - x_{hi}\beta_h)$  and  $\Sigma_{u|h} = \Sigma_{uu} - \Sigma_{uh}\Sigma_{hh}^{-1}\Sigma_{hu}$ .

5. Sample the covariance matrix  $\Sigma$ :

$$\Sigma | \Xi_{-\Sigma}, Data \sim IW[\rho + n, \rho R + (z_h - X_h\beta_h \quad z_u - X_u\beta_u - \overline{Q\theta})' \\ \times (z_h - X_h\beta_h \quad z_u - X_u\beta_u - \overline{Q\theta})]$$

6. Sample the school-specific returns to education  $\theta_s$ :

$$\theta_s | \Xi_{-\theta_s}, Data \sim N(D_{\theta_s} d_{\theta_s}, D_{\theta_s})$$

for  $s = 1, 2, 3, \dots, S$ , where  $D_{\theta_s} = \left( \Sigma_{\theta}^{-1} + \sum_{i=1}^I q_i' q_i \Sigma_{u|h}^{-1} \right)^{-1}$  and  $d_{\theta_s} = \Sigma_{\theta}^{-1} \theta + \sum_{i=1}^I q_i' [z_{ui} - x_{ui}\beta_u - \Sigma_{uh}\Sigma_{hh}^{-1}(z_{hi} - x_{hi}\beta_h)] \Sigma_{u|h}^{-1}$ .

7. Sample the returns to education common means  $\theta$ :

$$\theta | \Xi_{-\theta}, Data \sim N(D_{\theta} d_{\theta}, D_{\theta})$$

where  $D_{\theta} = (V_{\theta}^{-1} + \Sigma_{\theta}^{-1} S)^{-1}$  and  $d_{\theta} = V_{\theta}^{-1} \theta_0 + \Sigma_{\theta}^{-1} \sum_{s=1}^S \theta_s$ .

8. Sample the returns to education common covariance matrix  $\Sigma_{\theta}$ :

$$\Sigma_{\theta} | \Xi_{-\Sigma_{\theta}}, Data \sim IW \left[ \rho_{\theta} + S, \rho_{\theta} R_{\theta} + \sum_{s=1}^S (\theta_s - \theta)(\theta_s - \theta)' \right]$$

I run this algorithm with 20 000 iterations and discard the first 4000 iterations as the pre-convergence draws. To monitor the convergence of the algorithm, I simulate several independent sequences, with starting points sampled from an overdispersed distribution, following Gelman *et al.* (1995).

## 2.4. Results

In Table II, I present the posterior means  $E(\beta|Data)$ , standard deviations  $\text{Std}(\beta|Data)$  and probabilities of being positive  $P(\beta > 0|Data)$  of the coefficients, and the posterior marginal effects

Table II. Posterior means  $E(\beta|Data)$ , standard deviations  $Std(\beta|Data)$  and probabilities of being positive  $P(\beta > 0|Data)$  of the coefficients, and posterior marginal effects of the control variables on the probability of high school completion  $\Delta P(y_h = 4|\Delta x_h, Data)$  and unemployment  $\Delta E(w_u|\Delta x_u, Data)$

Explanatory variables <sup>a</sup>	High school completion $y_h$			
	$E(\beta D)^b$	$Std(\beta D)$	$P(\beta > 0 D)$	$\Delta P(y_h = 4 \Delta x_h, D)^c$
Parental income	0.0112	0.0254	0.674	0.00213
Base year cognitive test	0.433	0.0365	1	0.0633
Father's education	0.0445	0.0128	1	0.0085
Mother's education	0.062	0.0157	1	0.0117
Number of siblings	-0.0362	0.015	0.00775	-0.00727
Female	-0.0686	0.0523	0.0912	-0.014
Minority	0.38	0.066	1	0.0575
County employment growth	-0.0128	0.00457	0.00187	-0.00254
Age	-0.408	0.0807	0	-0.165
Age <sup>2</sup>	-0.192	0.0724	0.00475	
Age <sup>3</sup>	-0.0304	0.0412	0.245	
Age <sup>4</sup>	0.0305	0.0134	0.997	
Time eligible to drop out	0.0958	0.0672	0.925	0.0292
Time <sup>2</sup>	0.0896	0.0451	0.982	
Time <sup>3</sup>	-0.00918	0.0106	0.19	
Time <sup>4</sup>	-0.0094	0.0051	0.0299	
Constant	0.946	0.211	1	
	Proportion of time unemployed $w_u$			
	$E(\beta D)$	$Std(\beta D)$	$P(\beta > 0 D)$	$\Delta E(w_u \Delta x_u, D)^d$
Parental income	-0.0256	0.0055	0	-0.00915
Base year cognitive test	-0.0365	0.00686	0	-0.0128
Father's education	-0.00152	0.00238	0.265	-0.000565
Mother's education	-0.00344	0.00283	0.11	-0.00127
Number of siblings	0.00471	0.00326	0.925	0.00176
Post-secondary education	-0.0105	0.0136	0.222	-0.00387
Female	0.0604	0.0109	1	0.0246
Minority	0.0859	0.0131	1	0.0363
Age	-0.00567	0.0122	0.317	-0.00194
Completing ninth grade ( $\theta_1$ )	0.164	0.0677	0.991	0.0791
Completing tenth grade ( $\theta_2$ )	0.11	0.0445	0.994	0.0487
Completing eleventh grade ( $\theta_3$ )	0.0983	0.0394	0.994	0.0428
Constant ( $\theta_4$ )	-0.00643	0.0397	0.435	

<sup>a</sup> The descriptive statistics for the data are reported in Table I.

<sup>b</sup>  $D$  denotes the data.

<sup>c</sup>  $\Delta P(y_h = 4|\Delta x_h, Data)$  denotes the posterior marginal effects of the control variables in the schooling equation on the probability of completing high school.

<sup>d</sup>  $\Delta E(w_u|\Delta x_u, Data)$  denotes the posterior marginal effects of the control variables in the unemployment equation on the proportion of time the individual is unemployed.

of the control variables on the probability of high school completion  $\Delta P(y_h = 4|\Delta x_h, Data)$  and unemployment  $\Delta E(w_u|\Delta x_u, Data)$ . Formally,

1. Probability of high school completion ( $y_h = 4$ ):

$$P(y_h = 4) = \Phi(\infty|x_h\beta_h, \Sigma_{hh}) - \Phi(1|x_h\beta_h, \Sigma_{hh}) = 1 - \Phi(1|x_h\beta_h, \Sigma_{hh})$$

where  $x_h$  denotes the  $1 \times k_h$  vector of the individual-level characteristics of a representative individual.

2. Proportion of time unemployed ( $w_u$ ):

$$\begin{aligned}
 E(w_u) &= P(w_u \leq 0) \cdot 0 + P(0 < w_u < 1)E(w_u|0 < w_u < 1) + P(w_u \geq 1) \cdot 1 \\
 &= [\Phi(0|x_u\beta_u + q\theta, \Sigma_{uu}) - \Phi(-\infty|x_u\beta_u + q\theta, \Sigma_{uu})] \cdot 0 \\
 &\quad + [\Phi(1|x_u\beta_u + q\theta, \Sigma_{uu}) - \Phi(0|x_u\beta_u + q\theta, \Sigma_{uu})] \\
 &\quad \times \left[ x_u\beta_u + q\theta - \Sigma_{uu} \frac{\phi(1|x_u\beta_u + q\theta, \Sigma_{uu}) - \phi(0|x_u\beta_u + q\theta, \Sigma_{uu})}{\Phi(1|x_u\beta_u + q\theta, \Sigma_{uu}) - \Phi(0|x_u\beta_u + q\theta, \Sigma_{uu})} \right] \\
 &\quad + [\Phi(\infty|x_u\beta_u + q\theta, \Sigma_{uu}) - \Phi(1|x_u\beta_u + q\theta, \Sigma_{uu})] \cdot 1 \\
 &= [\Phi(1|x_u\beta_u + q\theta, \Sigma_{uu}) - \Phi(0|x_u\beta_u + q\theta, \Sigma_{uu})](x_u\beta_u + q\theta) \\
 &\quad - [\phi(1|x_u\beta_u + q\theta, \Sigma_{uu}) - \phi(0|x_u\beta_u + q\theta, \Sigma_{uu})]\Sigma_{uu} + 1 - \Phi(1|x_u\beta_u + q\theta, \Sigma_{uu})
 \end{aligned}$$

where  $x_u$  denotes the  $1 \times k_u$  vector of the individual-level characteristics of a representative individual, and  $q$  denotes the  $1 \times 4$  vector of dummy variables indicating the high school grade completed by the individual.

I first discuss the results for the probability of high school completion (reported in the top panel of Table II), while the marginal effects are reported in the last column of Table II.<sup>11</sup> The coefficients on the individual-level variables generally have the expected signs and are statistically significant. For example, a one standard deviation increase in the base year cognitive test score increases the probability of completing high school by 6.3 percentage points (from 89 to 95.3 percentage points, or 7.1%).<sup>12</sup> To capture the effects of family endowments, I also include the parental education, parental income and number of siblings in both the schooling and unemployment equations.<sup>13</sup> A 1 year increase in the father's education increases the probability of high school completion by 0.85 percentage points (or 0.96%), and a 1 year increase in the mother's education increases the probability by 1.2 percentage points (or 1.3%). Having one more sibling reduces the probability by

<sup>11</sup> To report the marginal effects of the control variables on the probability of high school completion, I define the representative individual to have a parental income of \$20 000, a base year cognitive test score of 0 (standardized test score with mean of zero and standard deviation of one), a father with an education of 12 years, a mother with an education of 12 years, two siblings, and to be male, white, residing in a county with an employment growth rate from 1980 to 1982 of zero percentage points, born on January 1, 1964, and eligible to drop out from January 1, 1980. The predictive probability of completing high school for this representative individual is 0.89.

<sup>12</sup> The cognitive test was conducted in the base year of the survey (1980) while the schooling and unemployment outcomes were collected in the follow-up surveys. The test serves as a proxy for the cognitive ability and has six test components: vocabulary, reading, math, science, writing and civics. Some researchers prefer to utilize the math test only, but the results of this paper are not sensitive to this alternative specification of the test components. For another robustness check, I re-estimated the model by excluding the cognitive test score from both the schooling and unemployment equations. The cognitive ability is then mainly captured by the disturbance terms. The instrumental variable approach allows me to control for both the endogeneity of schooling and the correlation between the disturbance terms in the two equations. The estimation results are again robust to this alternative specification.

<sup>13</sup> Following the convention in the empirical literature, I report the marginal effect of each individual family characteristic holding other family endowments constant. Various types of family characteristics may be correlated with each other. Consequently, the 'actual' effect of a single covariate is likely to be even bigger than the effect I report holding other covariates constant.

