Econometric Reviews, 26(5):529–556, 2007

ISSN: 0747-4938 print/1532-4168 online
DOI: 10.1080/07474930701509416

BAYESIAN PROPORTIONAL HAZARD ANALYSIS OF THE TIMING OF HIGH SCHOOL DROPOUT DECISIONS

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

□ *In this paper, I study the timing of high school dropout decisions using data from High School and Beyond. I propose a Bayesian proportional hazard analysis framework that takes into account the specification of piecewise constant baseline hazard, the time-varying covariate of dropout eligibility, and individual, school, and state level random effects in the dropout hazard. I find that students who have reached their state compulsory school attendance ages are more likely to drop out of high school than those who have not reached compulsory school attendance ages. Regarding the school quality effects, a student is more likely to drop out of high school if the school she attends is associated with a higher pupil–teacher ratio or lower district expenditure per pupil. An interesting finding of the paper that comes along with the empirical results is that failure to account for the* time-varying *heterogeneity in the hazard, in this application, results in upward biases in the duration dependence estimates. Moreover, these upward biases are comparable in magnitude to the well-known downward biases in the duration dependence estimates when the modeling of the* time-invariant *heterogeneity in the hazard is absent.*

Keywords Bayesian analysis; High school dropout behavior; Proportional hazard analysis.

JEL Classification C11; C41; I20.

1. INTRODUCTION

There has been a large literature that documents the negative correlations between the act of dropping out of high school and various outcomes of the individuals. For example, compared with high school graduates, high school dropouts are likely to suffer from lower wages and higher likelihood of unemployment when they enter the labor market (Blakemore and Low, 1984; Li, 2006; Stern et al., 1989). In terms of

Received June 12, 2005; Accepted May 18, 2006
Address correspondence to Mingliang Li, Department of Economics, 415 Fronczak Hall, State University of New York at Buffalo, Buffalo, New York, NY 14260, USA; E-mail: mli3@buffalo.edu

academic performance, high school dropouts on average have lower subject test scores (Li et al., 2004). High school dropouts are also more likely to be welfare recipients (McCaul et al., 1992), and as welfare recipients they tend to stay in welfare programs longer than others (Bane and Ellwood, 1983). In addition, evidence suggests that high school dropouts have a higher chance of engaging in illegal drug activities and marijuana initiation (Bray et al., 2000; Hsing, 1995), and female dropouts are more likely to be related to teenage pregnancy and single motherhood (Garfinkel and McLanahan, 1986; Olsen and Farkas, 1989).

Given the negative links between high school dropout behavior and individuals' outcomes, it is of substantial policy interest to study the timing of high school dropout decisions and various factors that can influence the decisions of dropping out of high school. For instance, finding out the times at which students are at high risk of dropping out of school helps the education policy makers to understand dropout behavior and design timely intervention strategy to prevent students from dropping out (DesJardins et al., 1999; Olsen and Farkas, 1989). Some researchers find that a substantial proportion of the potential dropouts remain in high school because of compulsory schooling laws (Angrist and Krueger, 1991). If compulsory schooling laws are indeed highly effective in preventing students from dropping out, enforcement of such laws will be desirable. Finally, a key question often addressed in the schooling literature is whether better school quality, as commonly measured by lower pupil–teacher ratio, higher district expenditure per pupil, etc. can enhance various student outcomes. So far, no consensus has been reached in the literature on the effectiveness of different school quality variables. In particular, little work has been done to examine the potential impacts of school quality variables on students' dropout behavior.

This paper contributes to the existing literature and casts light on the above important questions by using data from High School and Beyond (HSB) and adopting a Bayesian proportional hazard analysis approach.[1] In particular, I employ a piecewise constant baseline hazard

[1]As mentioned, there are several related papers written by this author that belong to the large literature on the dropout behavior of high school students. For example, Li et al. (2004) focus on the impacts of high school dropout decisions on students' subject test performance. Li (2006) studies the effects of students' dropout behavior on their unemployment outcomes in the future. Compared with those papers, this paper focuses mainly on the process of dropping out of high school itself. In particular, this paper attempts to address whether school resources are effective in keeping students in high school and the effectiveness of various determinants of the high school dropout decisions. In terms of methodology, the other papers employ a simultaneous equation framework where students' dropout behavior serves mainly as an endogenous covariate. In contrast, the timing of high school dropout decisions is the key outcome variable in this paper. To capture the unobserved heterogeneity in the dropout hazard at the individual, school and state levels, I also incorporate into this model random effects at various levels. Another interesting finding that comes along with the empirical results is the potential upward biases in the duration dependence estimates when the modeling of time-varying heterogeneity in the hazard is absent.

specification (Holford, 1976) to permit a flexible estimation of the duration dependence pattern of the dropout process and to address the concern that, during a particular period of time, students may have a relatively higher or lower chance of dropping out of high school. To take into account that compulsory schooling laws may prevent students from dropping out, I include in the model a time-varying covariate indicating each student's eligibility of dropping out of high school under the state compulsory schooling laws. I incorporate into the model this time-varying covariate and other individual and school level covariates by utilizing a proportional hazard analysis framework (Cox, 1972; Lancaster, 1979). The hazard of dropping out of high school may vary from one individual to another, from individuals in one school to those in another, and from schools in one state to those in another. To take into consideration both observed and unobserved heterogeneity in the dropout hazard at the individual, school, and state levels, I also model explicitly the random effects in the hazard at various levels and add these features to the proportional hazard framework (Bolstad and Manda, 2001; Guo and Rodriguez, 1992; Lancaster, 1979; Sastry, 1997).