0.73 percentage points (or 0.82%). A one percentage point increase in the county level employment growth rate from 1980 to 1982 decreases the probability by 0.25 percentage points (or 0.29%).<sup>14</sup>

The results for the time eligible to drop out are noteworthy. I create this variable by first identifying both the state of residence and the date of birth of each 1980 sophomore. I then construct the time from which an individual is eligible to drop out. Specifically, time eligible to drop out = (date eligible to drop out – January 1, 1980) ÷ 365 days. To separate the age effect from the compulsory attendance effect, I include both a fourth-order polynomial in age and a fourth-order polynomial in the time eligible to drop out in the schooling equation. I plot in Figure 1 the posterior means and standard deviations of the predictive probability of high school completion in response to compulsory attendance.

As expected, I find that the time eligible to drop out has a positive and statistically significant effect on the probability of completing high school, although the marginal effect of compulsory attendance diminishes. From the top panel of Table II, the coefficients of the fourth-order

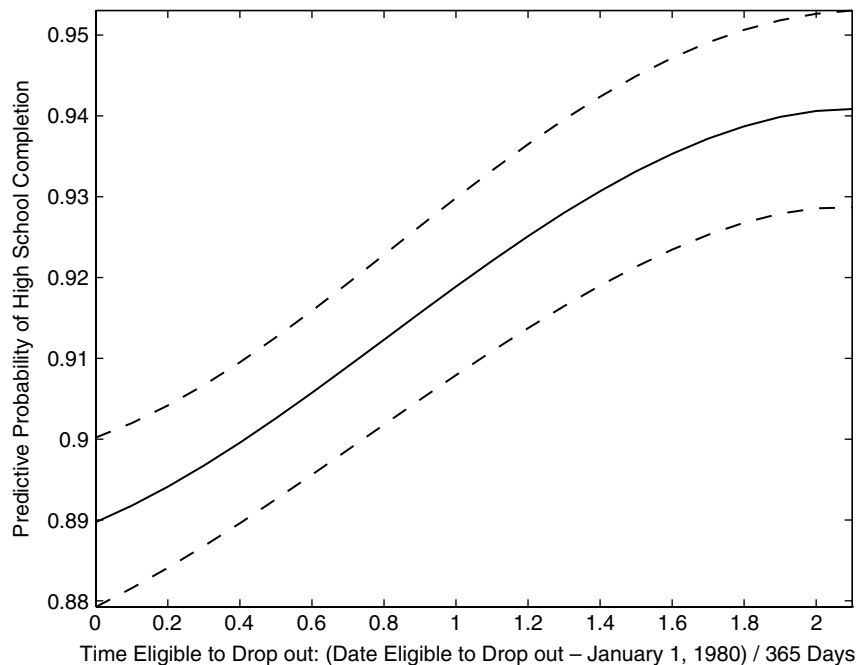


Figure 1. Posterior means (solid line) and posterior means plus and minus one posterior standard deviation (dashed lines) of the predictive probability of high school completion in response to the time eligible to drop out  $P(y_h = 4 | \text{Time eligible to drop out, Data})$

<sup>14</sup> The county level employment growth rate between 1980 and 1982 is only related to the 1980 sophomores when they were still in high school. I include this variable in the schooling equation to capture the opportunity cost of staying in high school as opposed to being in the labour market. Ideally, I would be able to incorporate in the unemployment equation the county level employment growth rate associated with the 1980 sophomores after they left high school. Unfortunately, HSB releases to the public neither this information nor the geographical locations of the individuals. If I assume both perfect correlation in the county level employment growth rates over time and no movement of the individuals between counties after high school, I can then include the county level employment growth rate between 1980 and 1982 in both the schooling and unemployment equations. However, the estimation results virtually do not change with this alternative specification.

polynomial in the time eligible to drop out (Time, Time<sup>2</sup>, Time<sup>3</sup>, Time<sup>4</sup>) are 0.096, 0.090, -0.0092 and -0.0094, respectively. A 1 year increase in the time eligible to drop out increases the probability of completing high school by 2.9 percentage points (or 3.3%), and a 2 year increase in the time eligible to drop out increases the probability by 5.1 percentage points (or 5.7%), suggesting a decreasing marginal effect of compulsory attendance on high school completion.<sup>15</sup>

The coefficients of the fourth-order polynomial in age (Age, Age<sup>2</sup>, Age<sup>3</sup>, Age<sup>4</sup>) are -0.41, -0.19, -0.03 and 0.03, respectively. Being 1 year older is associated with a 16.5 percentage point (or 19%) decrease in the probability of completing high school, which may reflect partly the effects of repeating or skipping grades. Regardless of the age effect, the sample is representative of the high school sophomores in 1980 and has limited variation in age. For example, 1.8% of the individuals were born in 1962, 21.2% born in 1963, 75.4% born in 1964 and 1.2% born in 1965. From Table I, the standard deviations of age and the time eligible to drop out are 0.5 and 0.84, respectively. Therefore, a substantial amount of the variation in the time eligible to drop out is attributable to the variation in compulsory attendance as opposed to the variation in age.

I now turn to the results for unemployment (bottom panel of Table II). An increase of \$10 000 in parental income decreases the proportion of time the individual is unemployed by 0.92 percentage points (from 8.95 to 8.03 percentage points, or 11%).<sup>16</sup> A one standard deviation increase in the base year cognitive test score decreases unemployment by 1.3 percentage points (or 15%). Having one more sibling increases unemployment by 0.18 percentage points (or 2.1%). Having some post-secondary education reduces unemployment by 0.39 percentage points (or 4.3%).<sup>17</sup> The four rows of  $\{\theta_k\}_{k=1}^4$  in the bottom panel of Table II report the common means of school-specific returns to education. As compared with completing high school, dropping out after completing the ninth grade, on average, increases the proportion of time the individual is unemployed by 7.9 percentage points (or 95%). The effect of dropping out after completing the tenth grade is 4.9 percentage points (or 58%), and the effect of dropping out after completing the eleventh grade is 4.3 percentage points (or 51%).

To measure the variability in returns to education across schools, the most relevant parameter is the  $4 \times 4$  common covariance matrix  $\Sigma_\theta$ , which captures the variation in the  $4 \times 1$  vectors of school-specific returns to education  $\{\theta_s\}_{s=1}^S$ . Small variance estimates lend little posterior support to the school-level heterogeneity, while big variance estimates provide strong evidence of variation across schools (Tobias and Li, 2003; Koop and Tobias, 2004). From Table III, the posterior means of the variance parameters in  $\Sigma_\theta$  are:  $E(\Sigma_{\theta_{11}}|Data) = 0.0068$ ,  $E(\Sigma_{\theta_{22}}|Data) = 0.0058$ ,  $E(\Sigma_{\theta_{33}}|Data) = 0.0058$  and  $E(\Sigma_{\theta_{44}}|Data) = 0.0034$ . The marginal posterior distributions of these parameters are located in areas away from zero. To see this, the posterior standard

<sup>15</sup> Acemoglu and Angrist (2000) use census data from 1960–80 and find that men exposed to more restrictive compulsory schooling laws were 1–4 percentage points more likely to complete grades 8–12. Lochner and Moretti (2003) use census data and find that in states/years requiring 11 or more years of compulsory attendance, the number of high school dropouts is 5.5% lower than in states/years requiring 8 years or less. Lleras-Muney (2002) uses 1960 census data and finds that legally requiring children to attend school for 1 year more increased educational attainment by about 5%.

<sup>16</sup> To report the marginal effects of the control variables on unemployment, I define the representative individual to have a parental income of \$20 000, a base year cognitive test score of 0, a father with an education of 12 years, a mother with an education of 12 years, two siblings, a high school diploma, no post-secondary education, and to be male, white and born on January 1, 1964. The predictive mean of the proportion of time the individual is unemployed is 0.0895.

<sup>17</sup> I choose not to model the post-secondary education as an endogenous covariate because the data currently available does not provide a good instrument for this variable. Since individuals who have graduated from high school are no longer constrained by the compulsory attendance, the time eligible to drop out only serves as a valid instrument for high school completion, not for post-secondary education.

deviations of these parameters are quite small:  $\text{Std}(\Sigma_{\theta_{11}}|Data) = 0.0085$ ,  $\text{Std}(\Sigma_{\theta_{22}}|Data) = 0.0056$ ,  $\text{Std}(\Sigma_{\theta_{33}}|Data) = 0.0058$  and  $\text{Std}(\Sigma_{\theta_{44}}|Data) = 0.0012$ .

The fact that the posterior mean of  $\theta_1$  is 0.16 (from the bottom panel of Table II) and the posterior mean of  $\Sigma_{\theta_{11}}$  is 0.0068 (from Table III) implies the school-specific effects of dropping out after completing the ninth grade  $\{\theta_{1s}\}_{s=1}^S$  are ‘drawn from’ a distribution with a mean of 0.16 and a variance of 0.0068. Given the multivariate normality assumption of the school-level heterogeneity, a 95% probability interval of these school-specific effects would be  $(-0.0016, 0.32)$ , indicating a substantial amount of heterogeneity in returns to education across schools. The 95% probability interval for the school-specific effects of dropping out after completing the tenth grade  $\{\theta_{2s}\}_{s=1}^S$  and the 95% interval for the effects of dropping out after completing the eleventh grade  $\{\theta_{3s}\}_{s=1}^S$  are  $(-0.039, 0.26)$  and  $(-0.051, 0.25)$ , respectively, again showing large amounts of school-level heterogeneity. To assess the appropriateness of the multivariate normality assumption of the school-level heterogeneity, I plot in Figure 2 the kernel-smoothed densities of the returns to education posterior means for every school in the sample  $\{E(\theta_s|Data)\}_{s=1}^S$ . The graphs suggest that normality serves as a good approximation to the underlying distributions of the school-level heterogeneity.

From the bottom few lines of Table III, the cutoff point estimate  $\gamma_3$  is 0.67. The variance estimate of the individual-level unobservables in the schooling equation  $\Sigma_{hh}$  is 0.92. The variance estimate in the unemployment equation  $\Sigma_{uu}$  is 0.11. The covariance estimate between the two unobservables  $\Sigma_{hu}$  is  $-0.0098$  with a posterior probability of being positive of 0.28, suggesting slight evidence of a negative correlation between the two unobservables.<sup>18</sup> To measure the potential bias resulting

Table III. Posterior means  $E(\cdot|Data)$ , standard deviations  $\text{Std}(\cdot|Data)$  and probabilities of being positive  $P(\cdot > 0|Data)$  of the elements of the covariance matrix  $\Sigma_{\theta}$ , cutoff point  $\gamma_3$  and covariance matrix  $\Sigma$

	$E(\cdot Data)$	$\text{Std}(\cdot Data)$	$P(\cdot > 0 Data)$
Covariance matrix $\Sigma_{\theta}$			
$\Sigma_{\theta_{11}}$	0.00677	0.00846	1
$\Sigma_{\theta_{12}}$	0.0027	0.00517	0.738
$\Sigma_{\theta_{13}}$	0.00362	0.00587	0.813
$\Sigma_{\theta_{14}}$	0.00254	0.00236	0.889
$\Sigma_{\theta_{22}}$	0.00576	0.00562	1
$\Sigma_{\theta_{23}}$	0.00198	0.00419	0.681
$\Sigma_{\theta_{24}}$	0.00183	0.00231	0.784
$\Sigma_{\theta_{33}}$	0.00579	0.00579	1
$\Sigma_{\theta_{34}}$	0.00172	0.00248	0.789
$\Sigma_{\theta_{44}}$	0.0034	0.00115	1
Cutoff point $\gamma_3$	0.671	0.0219	1
Covariance matrix $\Sigma$			
$\Sigma_{hh}$	0.918	0.0854	1
$\Sigma_{uu}$	0.109	0.0037	1
$\Sigma_{hu}$	$-0.00984$	0.0186	0.281

<sup>18</sup> The literature provides mixed evidence on the endogeneity bias in the estimates of economic returns to schooling. For example, Angrist and Krueger (1991) use the census data from various years and find that their instrumental variables estimate of the earnings return to education is close to the ordinary least squares estimate. Harmon and Walker (1995) use a large UK sample and find the presence of a large and negative bias in the least squares estimate of the schooling–earnings relationship. Oreopoulos (2003) compares results across the United States, Canada and the United Kingdom, and finds that the magnitude of the effect is similar across countries, and similar compared to corresponding OLS estimates.

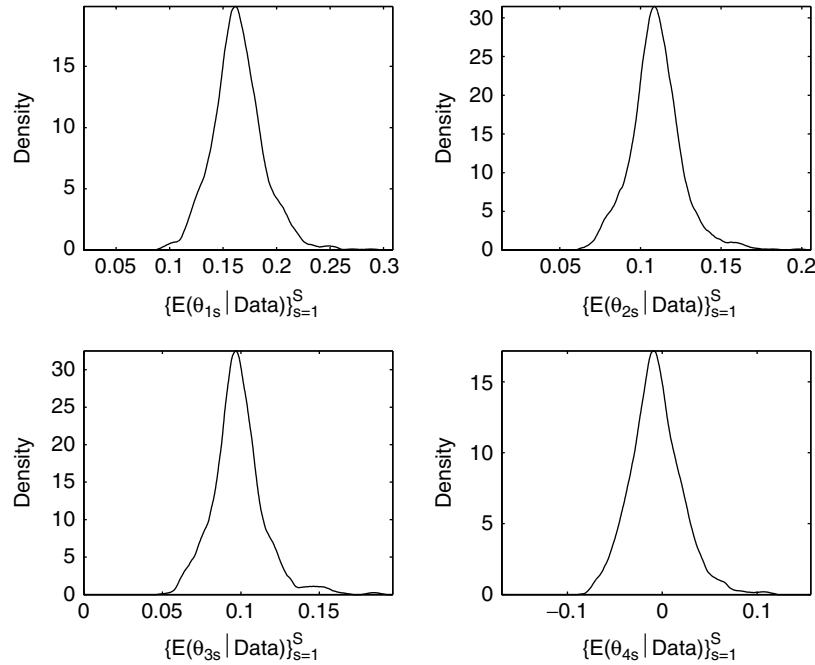


Figure 2. Kernel-smoothed densities of returns to education posterior means for all schools in the sample  $\{E(\theta_s | Data)\}_{s=1}^S$

from ignoring the correlation between the two unobservables, I re-estimated the model in single equations by restricting  $\Sigma_{hu} = 0$  in the original simultaneous equation model. To this end, I make slight changes to the priors and the Gibbs sampler in the last section. For the priors, instead of assuming  $\Sigma \sim IW(\rho, \rho R)$ , I assume  $\Sigma_{hh} \sim IG\left(\frac{\rho}{2}, \frac{2}{\rho R_{hh}}\right)$  and  $\Sigma_{uu} \sim IG\left(\frac{\rho}{2}, \frac{2}{\rho R_{uu}}\right)$ , where  $IG(\alpha, \beta)$  denotes the inverted Gamma distribution with mean  $\frac{1}{(\alpha-1)\beta}$  and variance  $\frac{1}{(\alpha-1)^2(\alpha-2)\beta^2}$ ,  $\rho = 6$ ,  $R_{hh} = 1$  and  $R_{uu} = 1$ . The Gibbs sampler remains unchanged, except for step 5. Instead of sampling the covariance matrix  $\Sigma$ , I sample the variance parameters  $\Sigma_{hh}$  and  $\Sigma_{uu}$ , where  $\Sigma_{hh} | \Xi_{-\Sigma_{hh}}, Data \sim IG\left[\frac{\rho+n}{2}, \frac{2}{\rho R_{hh} + (z_h - X_h \beta_h)'(z_h - X_h \beta_h)}\right]$  and  $\Sigma_{uu} | \Xi_{-\Sigma_{uu}}, Data \sim IG\left[\frac{\rho+n}{2}, \frac{2}{\rho R_{uu} + (z_u - X_u \beta_u - \bar{Q}\theta)'(z_u - X_u \beta_u - \bar{Q}\theta)}\right]$ .

The results obtained from the single equation model are very close to those obtained from the simultaneous equation model. There is slight evidence of upward bias in the estimated marginal effects of schooling on unemployment. In the single-equation model, dropping out after completing the ninth, tenth and eleventh grades increases the proportion of time the individual is unemployed by 9.4, 5.8 and 5.1 percentage points, respectively. In the simultaneous equation model, these marginal effects are 7.9, 4.9 and 4.3 percentage points, respectively. The bigger marginal effects estimated from the single-equation model are consistent with the hypothesis of a negative correlation between the two unobservables.

To forecast the policy effect of compulsory attendance on high school grade completion  $\{P(y_h = g)\}_{g=1}^4$  and unemployment  $E(w_u)$ , I simulate these predictive functions of interest for students with different characteristics (most disadvantaged, disadvantaged, average, advantaged

Table IV. Posterior marginal effects of different compulsory attendance ages (16 or 18) on high school grade completion  $\{P(y_h = g|Data)\}_{g=1}^4$  and unemployment  $E(w_u|Data)$  for students with different characteristics (most disadvantaged, disadvantaged, average, advantaged and most advantaged)

	Posterior functions of interest				
	$P(y_h = 1)^a$	$P(y_h = 2)^b$	$P(y_h = 3)^c$	$P(y_h = 4)^d$	$E(w_u)^e$
Most disadvantaged					
Compulsory 16	0.184	0.236	0.136	0.445	0.183
Compulsory 18	0.108	0.186	0.127	0.579	0.172
Disadvantaged					
Compulsory 16	0.0569	0.132	0.106	0.705	0.131
Compulsory 18	0.0278	0.0839	0.0789	0.809	0.125
Average student					
Compulsory 16	0.0117	0.0467	0.0519	0.89	0.0895
Compulsory 18	0.00468	0.0239	0.0308	0.941	0.0873
Advantaged					
Compulsory 16	0.00159	0.0106	0.016	0.972	0.0619
Compulsory 18	0.000524	0.00439	0.00762	0.987	0.0614
Most advantaged					
Compulsory 16	0.000146	0.00157	0.00315	0.995	0.0427
Compulsory 18	3.98e-005	0.000532	0.00122	0.998	0.0426

<sup>a</sup>  $P(y_h = 1|Data)$  denotes the predictive probability of dropping out after completing the ninth grade.

<sup>b</sup>  $P(y_h = 2|Data)$  denotes the predictive probability of dropping out after completing the tenth grade.

<sup>c</sup>  $P(y_h = 3|Data)$  denotes the predictive probability of dropping out after completing the eleventh grade.

<sup>d</sup>  $P(y_h = 4|Data)$  denotes the predictive probability of completing high school.

<sup>e</sup>  $E(w_u|Data)$  denotes the predictive mean of the proportion of time the individuals are unemployed.

and most advantaged) and list the results in Table IV.<sup>19</sup> For average students, increasing the compulsory attendance age from 16 to 18 decreases the probability of dropping out after completing the ninth grade by 0.7 percentage points (from 1.2 to 0.47 percentage points, or 60%), reduces the probability of dropping out after completing the tenth grade by 2.3 percentage points (or 49%), decreases the probability of dropping out after completing the eleventh grade by 2.1 percentage points (or 41%), increases the probability of completing high school by 5.1 percentage points (or 5.7%), and reduces the proportion of time the individuals are unemployed by 0.22 percentage points (or 2.5%).

The fact that increasing the compulsory attendance age from 16 to 18 increases the probability of completing high school from 89 to 94.1 percentage points for average students suggests the following: if potential dropouts were forced to remain in school under compulsory schooling laws for 2 more years after they reach age 16, then their graduation rate would increase to  $\frac{94.1-89}{100-89} = 46\%$ .

<sup>19</sup> The most disadvantaged students are defined to have a parental income of \$5000, a base year cognitive test score of negative two standard deviations, a father with an education of 8 years, a mother with an education of 8 years, two siblings, no post-secondary education, and to be male, white, residing in a county with an employment growth rate from 1980 to 1982 of zero percentage points, and born on January 1, 1964.

Everything else being the same, disadvantaged students have a parental income of \$10 000, a base year cognitive test score of negative one standard deviation, a father with an education of 10 years and a mother with an education of 10 years. Average students have a parental income of \$20 000, a base year cognitive test score of zero, a father with an education of 12 years and a mother with an education of 12 years. Advantaged students have a parental income of \$30 000, a base year cognitive test score of one standard deviation, a father with an education of 14 years and a mother with an education of 14 years. The most advantaged students have a parental income of \$40 000, a base year cognitive test score of two standard deviations, a father with an education of 16 years and a mother with an education of 16 years.



This number is 24%, 35%, 54% and 60% for the most disadvantaged, disadvantaged, advantaged and most advantaged students, respectively.<sup>20</sup> Angrist and Krueger (1991) use the census data and find that roughly 25% of the potential dropouts remain in school because of compulsory schooling laws. Eckstein and Wolpin (1999) develop and structurally estimate a sequential model of high school attendance and work decisions using data from the NLSY. They find that if dropouts were forced to remain in school for 5 years after entry without working, then their graduation rate would increase only to 13%.