The above features of the proportional hazard analysis, like the modeling of piecewise constant baseline hazard, the time-varying covariate, and the random effects in the hazard, are widely present in the empirical works of hazard analysis, especially among those analyzed in a non-Bayesian framework. In this paper, I propose a Bayesian analysis framework that incorporates all the above features. There are several advantages associated with the Bayesian approach I propose. First, the Bayesian estimation tools, such as the Metropolis–Hastings algorithm and the Gibbs sampler, are adept at dealing with multiple integrals in the likelihood function and random effects at many levels of the model. Bayesian estimation methods also avoid any asymptotic theory or approximation to the true posterior distribution and provide exact finite sample estimates of the model parameters. Moreover, since parameter uncertainty is formally introduced into a Bayesian analysis framework and parameters can be easily integrated out, a rich list of predictive functions of interest, including the marginal effect of each covariate and the duration dependence pattern, can be easily estimated to gain insights into the model.

An interesting finding that comes along with the empirical results of the model is the potential upward biases in the duration dependence estimates when the modeling of heterogeneity in the hazard is absent. Initially, this seems to contradict the well-established claim that failure to incorporate the observed and unobserved heterogeneity in the hazard results in downward biases in the duration dependence estimates (Heckman, 1991; Heckman and Singer, 1984a,b). However, a deeper look into the problem reveals a striking difference between the consequence of omitting the time-varying heterogeneity from the model and that of omitting the time-invariant heterogeneity. Specifically, the absence of the *time-invariant* heterogeneity from the model leads to the well-known and commonly observed downward biases in the duration dependence estimates. However, the omission of the *time-varying* heterogeneity from the model tells a different story. In this paper, it turns out that the failure to incorporate the time-varying heterogeneity in the hazard, i.e., the time-varying covariate of dropout eligibility, results in upward biases in the duration dependence estimates.

I analyze the timing of high school dropout decisions using data from High School and Beyond (HSB). HSB is a rich data set that offers an abundant array of individual and school level variables that have potential impacts on students' dropout behavior. I find that the hazard of dropping out of high school increases over time before students enter their senior year, and the hazard decreases over time afterwards. In addition, students who are eligible to drop out of high school under the compulsory schooling laws are 114 percent more likely to drop out than those who are not eligible. Regarding the school quality effects, a one unit increase in the pupil–teacher ratio increases the dropout hazard of a student by 1.3 percent. A one thousand dollars' increase in the district expenditure per pupil reduces the hazard by 8.1 percent. Finally, substantial heterogeneity in the dropout hazard exists at the student, school, and state levels. The remaining part of the paper is organized as follows. Section 2 gives a brief description of the HSB data used in the paper. Section 3 presents the proportional hazard analysis framework and discusses the Bayesian estimation methods. Section 4 discusses the empirical findings of the paper and some interesting results regarding the duration dependence estimates. Section 5 concludes.

2. DATA

In this paper, I use data from High School and Beyond (HSB), a national survey of US high school students that was conducted by the National Center for Education Statistics (NCES). In the spring of 1980, NCES administered the base year survey and interviewed up to 36 sophomores from each of about 1,000 high schools in the United States. A follow-up survey was conducted in the spring of 1982, which collected information on the high school dropout behavior of the 1980 sophomores. In particular, the survey investigators consulted school administrative records to determine each 1980 sophomore's dropout status at the time of the follow-up survey. In addition, the high school dropouts of the 1980 sophomores reported the year and month in which they left high school. Because of the nature of the survey that is mentioned above, it is natural to define that the high school duration is from the beginning of the base year survey (February 1980).[2] Likewise, the last month included in the sampling period is the end of the follow-up survey (June 1982).[3] This month also corresponds to the time point when the nondropouts of the 1980 sophomores were expected to graduate from high school. Hence it is reasonable to say that the high school duration is censored at this time point. Finally, the nature of the data suggests that it is reasonable to specify a month specific dropout hazard and assume that dropouts decide to drop out of high school in the middle of a month.

The HSB survey offers an abundant array of individual and family background variables that have influence on students' dropout behavior. The list of individual level variables I include in the model specification are gender, race, urbanicity, parental education, family income, and number of siblings in the household. I also include the base year cognitive test score, which can be used as a measure of the cognitive ability of a student. An attractive feature of the HSB survey is the availability of a rich set of variables characterizing the school a student attends. These variables provide an opportunity to study the effects of school quality variables on students' dropout behavior. The school level covariates I include in the model are pupil–teacher ratio, books per pupil, percentage of teachers with an M.A. or Ph.D. degree, percentage of teachers staying at school 10 years or more, and district expenditure per pupil. To control for the community characteristics associated with a given school, I also include in the model the average family income of the students in a school, the average base year cognitive test score of the students in a school, and the county level employment growth rate between 1980 and 1982.

One particular variable I include in the model that is worth emphasizing is the student's eligibility of dropping out of high school under the state compulsory schooling laws. First, this variable is of interest and relevance to the education policy makers and legislators. Indeed, it

[2]There are several different ways of defining high school duration. For example, one may define it to be from the time when a student enters high school, or the time when a student reaches a particular age, or a particular calendar date. In this paper, the duration is defined to be from February 1980. This month corresponds to the beginning of the base year survey. Since the 1980 sophomores included in the HSB survey were current high school sophomores in February 1980, no dropout behavior was observed for this group of students before this month. In addition, as far as the group of 1980 sophomores is concerned, February 1980 also corresponds to the middle point of the sophomore year. Taking into consideration all the above, it seems most natural to choose this month as the origin of the time scale.

[3]One of the main objectives of the HSB survey is to provide researchers with detailed information on the dropout behavior of the 1980 sophomores and an opportunity to study the various determinants of the decisions of dropping out of high school. To this end, the survey investigators made their best endeavors to retain both nondropouts and dropouts of the 1980 sophomores in the follow-up survey. As a result, although many students dropped out of school before June 1982, they did not drop out of the survey. The sample included in the follow-up survey is still representative of the 1980 sophomores, although the follow-up survey was conducted about two years after the base year survey.

is of substantial policy interest to study whether compulsory schooling laws are effective in keeping students from dropping out of high school. Importantly, there is considerable variation in the compulsory school attendance ages stipulated in the state compulsory schooling laws across the nation. Around 1980, the compulsory school attendance age of many states was 16 (Angrist and Krueger, 1991). This means that, in these states, students were not allowed to drop out of school until they reached the age of 16. However, there also existed many exceptions. Specifically, according to the Digest of Education Statistics published by the NCES, around 1980, there were one, thirty-seven, six, and another six states adopting the compulsory school attendance ages of 15, 16, 17, and 18, respectively. Moreover, in addition to the variation in the state compulsory school attendance ages across the nation, each student's eligibility of dropping out of school changes over time. As mentioned before, in most states where the compulsory school attendance age was 16, students younger than 16 were not allowed to drop out but those who were older than 16 were eligible. Therefore the variation in the dates of birth of the 1980 sophomores introduces an additional source of variation in students' dropout eligibility at a given time point, even among those who are from the states with the same compulsory school attendance age. In short, the dropout eligibility is a time-varying covariate that plays a role different from the other covariates.

One reason that so far, little work has been done to examine the impacts of compulsory schooling laws on students' dropout behavior using the HSB data is that the NCES does not publicize the information regarding the geographical locations of the 1980 sophomores included in the HSB survey. In this paper, I identify the state of each high school included in the HSB survey by following previous attempts of Hanushek and Taylor (1990), Rivkin (1991), Ganderton (1992), and Grogger (1996a,b). In particular, I make use of the local labor market conditions associated with each school contained in the HSB survey between 1980 and 1982. This data set was published by the NCES but it does not identify the location of any school or student in the HSB survey. Instead, the data set provides state level demographic information associated with each school in the survey. I match this data set to the publicly available demographic information and hence identify the state where each school was located. After I identify the state associated with each school and each student included in the survey, I combine this information with the date of birth of each student and construct the time-varying covariate indicating the dropout eligibility of a student under the state compulsory schooling laws.

A common feature of the survey data, which is also present in the data set used in this paper, is the prevalence of the problem of missing data. For example, for this data set, approximately 18%, 17%, 14%, 24%, 16%, 7%, 11%, 2%, 3%, and 25% of the observations are missing owing to missing information on base year cognitive test score, father's education, mother's education, family income, number of siblings, pupil–teacher ratio, books per pupil, percentage of teachers with an M.A. or Ph.D. degree, percentage of teachers staying at the school 10 years or more, and district expenditure per pupil, respectively. The exclusion of these observations from the sample will result in a substantial loss of the survey data. To minimize the impact of the missing data problem, I choose to create dummy variables indicating the missing status of each of the above covariates for all individuals in the sample and replace the missing observations associated with a particular covariate with the average of the nonmissing observations

TABLE 1 Descriptive statistics of the data

Variables[a]	Sample mean	Standard error
Dropout eligibility	0.742	0.439
Base year cognitive test score	0	1
Female	0.5	0.5
Minority	0.322	0.467
Father's education	12.4	3.64
Mother's education	12.1	3.14
Family income ($10,000)	2.09	0.989
Number of siblings	2.9	1.59
Urbanicity[b]		
Suburban	0.501	0.5
Rural	0.287	0.452
Pupil–teacher ratio	19.3	4.9
Books per pupil	15.3	13
% of teachers with M.A. or Ph.D. degree	48.2	23.3
% of teachers at school 10 years or more	39.4	23.5
District expenditure per pupil ($1,000)	1.92	0.711
Average family income ($10,000)[c]	2.09	0.533
Average test score[d]	0	1
County level employment growth rate 80–82 (%)	−0.0553	5.67
High school duration (month)[e]	27	4.11

[a]To reduce the number of missing observations, I create a set of dummy variables indicating whether an observation is missing for each of the following covariates: base year cognitive test score, father's education, mother's education, family income, number of siblings, pupil–teacher ratio, books per pupil, percentage of teachers with an M.A. or Ph.D. degree, percentage of teachers staying at school 10 years or more, and district expenditure per pupil. I do not present the descriptive statistics of these dummy variables to save space.

[b]For urbanicity, the excluded are individuals from the urban areas.

[c]Average family income is the average family income of the students within a school.

[d]Average cognitive test score is the average of base year cognitive test scores of the students within a school, standardized to have a mean of zero and a variance of one.

[e]High school duration is defined to be from February 1980.

of that covariate.[4] This helps to retain a significant number of observations in the sample, and the final sample contains 25,404 students from 930 schools in 50 states. In Table 1, I present the descriptive statistics of the data used in the paper.