From Table IV, the effects of compulsory attendance on high school completion and unemployment are most pronounced for the most disadvantaged students. For this group of students, increasing the compulsory attendance age from 16 to 18 decreases the probability of dropping out after completing the ninth, tenth and eleventh grade by 7.6 (or 41%), 5 (or 21%) and 0.92 (or 6.8%) percentage points, respectively, increases the probability of completing high school by 13 percentage points (or 31%), and reduces the proportion of time the individuals are unemployed by 1.1 percentage points (or 6.1%). The effect of compulsory attendance is also evident for disadvantaged students, and the corresponding marginal effects on high school grade completion and unemployment are -2.9 (or -51%), -4.8 (or -36%), -2.7 (or -26%), 10 (or 15%) and -0.62 (or -4.8%) percentage points, respectively. These marginal effects are much smaller for advantaged students: -0.11 (or -67%), -0.62 (or -58%), -0.84 (or -52%), 1.6 (or 1.6%) and -0.05 (or -0.8%) percentage points, respectively, and for the most advantaged students: -0.011 (or -73%), -0.1 (or -66%), -0.2 (or -61%), 0.31 (or 0.31%) and -0.007 (or -0.17%) percentage points, respectively.

Lleras-Muney (2002) finds that compulsory school attendance increases the education only of those in the lower percentiles of the education distribution, thereby decreasing education inequality, perhaps by as much as 15%. My results also suggest that compulsory attendance has an impact not only on average educational attainment and employment, but also on the distribution of education and employment as a whole. Compulsory schooling laws substantially lower the inequality in education and employment by increasing the education level and employment of those disadvantaged students.

### 3. HIGH SCHOOL DURATION AND FUTURE UNEMPLOYMENT DURATION

#### 3.1. Data

In the first follow-up survey of HSB, high school dropouts of the 1980 sophomores reported the month in which they left high school. Consequently, high school duration is defined to be from the beginning of the base year survey (February 1980), and censored at the end of the first follow-up survey (June 1982), when the nondropouts of the 1980 sophomores were expected to graduate.<sup>21</sup> Consequently, it is natural to specify a month-specific dropout hazard and assume that the dropouts leave high school in the middle of the month. An individual month-specific dummy variable is also constructed to indicate the eligibility of dropping out of each 1980 sophomore in each month.

<sup>20</sup> The experiments carried out are most likely to be heavily discounted by those who believe in potential general equilibrium effects caused by a change in the compulsory attendance regulations.

<sup>21</sup> The sample includes both high school graduates and dropouts. The high school durations of the graduates are considered censored at the time point they graduate. Because the focus of this section is on the effects of compulsory attendance on the dropout hazard and the hazard of leaving unemployment, the sample does not include the individuals who were never unemployed. This sample is more restricted than the sample used in the last section.

Table V. Descriptive statistics for the data: high school duration and unemployment duration

	Sample mean	Standard error
Parental income (\$10 000)	1.96	1.06
Base year cognitive test	0	1
Father's education	11.4	4.85
Mother's education	11.6	3.88
Number of siblings	2.78	1.88
Female	0.502	0.5
Minority	0.316	0.465
County level employment growth rate 80–82 (%)	−0.124	5.79
Age <sup>a</sup>	−0.28	0.551
Dropout eligibility	0.73	0.446
High school duration (month) <sup>b</sup>	25.8	6.4
Post-secondary education	0.583	0.493
Occupation <sup>c</sup>		
Professional	0.0241	0.154
Manager	0.0187	0.136
Salesman	0.619	0.486
Craftsman	0.0818	0.274
Industry <sup>d</sup>		
Agriculture	0.0305	0.172
Mining	0.0064	0.0798
Construction	0.0675	0.251
Manufacturing	0.0966	0.295
Transportation	0.0281	0.165
Trade	0.447	0.497
Unemployment duration (month)	2.65	3.19

<sup>a</sup> Age is the age of the individual on January 1, 1980, minus 16.

<sup>b</sup> High school duration is defined to be from February 1980.

<sup>c</sup> For occupation, the excluded group is labourers.

<sup>d</sup> For industry, the excluded group is the service industry.

The follow-up survey of HSB collected the number of weeks an individual was unemployed after his previous employment. The individual reported his occupation (professional, manager, salesman, craftsman or labourer) and the industry he worked in (agriculture, mining, construction, manufacturing, transportation, trade or service) before unemployment. The unemployment duration started from the end of the previous job and was censored at the end of the follow-up survey. I assume a month-specific hazard of leaving unemployment and convert the number of weeks unemployed into months. In the sample, 1052 individuals reported one unemployment spell, 318 individuals reported two unemployment spells, 90 individuals reported three unemployment spells, and 18 individuals reported four unemployment spells. The sample contains 1478 individuals with 2030 unemployment spells. In Table V, I present the descriptive statistics for the data used in this section.

### 3.2. Model

It is possible that unemployment behaviour and high school dropout behaviour are characterized by strong state dependence, so that the probabilities of leaving unemployment and leaving high school today depend on the amount of time an individual has been unemployed and has been in

school in the past. To estimate the conditional probability an individual leaves unemployment in any particular time period of his unemployment spell, and to determine the causes of variation between unemployed persons in the length of time they are unemployed, is a matter of great importance (Lancaster, 1979; Nickell, 1979; Lancaster and Nickell, 1980; Meyer, 1990; Han and Hausman, 1990; Campolieti, 1997, 2000, 2001).

High school duration also contains more information on dropout behaviour than the binary outcome of high school completion. The timing of high school dropout decisions can be considered as a sequence of binary outcomes of dropping out over time. Modelling high school duration and unemployment duration also captures the effects of time-varying covariates (e.g. the dropout eligibility) and the effects of spell-specific covariates (e.g. the occupation of the individual and the industry he worked in before unemployment).

The probability distribution of a duration can be specified by the distribution function  $F(t) = \Pr(T < t)$ , which specifies the probability that the random variable  $T$  is less than some value  $t$ . The corresponding density function is  $f(t) = dF(t)/dt$ . The survivor function  $S(t) = 1 - F(t) = \Pr(T \geq t)$  gives the probability that the random variable  $T$  will equal or exceed the value  $t$ . A particularly useful function for duration analysis is the hazard function  $\lambda(t) = f(t)/S(t)$ . Roughly,  $\lambda(t)$  is the rate at which a spell will be completed at duration  $t$ , given that they last until  $t$  (Cox, 1972; Kiefer, 1988). Following Meyer (1990), Han and Hausman (1990) and Campolieti (1997, 2001),<sup>22</sup> I assume a month-specific baseline dropout hazard  $\lambda_{hm}$  ( $h$  labels the dropout hazard and  $m$  indicates the month).

Consider a proportional hazard specification (Cox, 1972; Kiefer, 1988) in which the dropout hazard of individual  $i$  in month  $m$  is  $\lambda_{him} = \exp(x_{him}\beta_h)\xi_{hi}\lambda_{hm}$ , where  $x_{him}$  is a  $1 \times k_h$  vector of individual-level characteristics in month  $m$  (base year cognitive test score, parental income, parental education, number of siblings, gender, race, age, county level employment growth rate between 1980 and 1982, and individual month-specific dropout eligibility), and  $\xi_{hi}$  indicates the unobserved heterogeneity in the dropout hazard.

Similarly, I define a month-specific baseline hazard of leaving unemployment  $\lambda_{uw}$  ( $u$  labels the hazard of leaving unemployment and  $w$  denotes the month). Assume that individual  $i$  has  $S_i$  unemployment spells. The hazard of leaving unemployment of individual  $i$  in month  $w$  of unemployment spell  $s$  is  $\lambda_{uisw} = \exp(x_{uisw}\beta_u)\xi_{ui}\lambda_{uw}$ , where  $x_{uisw}$  is a  $1 \times k_u$  vector of individual-level variables in month  $w$  of unemployment spell  $s$  (base year cognitive test score, parental income, parental education, gender, race, age, dummy variable indicating any post-secondary education, occupation of the individual and the industry he worked in before unemployment, and high school duration), and  $\xi_{ui}$  indicates the unobserved heterogeneity in the hazard of leaving unemployment.

Note that the hazard of leaving unemployment is a function of high school duration. Therefore, the model is a bivariate duration model with endogenous covariates. While the concepts of simultaneous equations have been extended quite fruitfully to limited dependent variable models, their application to hazard or duration models is still a growing literature. To control for the endogeneity of high school duration, I follow Lillard (1993) and Van den Berg (2001) allowing for the correlation between  $\ln \xi_{hi}$  and  $\ln \xi_{ui}$ :<sup>23</sup>  $(\ln \xi_{hi} \quad \ln \xi_{ui})' \sim N(0_{2 \times 1}, \Sigma)$ , where  $\Sigma$  is the covariance matrix.

<sup>22</sup> Fully nonparametric specifications of the hazard rate (Ruggiero, 1994; Gørgens and Horowitz, 1999; Horowitz, 1999; Campolieti, 2000) are implementable, but are computationally demanding. The assumption of a month-specific hazard rate is reasonably flexible.

<sup>23</sup> It has become increasingly popular (Eckstein and Wolpin, 1999; Keane and Wolpin, 1997; Belzil and Hansen, 2002) to assume a multivariate discrete distribution with a finite mixture of supports for the unobserved heterogeneity. This is

Let  $t_{him}$  denote the high school duration in month  $m$  of individual  $i$ ,  $d_{him}$  denote the dummy variable indicating the act of dropping out in month  $m$ ,  $t_{uisw}$  denote the unemployment duration in month  $w$  of unemployment spell  $s$ , and  $d_{uisw}$  denote the dummy variable indicating the act of leaving unemployment in month  $w$  of unemployment spell  $s$ . Note that the hazard function  $\lambda(t)$  is the rate at which spells will be completed at duration  $t$ , given that they last until  $t$ :  $\lambda(t) = f(t)/S(t)$ , where  $S(t) = 1 - \int_0^t f(s)ds$ . It can be verified that  $S(t) = \exp\left[-\int_0^t \lambda(s)ds\right]$  by taking differentiation on both sides of the equation. Therefore  $f(t) = S(t)\lambda(t) = \exp\left[-\int_0^t \lambda(s)ds\right]\lambda(t)$ .

The likelihood function of this bivariate duration model with endogenous covariates is proportional to:

$$\begin{aligned}
 p(Data|\Xi) \propto & \left\{ \prod_{i=1}^I \prod_{m=1}^M \exp[-\exp(x_{him}\beta_h)\xi_{hi}\lambda_{hm}t_{him}][\exp(x_{him}\beta_h)\xi_{hi}\lambda_{hm}]^{d_{him}} \right\} \\
 & \times \left\{ \prod_{i=1}^I \prod_{s=1}^{S_i} \prod_{w=1}^W \exp[-\exp(x_{uisw}\beta_u)\xi_{ui}\lambda_{uw}t_{uisw}][\exp(x_{uisw}\beta_u)\xi_{ui}\lambda_{uw}]^{d_{uisw}} \right\} \\
 & \times \left\{ \prod_{i=1}^I \xi_{hi}^{-1} \xi_{ui}^{-1} |2\pi\Sigma|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\ln \xi_{hi} \quad \ln \xi_{ui})\Sigma^{-1}(\ln \xi_{hi} \quad \ln \xi_{ui})'\right] \right\} \quad (4)
 \end{aligned}$$

where  $I$  indicates the number of individuals. I develop a Bayesian estimation method, which avoids the direct evaluation of the nontrivial likelihood function, and draws instead from the exact posterior of the bivariate duration model. With some data augmentation steps of the unobserved heterogeneity in the dropout hazard and the hazard of leaving unemployment, the estimation does not rely on any approximation methods or asymptotic theory. In addition, any posterior function of interest can be estimated. For a bivariate duration model, the list of posterior functions of interest could include the marginal effects of the control variables on the dropout hazard  $E(\Delta\% \lambda_{hm}^* | \Delta x_h, Data)$ , their effects on the hazard of leaving unemployment  $E(\Delta\% \lambda_{uw}^* | \Delta x_u, Data)$ , duration density functions  $P(t_h | Data)$ , expected durations  $E(t_h | Data)$  and hazard elasticities  $E(\Delta\% \lambda_{hm}^* | \Delta\% x_h, Data)$ .