3. MODEL

It is possible that high school dropout behavior, like other types of economic and social behavior, is characterized by strong state dependence, so that the chance that a student leaves high school today depends on the length of time she has been in school in the past. There exists a large literature that exemplifies the applications of hazard analysis to various interesting problems in economics and social sciences (Campolieti, 1997, 2000, 2001; Han and Hausman, 1990; Lancaster, 1979, 1990; Lancaster and Nickell, 1980; Meyer, 1990; Nickell, 1979). Indeed, it is of great importance to study the conditional probabilities a student drops out of high school at different points of the high school duration and to find out the sources of variation between different individuals in the amount of time they remain in school. In addition, the timing of dropout decisions can be considered as a sequence of binary outcomes that record the dropout status of a student in each time period. Therefore the high school duration should contain richer information than the single binary outcome variable indicating the student's dropout status over the entire period of time.

Formally, let us define the survivor function as the probability that the duration T is equal to or bigger than a given value t, $S(t) = \text{Prob}(T \geq t)$.[5] Therefore the cumulative distribution function is $F(t) = \text{Prob}(T < t) = 1 - \text{Prob}(T \geq t) = 1 - S(t)$, and the probability density function is $f(t) = dF(t)/dt = d[1 - S(t)]/dt = -dS(t)/dt$. A key function of interest in the duration analysis literature is the hazard function (Cox, 1972; Kiefer, 1988). It roughly captures the instantaneous probability that the duration T is exactly of the value t, conditional on T being equal to or greater than t, $\lambda(t) = f(t)/S(t) = [-dS(t)/dt]/S(t) = -d\ln S(t)/dt$. Finally, the integrated hazard function is $\Lambda(t) = \int_0^t \lambda(u)du = \int_0^t [-d\ln S(u)/du]du = -\int_0^{ln S(t)} d\ln S(u) = -\ln S(t)$. This suggests that both $S(t)$ and $f(t)$ can

[4]An alternative approach to constructing the sample is to discard the observations with missing data altogether, which leads to a significant reduction in the sample size. This strategy was adopted in the last version of the paper. In this version, I take the suggestions from a referee to create dummy variables for missing data and substitute the means of the nonmissing observations for those missing observations. The results of the paper are robust to this alternative strategy of constructing the sample. The main results did not change after I chose to retain the observations with missing data in the sample.

[5]In this paper, I choose to analyze high school duration in a continuous time framework instead of using a discrete time model.

be expressed in terms of $\lambda(t)$, $S(t) = \exp[-\Lambda(t)] = \exp[-\int_0^t \lambda(u)du]$, and $f(t) = S(t)\lambda(t) = \exp[-\int_0^t \lambda(u)du]\lambda(t)$.

As discussed in the data section, the dropouts of the 1980 sophomores only reported the year and month in which they left high school. Therefore it is natural to assume a month-specific baseline dropout hazard λ_m, where m denotes month m. This approach is consistent with the commonly adopted specification of piecewise constant baseline hazard (Campolieti, 1997, 2001; Han and Hausman, 1990; Holford, 1976; Meyer, 1990).[6] Following the proportional hazard analysis approach (Cox, 1972; Kiefer, 1988), I incorporate into the model the individual, school, and state level covariates and random effects and a time-varying covariate indicating each student's eligibility of dropping out under the schooling laws.

To be more specific, define the dropout hazard of individual i from school c in state s and in month m as $\lambda_{scim} = \exp(x_{scim}\beta)\eta_s\xi_{sc}\alpha_{sci}\lambda_m$, where x_{scim} is a $1 \times k$ vector of covariates associated with the individual in month m, and η_s, ξ_{sc}, and α_{sci} denote the unobserved heterogeneity in the dropout hazard at the state, school, and individual levels, respectively. The set of observed variables x_{scim} includes a time-varying covariate that indicates each student's dropout eligibility over time, a group of individual specific variables (gender, race, urbanicity, base year cognitive test score, parental education, family income, and number of siblings), and an array of school level covariates (pupil–teacher ratio, books per pupil, percentage of teachers with an M.A. or Ph.D. degree, percentage of teachers staying at school 10 years or more, district expenditure per pupil, average family income of the students in a school, average base year cognitive test score of the students in a school, and county level employment growth rate between 1980 and 1982).

In addition, I incorporate into the model the individual, school, and state level random effects in the dropout hazard. It is a well-known result in the duration analysis literature that failure to account for observed and unobserved heterogeneity in the hazard results in downward biases in the duration dependence estimates. Although I include in the model a wide array of individual and school level covariates, it is not likely that the heterogeneity in the dropout hazard can be fully captured by these observed covariates. Therefore it is important to take into account both the observed and the unobserved heterogeneity in the hazard at various levels. Moreover, it is possible that a student drops out of school partly because his or her friends do so. In other words, the dropout behavior of

[6]It is possible to model the dropout hazard using full nonparametric specifications (Clayton, 1991; Campolieti, 2000, 2003; Gørgens and Horowitz, 1999; Horowitz, 1999; Kalbfleisch, 1978; Ruggiero, 1994), which are computationally more demanding. In this paper, I adopt the specification of month-specific dropout hazard, which has relatively lower computational costs and reasonable flexibility in modeling.

the students within the same school may not be independent, or there may exist some interaction effects between students from the same school. The set of school level random effects considered in this model is an attempt to capture such interaction or peer group effects. Since students from the same school share a common school level random effect in the hazard, the dropout hazards of these students are not independent from each other.[7]

In practice, I follow many previous studies of random effects in the hazard analysis literature (Abbring and Van den Berg, 2007; Bolstad and Manda, 2001; Clayton, 1978; Guo and Rodriguez, 1992; Han and Hausman, 1990; Lancaster, 1979; Meyer, 1990; Oakes, 1982; Sastry, 1997; Vaupel et al., 1979; Van den Berg, 2001) and assume that the individual, school, and state level random effects follow gamma distributions, $\alpha_{sci} \overset{iid}{\sim} G(r, r^{-1}), \xi_{sc} \overset{iid}{\sim} G(u, u^{-1}), \eta_s \overset{iid}{\sim} G(v, v^{-1})$, where $G(\alpha, \beta)$ denotes the gamma distribution[8] with a mean of $\alpha\beta$ and a variance of $\alpha\beta^2$.[9] These assumptions reflect the standard normalizations adopted in the literature, i.e., the random effects at various levels of the model have a mean of one, $E(\alpha_{sci}) = E(\xi_{sc}) = E(\eta_s) = 1$. In addition, the parameters r^{-1}, u^{-1}, and v^{-1} have a ready interpretation in the model as they are the variance parameters of the individual, school, and state level random effects, $Var(\alpha_{sci}) = r^{-1}$, $Var(\xi_{sc}) = u^{-1}$, and $Var(\eta_s) = v^{-1}$, respectively.

To derive the likelihood function of the model, let t_{scim} denote the high school duration associated with individual i from school c in state s and in month m, and let d_{scim} denote the binary outcome variable indicating whether the individual drops out of high school in month m. Following the previous arguments, the survivor function of the individual in month m is $S_{scim}(t) = \exp[-\int_0^t \lambda_{scim}(u)du] = \exp[-\int_0^t \exp(x_{scim}\beta)\eta_s\xi_{sc}\alpha_{sci}\lambda_m du] = \exp[-\exp(x_{scim}\beta)\eta_s\xi_{sc}\alpha_{sci}\lambda_m t]$, and the related probability density function is $f_{scim}(t) = S_{scim}(t)\lambda_{scim}(t) = \exp[-\exp(x_{scim}\beta)\eta_s\xi_{sc}\alpha_{sci}\lambda_m t]\exp(x_{scim}\beta)\eta_s\xi_{sc}\alpha_{sci}\lambda_m$. For an individual who does not drop out in month m (i.e., $d_{scim} = 0$), the contribution of this observation to the likelihood function is captured by the survivor function $S_{scim}(t)$ evaluated at $t = t_{scim}$. In contrast, for an individual who drops out in month m (i.e., $d_{scim} = 1$), the contribution of

[7]In this paper, I do not model in an explicit way the interaction or peer effects. For example, one way to explain the dropout outcome of a student is to include the dropout outcomes of her peer students as endogenous covariates. I do not pursue this approach but instead attempt to capture the peer group effects by incorporating the school level random effects in the dropout hazard.

[8]See, e.g., Poirier (1995, p. 98).

[9]Several researchers in the literature (Campolieti, 2001, 2003; Elbers and Ridder, 1982; Gørgens and Horowitz, 1999; Horowitz, 1999; Heckman and Singer, 1984a,b; Lancaster and Nickell, 1980) have studied the nonparametric identification of random effects in the proportional hazard analysis framework and proposed some nonparametric estimation methods. However, to most empirical researchers and from a computational point of view, it is not a trivial task to take into account simultaneously the piecewise constant baseline hazard and the Heckman and Singer type of unobserved heterogeneity distribution. Hence I choose to follow the more common approach and assume that the random effects in the hazard follow gamma distributions.

the observation to the likelihood function is captured by the probability density function $f_{scim}(t)$ evaluated at $t = t_{scim}$.

Following these arguments, the likelihood function of the model is the enumeration of the individual likelihood function associated with each student in each month. In addition, the likelihood function explicitly incorporates the previously discussed individual, school, and state level random effects in the hazard:

$$p(Data \mid \Xi) \propto \left\{ \prod_{s=1}^{S} \prod_{c=1}^{C_s} \prod_{i=1}^{I_{sc}} \exp\left[-\sum_{m=1}^{M} \exp(x_{scim}\beta)\eta_s \xi_{sc} \alpha_{sci} \lambda_m t_{scim} \right] \right.$$
$$\left. \times \prod_{m=1}^{M} [\exp(x_{scim}\beta)\eta_s \xi_{sc} \alpha_{sci} \lambda_m]^{d_{scim}} \right\}$$
$$\times \left[\prod_{s=1}^{S} v^v \Gamma(v)^{-1} \eta_s^{v-1} \exp(-\eta_s v) \prod_{c=1}^{C_s} u^u \Gamma(u)^{-1} \xi_{sc}^{u-1} \exp(-\xi_{sc} u) \right.$$
$$\left. \times \prod_{i=1}^{I_{sc}} r^r \Gamma(r)^{-1} \alpha_{sci}^{r-1} \exp(-\alpha_{sci} r) \right],$$

where Ξ denotes the set of parameters, S the number of states, C_s the number of schools in state s, I_{sc} the number of students from school c in state s, and M the number of months in the sampling period.[10]

To estimate the model, I develop a Bayesian estimation method, which avoids a direct evaluation of the above nontrivial likelihood function and draws instead from the exact posterior of the model (Chib and Greenberg, 1995; Gamerman, 1997; Gelman et al., 1995; Gilks et al., 1996). For example, the Bayesian estimation method allows me to simulate the random effects in the hazard at various levels directly from the posterior of the model, so the estimation does not rely on any approximation methods or asymptotic theory. A wide array of posterior functions of interest, such as the duration dependence pattern, the marginal effect of each covariate on the dropout hazard, and the school level random effects in the dropout hazard, can be easily estimated.

Following the Bayes rule, the joint posterior distribution of the set of parameters $p(\Xi \mid Data)$ is proportional to the product of the above likelihood function of the model $p(Data \mid \Xi)$ and the joint prior distribution of the set of parameters $p(\Xi)$, $p(\Xi \mid Data) \propto p(Data \mid \Xi)p(\Xi)$.