Using the Bayes' rule, the joint posterior distribution of the parameters  $p(\Xi | Data)$  is proportional to the product of the data augmented likelihood function  $p(Data | \Xi)$  in equation (4) and the joint prior distribution of the parameters  $p(\Xi)$ :  $p(\Xi | Data) \propto p(Data | \Xi)p(\Xi)$ . To complete the Bayesian analysis and obtain the posterior, it is necessary to discuss the priors  $p(\Xi)$ . I assume prior independence across parameters:  $p(\Xi) = [\prod_{m=1}^M p(\lambda_{hm})][\prod_{w=1}^W p(\lambda_{uw})]p(\beta_h)p(\beta_u)p(\Sigma)$ . Prior distributions of all the parameters  $\{p(\lambda_{hm})\}_{m=1}^M$ ,  $\{p(\lambda_{uw})\}_{w=1}^W$ ,  $p(\beta_h)$ ,  $p(\beta_u)$  and  $p(\Sigma)$  are specified as follows:  $\lambda_{hm} \sim G(a_h, b_h)$ , for  $m = 1, 2, 3, \dots, M$ ,  $\lambda_{uw} \sim G(a_u, b_u)$ , for  $w = 1, 2, 3, \dots, W$ ,

---

in the spirit of Heckman and Singer (1984), who suggest that making a parametric assumption about the distribution of the unobserved heterogeneity can bias the estimates. However, most empirical researchers have reported computational difficulties when trying to estimate a hazard model that includes both the piecewise constant baseline hazard and Heckman and Singer's discrete unobserved heterogeneity distribution.

The approach that has been used to solve these problems is to either place restrictions on the baseline hazard (i.e., a parametric specification) or place restrictions on the unobserved heterogeneity distribution (i.e., a parametric assumption instead of the Heckman and Singer style unobserved heterogeneity distribution). Given the complexity of my model, it is more realistic from a computational perspective to rely on the parametric assumption rather than on the discrete distribution. I thank the two anonymous referees for pointing out this trade-off to me.

$\beta_h \sim N(\beta_{h0}, V_{\beta_h})$ ,  $\beta_u \sim N(\beta_{u0}, V_{\beta_u})$  and  $\Sigma \sim IW(\rho, \rho R)$ , where  $G(\alpha, \beta)$  denotes the Gamma distribution with mean  $\alpha\beta$  and variance  $\alpha\beta^2$ ,  $a_h = 0.01$ ,  $b_h = 1$ ,  $a_u = 0.01$ ,  $b_u = 1$ ,  $\beta_{h0} = 0_{k_h \times 1}$ ,  $V_{\beta_h} = 1000I_{k_h}$ ,  $\beta_{u0} = 0_{k_u \times 1}$ ,  $V_{\beta_u} = 1000I_{k_u}$ ,  $\rho = 6$  and  $R = I_2$ .<sup>24</sup>

### 3.3. Gibbs Sampler

I use a Metropolis–Hastings within Gibbs algorithm to implement the Bayesian analysis, and iterate through sampling from the following complete conditional posterior distributions, which are derived from the joint posterior distribution  $p(\Xi|Data)$ .

1. Sample the baseline dropout hazard  $\lambda_{hm}$ :

$$\lambda_{hm} | \Xi_{-\lambda_{hm}}, Data \sim G \left\{ a_h + \sum_{i=1}^I d_{him}, \left[ b_h^{-1} + \sum_{i=1}^I \exp(x_{him}\beta_h)\xi_{hi}t_{him} \right]^{-1} \right\}$$

for  $m = 1, 2, 3, \dots, M$ .

2. Sample the baseline hazard of leaving unemployment  $\lambda_{uw}$ :

$$\lambda_{uw} | \Xi_{-\lambda_{uw}}, Data \sim G \left\{ a_u + \sum_{i=1}^I \sum_{s=1}^{S_i} d_{uisw}, \left[ b_u^{-1} + \sum_{i=1}^I \sum_{s=1}^{S_i} \exp(x_{uisw}\beta_u)\xi_{ui}t_{uisw} \right]^{-1} \right\}$$

for  $w = 1, 2, 3, \dots, W$ .

3. Sample the unobserved heterogeneity in the dropout hazard  $\xi_{hi}$ :

$$p[\xi_{hi} | \Xi_{-\xi_{hi}}, Data] \propto \left\{ \prod_{m=1}^M \xi_{hi}^{d_{him}} \exp[-\exp(x_{him}\beta_h)\xi_{hi}\lambda_{hm}t_{him}] \right\} \\ \times \xi_{hi}^{-1} \exp \left[ -\frac{1}{2} (\ln \xi_{hi} \quad \ln \xi_{ui}) \Sigma^{-1} (\ln \xi_{hi} \quad \ln \xi_{ui})' \right]$$

for  $i = 1, 2, 3, \dots, I$ . This conditional posterior distribution cannot be sampled directly, so I use the proposal density:  $\xi_{hi}^* | \Xi_{-\xi_{hi}}, Data \sim G \left\{ \alpha_h + \sum_{m=1}^M d_{him}, \left[ \sum_{m=1}^M \exp(x_{him}\beta_h)\lambda_{hm}t_{him} \right]^{-1} \right\}$ , where  $\alpha_h = 1$  is the tuning parameter. The probability of accepting the candidate draw is  $\min(R, 1)$ , where

$$R = \frac{\xi_{hi}^{*\alpha_h} \exp \left[ -\frac{1}{2} (\ln \xi_{hi}^* \quad \ln \xi_{ui}) \Sigma^{-1} (\ln \xi_{hi}^* \quad \ln \xi_{ui})' \right]}{\xi_{hi, j-1}^{-\alpha_h} \exp \left[ -\frac{1}{2} (\ln \xi_{hi, j-1} \quad \ln \xi_{ui}) \Sigma^{-1} (\ln \xi_{hi, j-1} \quad \ln \xi_{ui})' \right]}$$

<sup>24</sup> These priors are also very diffuse and essentially noninformative. With alternative assignments of the hyperparameters such as  $a_h = 0.001$  or  $a_h = 0.1$ ,  $a_u = 0.001$  or  $a_u = 0.1$ ,  $V_{\beta_h} = 100I_k$  or  $V_{\beta_h} = 10000I_k$ ,  $V_{\beta_u} = 100I_k$  or  $V_{\beta_u} = 10000I_k$ ,  $\rho = 4$  or  $\rho = 12$ ,  $R = 0.5I_2$  or  $R = 2I_2$ , estimation results virtually do not change.

4. Sample the unobserved heterogeneity in the hazard of leaving unemployment  $\xi_{ui}$ :

$$p[\xi_{ui} | \Xi_{-\xi_{ui}}, Data] \propto \left\{ \prod_{s=1}^{S_i} \prod_{w=1}^W \xi_{ui}^{d_{uisw}} \exp[-\exp(x_{uisw}\beta_u)\xi_{ui}\lambda_{uw}t_{uisw}] \right\} \\ \times \xi_{ui}^{-1} \exp \left[ -\frac{1}{2} (\ln \xi_{hi} \quad \ln \xi_{ui}) \Sigma^{-1} (\ln \xi_{hi} \quad \ln \xi_{ui})' \right]$$

for  $i = 1, 2, 3, \dots, I$ . This conditional posterior distribution also cannot be sampled directly, so I use the proposal density:  $\xi_{ui}^* | \Xi_{-\xi_{ui}}, Data \sim G \left\{ \alpha_u + \sum_{s=1}^{S_i} \sum_{w=1}^W d_{uisw}, \left[ \sum_{s=1}^{S_i} \sum_{w=1}^W \exp(x_{uisw}\beta_u) \lambda_{uw}t_{uisw} \right]^{-1} \right\}$ , where  $\alpha_u = 1$  is the tuning parameter. The probability of accepting the candidate draw is  $\min(R, 1)$ , where

$$R = \frac{\xi_{ui}^{*\alpha_u} \exp \left[ -\frac{1}{2} (\ln \xi_{hi} \quad \ln \xi_{ui}^*) \Sigma^{-1} (\ln \xi_{hi} \quad \ln \xi_{ui}^*)' \right]}{\xi_{ui,j-1}^{-\alpha_u} \exp \left[ -\frac{1}{2} (\ln \xi_{hi} \quad \ln \xi_{ui,j-1}) \Sigma^{-1} (\ln \xi_{hi} \quad \ln \xi_{ui,j-1})' \right]}$$

5. Sample the covariance matrix  $\Sigma$ :

$$\Sigma | \Xi_{-\Sigma}, Data \sim IW \left[ \rho + I, \rho R + \sum_{i=1}^I (\ln \xi_{hi} \quad \ln \xi_{ui})' (\ln \xi_{hi} \quad \ln \xi_{ui}) \right]$$

6. Sample the coefficients in the dropout hazard  $\beta_h$ :

$$p(\beta_h | \Xi_{-\beta_h}, Data) \propto \left\{ \prod_{i=1}^I \prod_{m=1}^M \exp[-\exp(x_{him}\beta_h)\xi_{hi}\lambda_{hm}t_{him}] [\exp(x_{him}\beta_h)\xi_{hi}\lambda_{hm}]^{d_{him}} \right\} \\ \times \exp \left[ -\frac{1}{2} (\beta_h - \beta_{h0})' V_{\beta_h}^{-1} (\beta_h - \beta_{h0}) \right]$$