[10] As previously discussed in the data section, the first month of the sampling period is February 1980. This month corresponds to the beginning of the base year survey. Likewise, the last month included in the sampling period is June 1982. This month corresponds to the end of the follow-up survey and is also the point when the nondropouts of the 1980 sophomores were expected to graduate from high school. Therefore the number of months included in the sampling period is $M = 28$.

I specify the following priors to complete the Bayesian analysis. First, I assume prior independence between different groups of parameters, $p(\Xi) = [\prod_{m=1}^{M} p(\lambda_m)]p(\beta)p(r)p(u)p(v)$. The priors for the baseline hazard parameters are $\lambda_m \sim G(a_\lambda, b_\lambda)$, for $m = 1, 2, 3, \ldots, M$, where $a_\lambda = 0.01$ and $b_\lambda = 100$. For the coefficient parameters, I assume that $\beta \sim N(\beta_0, V_\beta)$, where $\beta_0 = 0_{k\times 1}$ and $V_\beta = 1000I_k$. Finally, the priors for the individual, school, and state level variance parameters are specified as: $r \sim G(a_r, b_r)$, $u \sim G(a_u, b_u)$, and $v \sim G(a_v, b_v)$, where $a_r = a_u = a_v = 0.01$, $b_r = b_u = b_v = 100$.[11]

3.1. Gibbs Sampler

In practice, I use the following Metropolis–Hastings within Gibbs algorithm to implement the above Bayesian estimation method. The algorithm involves an iterative sampling from the following complete conditional posterior distributions that are derived from the previously discussed joint posterior distribution $p(\Xi \mid Data)$.

1. Sample the month specific baseline dropout hazard λ_m:

$$\lambda_m \mid \Xi_{-\lambda_m}, D \sim G\left\{a_\lambda + \sum_{s=1}^{S}\sum_{c=1}^{C_s}\sum_{i=1}^{I_{sc}} d_{scim}, \left[b_\lambda^{-1} + \sum_{s=1}^{S}\sum_{c=1}^{C_s}\sum_{i=1}^{I_{sc}} \exp(x_{scim}\beta)\eta_s\xi_{sc}\alpha_{sci}t_{scim}\right]^{-1}\right\}$$

for $m = 1, 2, \ldots, M$.

2. Sample the state level random effects in the dropout hazard η_s:

$$\eta_s \mid \Xi_{-\eta_s}, D \sim G\left\{v + \sum_{c=1}^{C_s}\sum_{i=1}^{I_{sc}}\sum_{m=1}^{M} d_{scim}, \left[v + \sum_{c=1}^{C_s}\sum_{i=1}^{I_{sc}}\sum_{m=1}^{M} \exp(x_{scim}\beta)\xi_{sc}\alpha_{sci}\lambda_m t_{scim}\right]^{-1}\right\}$$

for $s = 1, 2, \ldots, S$.

3. Sample the school level random effects in the dropout hazard ξ_{sc}:

$$\xi_{sc} \mid \Xi_{-\xi_{sc}}, D \sim G\left\{u + \sum_{i=1}^{I_{sc}}\sum_{m=1}^{M} d_{scim}, \left[u + \sum_{i=1}^{I_{sc}}\sum_{m=1}^{M} \exp(x_{scim}\beta)\eta_s\alpha_{sci}\lambda_m t_{scim}\right]^{-1}\right\}$$

for $c = 1, 2, \ldots, C_s$, and $s = 1, 2, \ldots, S$.

[11]These prior specifications are essentially noninformative and very diffuse. For example, I have tried the alternative prior hyperparameters $a_\lambda = a_r = a_u = a_v = 0.001$, or $a_\lambda = a_r = a_u = a_v = 0.1$; $b_\lambda = b_r = b_u = b_v = 1000$, or $b_\lambda = b_r = b_u = b_v = 10$; $V_\beta = 100I_k$ or $V_\beta = 10000I_k$ and found that estimation results virtually do not change.

4. Sample the individual level random effects in the dropout hazard α_{sci}:

$$\alpha_{sci} \mid \Xi_{-\alpha_{sci}}, D \sim G\left\{ r + \sum_{m=1}^{M} d_{scim}, \left[r + \sum_{m=1}^{M} \exp(x_{scim}\beta)\eta_s \xi_{sc} \lambda_m t_{scim} \right]^{-1} \right\}$$

for $i = 1, 2, \ldots, I_{sc}$, $c = 1, 2, \ldots, C_s$, and $s = 1, 2, \ldots, S$.

5. Sample the state level variance parameter v^{-1}:

$$p(v \mid D) \propto b_v^{-a_v} \Gamma(a_v)^{-1} v^{a_v - 1} \exp(-v b_v^{-1}) \prod_{s=1}^{S} v^v \Gamma(v)^{-1} \eta_s^{v-1} \exp(-\eta_s v).$$

Since this conditional posterior distribution cannot be sampled directly, I sample a candidate draw from the proposal density

$$v^* \mid \Xi_{-v}, D \sim G[a_v + \theta_v, (b_v^{-1} + \theta_v s_\eta^2)^{-1}],$$

where * denotes the candidate draw, $\theta_v = 10$ is the tuning parameter, $s_\eta^2 = \frac{1}{S} \sum_{s=1}^{S} (\eta_s - \bar{\eta})^2$, and $\bar{\eta} = \frac{1}{S} \sum_{s=1}^{S} \eta_s$. The probability of accepting the candidate draw is $\min(R, 1)$, where

$$R = \frac{v^{*-\theta_v} \exp(\theta_v s_\eta^2 v^*)}{v_{j-1}^{-\theta_v} \exp(\theta_v s_\eta^2 v_{j-1})} \prod_{s=1}^{S} \frac{v^{*v^*} \Gamma(v^*)^{-1} \eta_s^{v^*-1} \exp(-\eta_s v^*)}{v_{j-1}{}^{v_{j-1}} \Gamma(v_{j-1})^{-1} \eta_s^{v_{j-1}-1} \exp(-\eta_s v_{j-1})},$$

and $_{j-1}$ denotes the draw accepted in the last iteration.

6. Sample the school level variance parameter u^{-1}:

$$p(u \mid D) \propto b_u^{-a_u} \Gamma(a_u)^{-1} u^{a_u - 1} \exp(-u b_u^{-1}) \prod_{s=1}^{S} \prod_{c=1}^{C_s} u^u \Gamma(u)^{-1} \xi_{sc}^{u-1} \exp(-\xi_{sc} u).$$

Again, since this conditional posterior distribution cannot be sampled directly, I sample a candidate draw from the proposal density

$$u^* \mid \Xi_{-u}, D \sim G[a_u + \theta_u, (b_u^{-1} + \theta_u s_\xi^2)^{-1}],$$

where $\theta_u = 10$ is the tuning parameter, $s_\xi^2 = (\sum_{s=1}^{S} \sum_{c=1}^{C_s} 1)^{-1} \sum_{s=1}^{S} \sum_{c=1}^{C_s} (\xi_{sc} - \bar{\xi})^2$, and $\bar{\xi} = (\sum_{s=1}^{S} \sum_{c=1}^{C_s} 1)^{-1} \sum_{s=1}^{S} \sum_{c=1}^{C_s} \xi_{sc}$. The probability of accepting the candidate draw is $\min(R, 1)$, where

$$R = \frac{u^{*-\theta_u}\exp(\theta_u s_\xi^2 u^*)}{u_{j-1}^{-\theta_u}\exp(\theta_u s_\xi^2 u_{j-1})}\prod_{s=1}^{S}\prod_{c=1}^{C_s}\frac{u^{*u^*}\Gamma(u^*)^{-1}\xi_{sc}^{u^*-1}\exp(-\xi_{sc}u^*)}{u_{j-1}{}^{u_{j-1}}\Gamma(u_{j-1})^{-1}\xi_{sc}^{u_{j-1}-1}\exp(-\xi_{sc}u_{j-1})}.$$

7. Sample the individual level variance parameter r^{-1}:

$$p(r \mid D) \propto b_r^{-a_r}\Gamma(a_r)^{-1}r^{a_r-1}\exp(-rb_r^{-1})\prod_{s=1}^{S}\prod_{c=1}^{C_s}\prod_{i=1}^{I_{sc}} r^r\Gamma(r)^{-1}\alpha_{sci}^{r-1}\exp(-\alpha_{sci}r).$$

Once again, since this conditional posterior distribution cannot be sampled directly, I sample a candidate draw from the proposal density

$$r^* \mid \Xi_{-r}, D \sim G[a_r + \theta_r, (b_r^{-1} + \theta_r s_\alpha^2)^{-1}],$$

where $\theta_r = 10$ is the tuning parameter, $s_\alpha^2 = \left(\sum_{s=1}^{S}\sum_{c=1}^{C_s}\sum_{i=1}^{I_{sc}}\right)^{-1} \sum_{s=1}^{S}\sum_{c=1}^{C_s}\sum_{i=1}^{I_{sc}}(\alpha_{sci} - \bar{\alpha})^2$, and $\bar{\alpha} = \left(\sum_{s=1}^{S}\sum_{c=1}^{C_s}\sum_{i=1}^{I_{sc}}\right)^{-1}\sum_{s=1}^{S}\sum_{c=1}^{C_s}\sum_{i=1}^{I_{sc}} \alpha_{sci}$. The probability of accepting the candidate draw is $\min(R, 1)$, where

$$R = \frac{r^{*-\theta_r}\exp(\theta_r s_\alpha^2 r^*)}{r_{j-1}^{-\theta_r}\exp(\theta_r s_\alpha^2 r_{j-1})}\prod_{s=1}^{S}\prod_{c=1}^{C_s}\prod_{i=1}^{I_{sc}}\frac{r^{*r^*}\Gamma(r^*)^{-1}\alpha_{sci}^{r^*-1}\exp(-\alpha_{sci}r^*)}{r_{j-1}{}^{r_{j-1}}\Gamma(r_{j-1})^{-1}\alpha_{sci}^{r_{j-1}-1}\exp(-\alpha_{sci}r_{j-1})}.$$

8. Sample the coefficients β:

$$\begin{aligned} p(\beta \mid \Xi_{-\beta}, D) \propto & \left\{\prod_{s=1}^{S}\prod_{c=1}^{C_s}\prod_{i=1}^{I_{sc}}\exp\left[-\sum_{m=1}^{M}\exp(x_{scim}\beta)\eta_s\xi_{sc}\alpha_{sci}\lambda_m t_{scim}\right]\right. \\ & \left.\times\prod_{m=1}^{M}[\exp(x_{scim}\beta)\eta_s\xi_{sc}\alpha_{sci}\lambda_m]^{d_{scim}}\right\} \\ & \times\exp\left[-\frac{1}{2}(\beta-\beta_0)'V_\beta^{-1}(\beta-\beta_0)\right]. \end{aligned}$$