This conditional posterior distribution cannot be sampled directly, so I use the proposal density:  $\beta_h^* | \beta_{h,j-1} \sim N(\beta_{h,j-1}, \Sigma_{\beta_h})$ , where  $\Sigma_{\beta_h} = \alpha_{\beta_h} \left( \sum_{i=1}^I \sum_{m=1}^M 1 \right)^{-1} \sum_{i=1}^I \sum_{m=1}^M x'_{him} x_{him}$  and  $\alpha_{\beta_h} = 0.001$  is the tuning parameter. The probability of accepting the candidate draw is  $\min(R, 1)$ , where

$$R = \left\{ \prod_{i=1}^I \prod_{m=1}^M \frac{\exp[-\exp(x_{him}\beta_h^*)\xi_{hi}\lambda_{hm}t_{him}] [\exp(x_{him}\beta_h^*)\xi_{hi}\lambda_{hm}]^{d_{him}}}{\exp[-\exp(x_{him}\beta_{h,j-1})\xi_{hi}\lambda_{hm}t_{him}] [\exp(x_{him}\beta_{h,j-1})\xi_{hi}\lambda_{hm}]^{d_{him}}} \right\} \\ \times \frac{\exp \left[ -\frac{1}{2} (\beta_h^* - \beta_{h0})' V_{\beta_h}^{-1} (\beta_h^* - \beta_{h0}) \right]}{\exp \left[ -\frac{1}{2} (\beta_{h,j-1} - \beta_{h0})' V_{\beta_h}^{-1} (\beta_{h,j-1} - \beta_{h0}) \right]}$$

7. Sample the coefficients in the hazard of leaving unemployment  $\beta_u$ :

$$p(\beta_u | \Xi_{-\beta_u}, Data) \propto \left\{ \prod_{i=1}^I \prod_{s=1}^{S_i} \prod_{w=1}^W \exp[-\exp(x_{uisw}\beta_u)\xi_{ui}\lambda_{uw}t_{uisw}][\exp(x_{uisw}\beta_u)\xi_{ui}\lambda_{uw}]^{d_{uisw}}] \right\} \\ \times \exp\left[-\frac{1}{2}(\beta_u - \beta_{u0})'V_{\beta_u}^{-1}(\beta_u - \beta_{u0})\right]$$

This conditional posterior distribution also cannot be sampled directly, so I use the proposal density:  $\beta_u^* | \beta_{u,j-1} \sim N(\beta_{u,j-1}, \Sigma_{\beta_u})$ , where  $\Sigma_{\beta_u} = \alpha_{\beta_u} \left( \sum_{i=1}^I \sum_{s=1}^{S_i} \sum_{w=1}^W 1 \right)^{-1} \sum_{i=1}^I \sum_{s=1}^{S_i} \sum_{w=1}^W x'_{uisw} x_{uisw}$  and  $\alpha_{\beta_u} = 0.0001$  is the tuning parameter. The probability of accepting the candidate draw is  $\min(R, 1)$ , where

$$R = \left\{ \frac{\prod_{i=1}^I \prod_{s=1}^{S_i} \prod_{w=1}^W \frac{\exp[-\exp(x_{uisw}\beta_u^*)\xi_{ui}\lambda_{uw}t_{uisw}][\exp(x_{uisw}\beta_u^*)\xi_{ui}\lambda_{uw}]^{d_{uisw}}]}{\exp[-\exp(x_{uisw}\beta_{u,j-1})\xi_{ui}\lambda_{uw}t_{uisw}][\exp(x_{uisw}\beta_{u,j-1})\xi_{ui}\lambda_{uw}]^{d_{uisw}}]} \right\} \\ \times \frac{\exp\left[-\frac{1}{2}(\beta_u^* - \beta_{u0})'V_{\beta_u}^{-1}(\beta_u^* - \beta_{u0})\right]}{\exp\left[-\frac{1}{2}(\beta_{u,j-1} - \beta_{u0})'V_{\beta_u}^{-1}(\beta_{u,j-1} - \beta_{u0})\right]}$$

I run this algorithm with 20000 iterations and discard the first 4000 iterations as the pre-convergence draws. To monitor the convergence of the algorithm, I simulate several independent sequences, with starting points sampled from an overdispersed distribution following Gelman *et al.* (1995).

### 3.4. Results

To examine the duration dependence of the dropout hazard and the hazard of leaving unemployment, I plot the posterior means and standard deviations of the dropout hazard  $\lambda_{hm}^*$ , for  $m = 1, 2, 3, \dots, M$ , and the hazard of leaving unemployment  $\lambda_{uw}^*$ , for  $w = 1, 2, 3, \dots, W$ , of a representative individual in Figures 3 and 4, respectively, where \* denotes a representative individual.<sup>25</sup>

The dropout hazard climbs up slightly from the first month, February 1980, peaks at the twenty-first month, October 1981, and dwindles afterwards. There are two dips in the dropout hazard, one around the sixth month, July 1980, and the other around the eighteenth and nineteenth months, July and August of 1981. The pattern of the dropout hazard may be consistent with the following conjecture. In the sophomore and junior years, the students are more challenged by the curriculum as they progress and, consequently, they are more likely to drop out. In the senior year, although

<sup>25</sup> To report the dropout hazard, I define the representative individual to have a parental income of \$20000, a base year cognitive test score of 0, a father with an education of 12 years, a mother with an education of 12 years, two siblings, and to be male, white, residing in a county with an employment growth rate from 1980 to 1982 of zero percentage points, born on January 1, 1964, and ineligible to drop out.

To report the hazard of leaving unemployment, I define the representative individual to have a parental income of \$20000, a base year cognitive test score of 0, a father with an education of 12 years, a mother with an education of 12 years, two siblings, a high school duration of 25 months (from February 1980), no post-secondary education, and to be male, white, born on January 1, 1964, and working as a labourer in the service industry before unemployment.

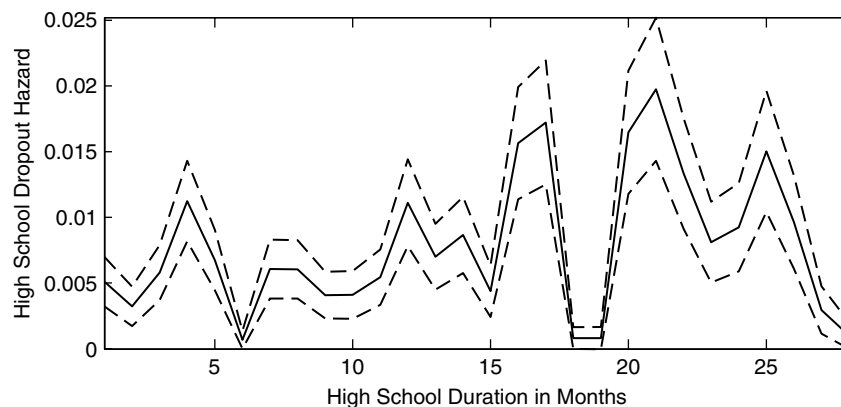


Figure 3. Posterior means (solid line) and posterior means plus and minus one posterior standard deviation (dashed lines) of the dropout hazard of a representative individual  $\lambda_{hm}^*$ , for  $m = 1, 2, 3, \dots, M = 28$



Figure 4. Posterior means (solid line) and posterior means plus and minus one posterior standard deviation (dashed lines) of the hazard of leaving unemployment of a representative individual  $\lambda_{uw}^*$ , for  $w = 1, 2, 3, \dots, W = 18$

the students are still challenged by the curriculum, they are also more enticed by the benefits associated with graduation as they approach the end of the curriculum, and, consequently, they are less likely to drop out. The hazard of leaving unemployment decreases slightly during the first year of unemployment, but rises at the end of the year. The hazard estimates for the second year are less precise because there are many fewer individuals who were still unemployed and remained in the risk set of leaving unemployment during the year.

In Table VI, I present the posterior means  $E(\beta|Data)$ , standard deviations  $Std(\beta|Data)$  and probabilities of being positive  $P(\beta > 0|Data)$  of the coefficients. For a bivariate duration model, the posterior means of the coefficients  $E(\beta|Data)$  can be interpreted as the marginal effects of the control variables on the dropout hazard  $E(\Delta\% \lambda_{hm}^* | \Delta x_h, Data)$  and the hazard of leaving unemployment  $E(\Delta\% \lambda_{uw}^* | \Delta x_u, Data)$ , where  $x_h$  and  $x_u$  denote the  $1 \times k_h$  and  $1 \times k_u$  vectors of individual-level characteristics of a representative individual. Nevertheless, this interpretation is



only a good approximation when the marginal effects are close to zero. To this end, I also calculate the *exact* marginal effects and list them in the last column of Table VI.

I first discuss the results for the dropout hazard (reported in the top panel of Table VI). Again, the coefficients on the individual-level variables generally have the expected signs and are statistically significant. A one standard deviation increase in the base year cognitive test score reduces the dropout hazard by 42%. A 1 year increase in the father's education decreases the hazard by 3.9%, and a 1 year increase in the mother's education reduces the hazard by 5.6%. A one percentage point increase in the county level employment growth rate from 1980 to 1982 increases the hazard

Table VI. Posterior means  $E(\beta|Data)$ , standard deviations  $Std(\beta|Data)$  and probabilities of being positive  $P(\beta > 0|Data)$  of the coefficients, and posterior marginal effects of the control variables on the dropout hazard  $E(\Delta\% \lambda_{hm}^* | \Delta x_h, Data)$  and the hazard of leaving unemployment  $E(\Delta\% \lambda_{uw}^* | \Delta x_u, Data)$

Explanatory variables <sup>a</sup>	High school dropout hazard $\lambda_{hm}^*$			
	$E(\beta D)^b$	$Std(\beta D)$	$P(\beta > 0 D)$	$E(\Delta\% \lambda_{hm}^*   \Delta x_h, D)^c$
Parental income	-0.00353	0.0614	0.481	-0.00164
Base year cognitive test	-0.548	0.0764	0	-0.42
Father's education	-0.04	0.0279	0.0838	-0.0389
Mother's education	-0.0584	0.0302	0.0323	-0.0563
Number of siblings	0.0182	0.038	0.699	0.0191
Female	-0.405	0.122	0	-0.328
Minority	-0.537	0.142	0	-0.409
County employment growth	0.0224	0.0109	0.979	0.0228
Age	0.792	0.105	1	1.22
Dropout eligibility	0.354	0.168	0.975	0.445
	Hazard of leaving unemployment $\lambda_{uw}^*$			
	$E(\beta D)$	$Std(\beta D)$	$P(\beta > 0 D)$	$E(\Delta\% \lambda_{uw}^*   \Delta x_u, D)^d$
Parental income	0.0456	0.0279	0.951	0.0471
Base year cognitive test	0.106	0.0351	1	0.113
Father's education	0.014	0.0175	0.817	0.0142
Mother's education	0.0308	0.0123	0.992	0.0313
Number of siblings	0.00318	0.0171	0.564	0.00333
High school duration	0.0141	0.00494	1	0.0142
Post-secondary education	0.0836	0.0527	0.958	0.0887
Female	-0.0266	0.0696	0.402	-0.0239
Minority	-0.0934	0.0688	0.0801	-0.087
Age	-0.0842	0.0391	0.035	-0.0801
Professional	-0.0423	0.218	0.446	-0.0188
Manager	0.257	0.201	0.898	0.319
Salesman	0.0921	0.091	0.846	0.101
Craftsman	0.152	0.122	0.89	0.173
Agriculture	-0.259	0.183	0.0843	-0.215
Mining	-0.0424	0.364	0.451	0.023
Construction	0.216	0.127	0.948	0.251
Manufacturing	0.0172	0.0971	0.586	0.0221
Transportation	-0.0783	0.15	0.293	-0.0648
Trade	0.0237	0.0656	0.632	0.0262

<sup>a</sup> The descriptive statistics for the data are reported in Table V.

<sup>b</sup>  $D$  denotes the data.

<sup>c</sup>  $E(\Delta\% \lambda_{hm}^* | \Delta x_h, Data)$  denotes the marginal effects of the control variables on the dropout hazard.

<sup>d</sup>  $E(\Delta\% \lambda_{uw}^* | \Delta x_u, Data)$  denotes the marginal effects of the control variables on the hazard of leaving unemployment.

by 2.3%. Being 1 year older raises the hazard by 122%. Being eligible to drop out increases the hazard by 45%.

Turning to the results for the hazard of leaving unemployment (bottom panel of Table VI), I find that an increase of \$10 000 in parental income increases the hazard of leaving unemployment by 4.7%. A one standard deviation increase in the base year cognitive test score raises the hazard by 11%. A 1 year increase in the father's education increases the hazard by 1.4% and a 1 year increase in the mother's education increases the hazard by 3.1%. A 1 month increase in the high school duration raises the hazard by 1.4%. Having some post-secondary education increases the hazard by 8.9%. As compared with working as a labourer, working as a manager or working as a craftsman before unemployment is associated with a higher chance of leaving unemployment. As compared with working in the service industry, working in the construction industry before unemployment is linked to a higher chance of leaving unemployment, while working in the agricultural industry is associated with a lower chance of leaving unemployment.

From Table VII, the variance estimate of the log unobserved heterogeneity in the dropout hazard  $\Sigma_{hh}$  is 0.31. The variance estimate of the log unobserved heterogeneity in the hazard of leaving unemployment  $\Sigma_{uu}$  is 0.306. The covariance estimate between the two unobservables  $\Sigma_{hu}$  is 0.019, with a posterior probability of being positive of 0.73, suggesting slight evidence of a positive correlation between the two unobservables. To measure the potential bias resulting from ignoring the correlation between the two unobservables, I re-estimated the model in single equations by restricting  $\Sigma_{hu} = 0$  in the original simultaneous equation model. To this end, I make slight changes in the priors and the Gibbs algorithm. For the priors, instead of assuming  $\Sigma \sim IW(\rho, \rho R)$ , I assume  $\Sigma_{hh} \sim IG\left(\frac{\rho}{2}, \frac{2}{\rho R_{hh}}\right)$  and  $\Sigma_{uu} \sim IG\left(\frac{\rho}{2}, \frac{2}{\rho R_{uu}}\right)$ , where  $\rho = 6$ ,  $R_{hh} = 1$  and  $R_{uu} = 1$ . The Gibbs sampler remains unchanged except for step 5. Instead of sampling the covariance matrix  $\Sigma$ , I sample the variance parameters  $\Sigma_{hh}$  and  $\Sigma_{uu}$ , where  $\Sigma_{hh} | \Xi_{-\Sigma_{hh}}, Data \sim IG\left[\frac{\rho+I}{2}, \frac{2}{\rho R_{hh} + \sum_{i=1}^I (\ln \xi_{hi})^2}\right]$  and  $\Sigma_{uu} | \Xi_{-\Sigma_{uu}}, Data \sim IG\left[\frac{\rho+I}{2}, \frac{2}{\rho R_{uu} + \sum_{i=1}^I (\ln \xi_{ui})^2}\right]$ .

The results obtained from the single-equation model are quite similar to those obtained from the simultaneous equation model. There is slight evidence of a downward bias in the estimated marginal effects of schooling on unemployment. In the single-equation model, a 1 month increase in the high school duration raises the hazard of leaving unemployment by 1.21%. In the simultaneous equation model, this marginal effect is 1.42%. The smaller marginal effect obtained from the single-equation model is consistent with the hypothesis of a positive correlation between the two unobservables.

Table VII. Posterior means  $E(\cdot|Data)$ , standard deviations  $Std(\cdot|Data)$  and probabilities of being positive  $P(\cdot > 0|Data)$  of the elements of the covariance matrix  $\Sigma$

	$E(\cdot Data)$	$Std(\cdot Data)$	$P(\cdot > 0 Data)$
Covariance matrix $\Sigma$			
$\Sigma_{hh}$	0.31	0.0592	1
$\Sigma_{uu}$	0.306	0.0462	1
$\Sigma_{hu}$	0.0193	0.0326	0.734

To forecast the policy effects of compulsory attendance on the probability of high school survival  $P(t_h = M|Data)$ , the high school duration  $E(t_h|Data)$ , the hazard of leaving unemployment in the first month of unemployment  $E(\lambda_{u1}^*|Data)$  and the unemployment duration  $E(t_u|Data)$ , I simulate these predictive functions of interest for students with different characteristics (most disadvantaged, disadvantaged, average, advantaged and most advantaged), and list the results in Table VIII. Formally,

1. The probability of high school survival:

$$P(t_h = M) = \exp\left(-\sum_{m=1}^M \lambda_{hm}^*\right) = \exp\left[-\sum_{m=1}^M \exp(x_h \beta_h) \lambda_{hm}\right]$$

where  $\lambda_{hm}^*$  indicates the dropout hazard of a representative individual,  $x_h$  is the  $1 \times k_h$  vector of individual-level characteristics of this representative individual, and  $\lambda_{hm}$  denotes the baseline dropout hazard.

2. The high school duration:

$$\begin{aligned} E(t_h) &= \int_0^\infty t f_h(t) dt = P(t_h = M)M + \sum_{m=1}^M \int_{m-1}^m t f_h(t) dt \\ &= \exp\left(-\sum_{m=1}^M \lambda_{hm}^*\right) M + \sum_{m=1}^M \int_{m-1}^m t \exp\left[-\int_0^t \lambda_h^*(\tau) d\tau\right] \lambda_h^*(t) dt \end{aligned}$$

Table VIII. Posterior effects of different compulsory attendance ages (16 and 18) on the probability of high school survival  $P(t_h = M|Data)$ , the high school duration  $E(t_h|Data)$ , the hazard of leaving unemployment in the first month of unemployment  $E(\lambda_{u1}^*|Data)$  and the unemployment duration  $E(t_u|Data)$  for students with different characteristics (most disadvantaged, disadvantaged, average, advantaged and most advantaged)

	Posterior functions of interest			
	$P(t_h = M Data)^a$	$E(t_h Data)^b$	$E(\lambda_{u1}^* Data)^c$	$E(t_u Data)^d$
Most disadvantaged				
Compulsory 16	0.255	16.4	0.214	5.21
Compulsory 18	0.355	18.8	0.221	5.06
Disadvantaged				
Compulsory 16	0.52	21.3	0.283	4
Compulsory 18	0.609	22.9	0.29	3.92
Average student				
Compulsory 16	0.734	24.5	0.376	3.01
Compulsory 18	0.791	25.4	0.381	2.97
Advantaged				
Compulsory 16	0.863	26.2	0.491	2.29
Compulsory 18	0.894	26.7	0.494	2.27
Most advantaged				
Compulsory 16	0.932	27.1	0.633	1.75
Compulsory 18	0.948	27.4	0.636	1.74

<sup>a</sup>  $P(t_h = M|D)$  denotes the predictive probability of high school survival.

<sup>b</sup>  $E(t_h|D)$  denotes the predictive mean of the high school duration.

<sup>c</sup>  $E(\lambda_{u1}^*|D)$  denotes the predictive mean of the hazard of leaving unemployment in the first month of unemployment.

<sup>d</sup>  $E(t_u|D)$  denotes the predictive mean of the unemployment duration.

$$\begin{aligned}
&= \exp\left(-\sum_{m=1}^M \lambda_{hm}^*\right) M + \sum_{m=1}^M \int_0^1 (t+m-1) \exp\left[-\int_0^{t+m-1} \lambda_h^*(\tau) d\tau\right] \lambda_{hm}^* dt \\
&= \exp\left(-\sum_{m=1}^M \lambda_{hm}^*\right) M + \sum_{m=1}^M \int_0^1 (t+m-1) \exp\left[-\lambda_{hm}^* t - \sum_{j=1}^{m-1} \lambda_{hj}^*\right] \lambda_{hm}^* dt \\
&= \exp\left(-\sum_{m=1}^M \lambda_{hm}^*\right) M + \sum_{m=1}^M \exp\left(-\sum_{j=1}^{m-1} \lambda_{hj}^*\right) [1 - \exp(-\lambda_{hm}^*)] \\
&\quad \times \left[m - 1 + \lambda_{hm}^{*-1} - \frac{\exp(-\lambda_{hm}^*)}{1 - \exp(-\lambda_{hm}^*)}\right]
\end{aligned}$$

3. The hazard of leaving unemployment in the first month of unemployment:

$$E(\lambda_{u1}^*) = E[\exp(x_u \beta_u) \lambda_{u1}]$$

where  $\lambda_{uw}^*$  indicates the hazard of leaving unemployment of a representative individual,  $x_u$  is the  $1 \times k_u$  vector of individual-level characteristics of this representative individual, and  $\lambda_{uw}$  denotes the baseline hazard of leaving unemployment.

4. The unemployment duration:

$$\begin{aligned}
E(t_u) &= \exp\left(-\sum_{w=1}^W \lambda_{uw}^*\right) W + \sum_{w=1}^W \exp\left(-\sum_{j=1}^{w-1} \lambda_{uj}^*\right) [1 - \exp(-\lambda_{uw}^*)] \\
&\quad \times \left[w - 1 + \lambda_{uw}^{*-1} - \frac{\exp(-\lambda_{uw}^*)}{1 - \exp(-\lambda_{uw}^*)}\right]
\end{aligned}$$

From Table VIII, the effects of compulsory attendance on the probability of high school survival, the high school duration, the hazard of leaving unemployment in the first month of unemployment and the unemployment duration are most pronounced for the most disadvantaged students.<sup>26</sup> For example, increasing the compulsory attendance age from 16 to 18 increases the probability of high school survival by 10 percentage points (from 25.5 to 35.5 percentage points, or 39%), lengthens the high school duration by 2.4 months (or 15%), increases the hazard of leaving unemployment in the first month of unemployment by 0.7 percentage points (or 3.3%), and decreases the

<sup>26</sup> The most disadvantaged students are defined to have a parental income of \$5000, a base year cognitive test score of negative two standard deviations, a father with an education of 8 years, a mother with an education of 8 years, two siblings, no post-secondary education, and to be male, white, residing in a county with an employment growth rate from 1980 to 1982 of zero percentage points, born on February 1, 1964, and working as a labourer in the service industry before unemployment.

Everything else being the same, disadvantaged students have a parental income of \$10000, a base year cognitive test score of negative one standard deviation, a father with an education of 10 years and a mother with an education of 10 years. Average students have a parental income of \$20000, a base year cognitive test score of zero, a father with an education of 12 years and a mother with an education of 12 years. Advantaged students have a parental income of \$30000, a base year cognitive test of one standard deviation, a father with an education of 14 years and a mother with an education of 14 years. The most advantaged students have a parental income of \$40000, a base year cognitive test of two standard deviations, a father with an education of 16 years and a mother with an education of 16 years.

unemployment duration by 0.15 months (or 2.9%).<sup>27</sup> The effects of compulsory attendance are also evident for disadvantaged students. For this group of students, the marginal effects on the probability of high school survival, the high school duration, the hazard of leaving unemployment in the first month of unemployment and the unemployment duration are 8.9 percentage points (or 17%), 1.6 months (or 7.5%), 0.7 percentage points (or 2.5%) and  $-0.08$  months (or  $-2\%$ ), respectively.

For average students, these marginal effects are relatively smaller: 5.7 percentage points (or 7.8%), 0.9 months (or 3.7%), 0.5 percentage points (or 1.3%) and  $-0.04$  months (or  $-1.3\%$ ), respectively. These marginal effects are even smaller for advantaged students: 3.1 percentage points (or 3.6%), 0.5 months (or 1.9%), 0.3 percentage points (or 0.61%) and  $-0.02$  months (or 0.87%), respectively, and for the most advantaged students: 1.6 percentage points (or 1.7%), 0.3 months (or 1.1%), 0.3 percentage points (or 0.47%) and  $-0.01$  months (or  $-0.57\%$ ), respectively. Again, my results suggest that the compulsory attendance has an impact not only on average schooling and employment, but also on the distribution of schooling and employment as a whole. Compulsory schooling laws lower the inequality in education and employment by increasing mainly the schooling and employment levels of those disadvantaged students.

#### 4. CONCLUSION

In this paper, I provide new evidence from High School and Beyond (HSB) on the effects of compulsory attendance on high school completion and future youth unemployment. I develop a Bayesian estimation approach to the simultaneous equation model with ordered probit and two-limit censored regression to study the joint outcomes of high school completion and unemployment, accounting for the heterogeneity in returns to education and nonlinearity in the effects of compulsory attendance. The simulation results suggest that for average students, increasing the compulsory attendance age from 16 to 18 increases the probability of completing high school by 5.1 percentage points (from 89 to 94.1 percentage points, or 5.7%) and reduces the proportion of time the individuals are unemployed by 0.22 percentage points (or 2.5%). These effects are much more pronounced for disadvantaged students, but less pronounced for advantaged students, suggesting the potential effects of compulsory attendance on reducing the inequality in education and employment.

It is possible that unemployment behaviour and high school dropout behaviour are characterized by strong state dependence, so that the probabilities of leaving unemployment and leaving high school today depend on the amount of time an individual has been unemployed and has been in school in the past. To this end, I also develop a Bayesian estimation approach to the bivariate duration model to study simultaneously the high school duration and future unemployment duration. The simulation results suggest that for average students, increasing the compulsory attendance age from 16 to 18 lengthens the high school duration by 0.9 months (from 24.5 to 25.4 months, or 3.7%) and increases the hazard of leaving unemployment in the first month of unemployment by 0.5 percentage points (or 1.3%). Again, these effects are much more pronounced for disadvantaged students, but less pronounced for advantaged students.

<sup>27</sup> Comparing the column indicating the probability of high school survival  $P(t_h = M|Data)$  in Table VIII with the column indicating the probability of high school completion  $P(y_h = 4|Data)$  in Table IV, it is evident that the sample of individuals who experienced some unemployment have relatively lower high school completion rates than those who experienced no unemployment.

## ACKNOWLEDGEMENTS

I would like to thank David Brownstone, Philip Oreopoulos, Dale Poirier and Justin Tobias for their helpful comments and suggestions. I am very grateful to three anonymous referees and the editor John Rust for their constructive and insightful comments which have significantly improved my paper. All errors are, of course, mine.

## REFERENCES

- Acemoglu D, Angrist J. 2000. How large are human capital externalities? Evidence from compulsory schooling laws. *NBER Macroannual*: 9–59.
- Albert JH, Chib S. 1993. Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association* **88**: 669–679.
- Angrist JD, Krueger AB. 1991. Does compulsory school attendance affect schooling and earnings? *Quarterly Journal of Economics* **106**(4): 979–1014.
- Belzil C, Hansen J. 2002. Unobserved ability and the return to schooling. *Econometrica* **70**(5): 2075–2091.
- Campolieti M. 1997. Bayesian estimation of duration models: an application of the multiperiod probit model. *Empirical Economics* **22**: 461–480.
- Campolieti M. 2000. Bayesian estimation and smoothing of the baseline hazard in discrete time duration models. *Review of Economics and Statistics* **82**(4): 685–701.
- Campolieti M. 2001. Bayesian semiparametric estimation of discrete duration models: an application of the Dirichlet process prior. *Journal of Applied Econometrics* **16**: 1–22.
- Chib S, Greenberg E. 1995. Understanding the Metropolis–Hastings algorithm. *The American Statistician* **49**(4): 327–335.
- Cowles MK. 1996. Accelerating Monte Carlo Markov chain convergence for cumulative-link generalized linear models. *Statistics and Computing* **6**: 101–111.
- Cox DR. 1972. Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, Series B* **34**: 17–30.
- Eckstein Z, Wolpin KI. 1999. Why youths drop out of high school: the impact of preferences, opportunities, and abilities. *Econometrica* **67**(6): 1295–1339.
- Gamerman D. 1997. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Chapman and Hall: London.
- Ganderton PT. 1992. The effect of subsidies in kind on the choice of college. *Journal of Public Economics* **48**: 269–292.
- Gelman A, Carlin J, Stern H, Rubin DB. 1995. *Bayesian Data Analysis*. Chapman and Hall: New York.
- Gilks WR, Richardson S, Spiegelhalter DJ. 1996. *Markov Chain Monte Carlo in Practice*. Chapman and Hall: London.
- Goldin C, Katz L. 2003. Mass secondary schooling and the state: the role of state compulsion in the high school movement. NBER Working Paper Series #10075.
- Gørgens T, Horowitz JL. 1999. Semiparametric estimation of a censored regression model with an unknown transformation of the dependent variable. *Journal of Econometrics* **90**: 155–191.
- 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.
- Han A, Hausman JA. 1990. Flexible parametric estimation of duration and competing risk models. *Journal of Applied Econometrics* **5**: 1–28.
- Hanushek EA, Taylor LL. 1990. Alternative assessments of the performance of schools: measurement of state variation in achievement. *Journal of Human Resources* **25**(2): 179–201.
- Harmon C, Walker I. 1995. Estimates of the economic return to schooling for the United Kingdom. *American Economic Review* **85**(5): 1278–1286.
- Heckman JJ, Singer B. 1984. A method of minimizing the impact of distributional assumption for duration data. *Econometrica* **52**: 271–320.

- Horowitz JL. 1999. Semiparametric estimation of a proportional hazard model with unobserved heterogeneity. *Econometrica* **67**(5): 1001–1028.
- Keane MP, Wolpin KI. 1994. The solution and estimation of discrete choice dynamic programming models by simulation and interpolation: Monte Carlo evidence. *Review of Economics and Statistics* **76**(4): 648–672.
- Keane MP, Wolpin KI. 1997. The career decisions of young men. *Journal of Political Economy* **105**(3): 437–522.
- Kiefer NM. 1988. Economic duration data and hazard functions. *Journal of Economic Literature* **26**: 646–679.
- Koop G, Tobias JL. 2004. Learning about heterogeneity in returns to schooling. *Journal of Applied Econometrics* **19**(7): 827–849.
- Lancaster T. 1979. Econometric methods for the duration of unemployment. *Econometrica* **47**(4): 939–956.
- Lancaster T, Nickell S. 1980. The analysis of re-employment probabilities for the unemployed. *Journal of the Royal Statistical Society, Series A* **143**(2): 141–165.
- Lang K, Kropp D. 1986. Human capital versus sorting: the effects of compulsory attendance laws. *Quarterly Journal of Economics* **101**(3): 609–624.
- Li K. 1998. Bayesian inference in a simultaneous equation model with limited dependent variables. *Journal of Econometrics* **85**: 387–400.
- Lillard LA. 1993. Simultaneous equations for hazards. *Journal of Econometrics* **56**: 189–217.
- Lleras-Muney A. 2002. Were compulsory attendance and child labor laws effective? An analysis from 1915 to 1939. *Journal of Law and Economics* **45**: 401–435.
- Lochner L, Moretti E. 2004. The effect of education on crime: evidence from prison inmates, arrests, and self-reports. *American Economic Review* **94**(1).
- Meyer BD. 1990. Unemployment insurance and unemployment spells. *Econometrica* **58**(4): 757–782.
- Nandram B, Chen M-H. 1996. Reparameterizing the generalized linear model to accelerate Gibbs sampler convergence. *Journal of Statistical Computation and Simulation* **54**: 129–144.
- Nickell S. 1979. Estimating the probability of leaving unemployment. *Econometrica* **47**(5): 1249–1266.
- Nobile A. 2000. Comment: Bayesian multinomial probit models with a normalization constraint. *Journal of Econometrics* **99**: 335–345.
- Oreopoulos P. 2003. Do dropouts drop out too soon? International evidence from changes in school leaving laws. NBER Working Paper Series #10155.
- Poirier D. 1995. *Intermediate Statistics and Econometrics*. MIT Press: Cambridge, MA.
- Rivkin SG. 1991. Schooling and employment in the 1980's: who succeeds? PhD dissertation, UCLA Department of Economics.
- Rosenzweig MR, Wolpin KI. 2000. Natural 'natural experiments' in economics. *Journal of Economic Literature* **38**(4): 827–874.
- Ruggiero M. 1994. Bayesian semiparametric estimation of proportional hazards models. *Journal of Econometrics* **62**: 277–300.
- Rust J. 1987. Optimal replacement of GMC bus engines: an empirical model of Harold Zurcher. *Econometrica* **55**(5): 999–1033.
- Stern D, Paik I-W, Catterall JS, Nakata Y-F. 1989. Labor market experience of teenagers with and without high school diplomas. *Economics of Education Review* **8**(3): 233–246.
- Tobias JL, Li M. 2003. A finite-sample hierarchical analysis of wage variation across public high school: evidence from the NLSY and High School and Beyond. *Journal of Applied Econometrics* **18**: 315–336.
- Van den Berg GJ. 2001. Duration models: specification, identification and multiple durations. In *Handbook of Econometrics*, Vol. 5, Heckman JJ, Leamer E (eds). North-Holland: Amsterdam.