Since this conditional posterior distribution cannot be sampled directly, I sample a candidate draw from the proposal density

$$\beta^* \mid \beta_{j-1} \sim N(\beta_{j-1}, \Sigma_\beta),$$

where $\Sigma_\beta = \alpha_\beta(\sum_{s=1}^{S}\sum_{c=1}^{C_s}\sum_{i=1}^{I_{sc}} x'_{scim}x_{scim})^{-1}\sum_{s=1}^{S}\sum_{c=1}^{C_s}\sum_{i=1}^{I_{sc}} 1$ and $\alpha_\beta = 0.0001$ is the tuning parameter. The probability of accepting the candidate draw is $\min(R, 1)$, where

$$R = \left\{ \prod_{s=1}^{S} \prod_{c=1}^{C_s} \prod_{i=1}^{I_{sc}} \frac{\exp[-\sum_{m=1}^{M} \exp(x_{scim}\beta^*)\eta_s \xi_{sc} \alpha_{sci} \lambda_m t_{scim}]}{\exp[-\sum_{m=1}^{M} \exp(x_{scim}\beta_{j-1})\eta_s \xi_{sc} \alpha_{sci} \lambda_m t_{scim}]} \right.$$
$$\left. \times \frac{\prod_{m=1}^{M} [\exp(x_{scim}\beta^*)\eta_s \xi_{sc} \alpha_{sci} \lambda_m]^{d_{scim}}}{\prod_{m=1}^{M} [\exp(x_{scim}\beta_{j-1})\eta_s \xi_{sc} \alpha_{sci} \lambda_m]^{d_{scim}}} \right\}$$
$$\times \frac{\exp[-\frac{1}{2}(\beta^* - \beta_0)' V_\beta^{-1} (\beta^* - \beta_0)]}{\exp[-\frac{1}{2}(\beta_{j-1} - \beta_0)' V_\beta^{-1} (\beta_{j-1} - \beta_0)]}.$$

In the implementation, I simulate 20,000 iterations from the above algorithm and discard the first 4,000 iterations as preconvergence draws. To monitor and ensure the convergence of the algorithm, I also run iterations from several independent sequences, with starting values sampled from an overdispersed distribution following Gelman et al. (1995).

4. RESULTS

To examine the duration dependence pattern of the high school dropout hazard, I plot in Figure 1 the posterior mean of each month's dropout hazard of a representative individual $E(\lambda_m^* \mid Data)$ and the corresponding Kaplan–Meier estimate of the empirical hazard $\hat{\lambda}_m$, for $m = 1, 2, 3, \ldots, M$ and $M = 28$.[12] The Kaplan–Meier estimate (Kaplan and Meier, 1958; Kiefer, 1988) in month m is defined as $\hat{\lambda}_m = h_m/n_m$, where h_m is the number of students who dropped out of high school in month m and n_m is the number of students in the sample who were at risk of dropping out in month m. Note that, as compared with the estimates of the dropout hazard obtained from the model considered in this paper, the Kaplan–Meier estimates do not account for time-invariant or time-variant or observed or unobserved heterogeneity in the dropout hazard.

From Figure 1, the estimated dropout hazard increases during the first five months of the sampling period (February 1980 to June 1980), when the 1980 sophomores were still in their sophomore year. For the next

[12]The dropout hazard reported is associated with a representative individual who is male, white, from an urban area, and has a base year cognitive test score equal to the sample mean, a family income of $20,000, and three siblings. Both his parents have twelve years of education. His high school has a pupil–teacher ratio of 19, 15 books per pupil in the school's library, 48% of teachers with an M.A. or Ph.D. degree, 40% of teachers staying at school 10 years or more, and district expenditure per pupil of $2,000. The average family income of the students in the school is $20,000, and the average base year cognitive test score of the students in the school is equal to the sample mean. The representative individual is from a county with an employment growth rate from 1980 to 1982 of zero percent and is not eligible to drop out from high school under the state compulsory schooling laws.

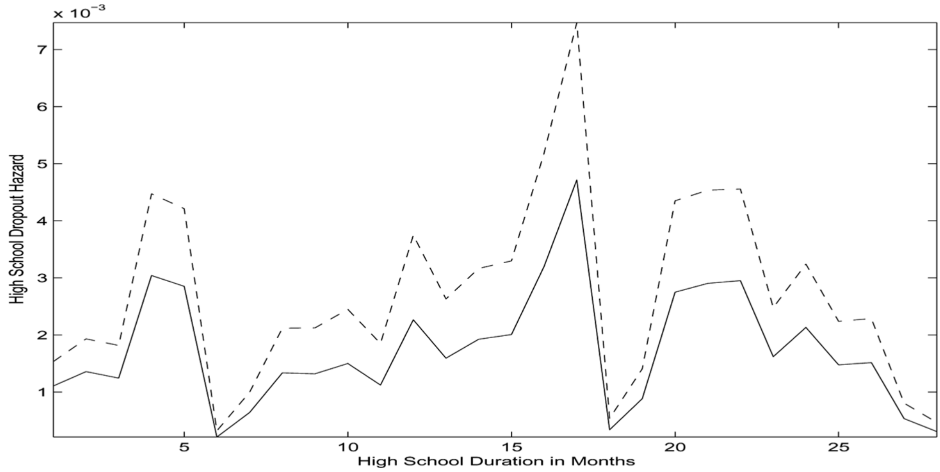

FIGURE 1 Posterior mean of each month's dropout hazard of a representative individual $E(\lambda_m^* \mid Data)$ (solid line) and the corresponding Kaplan–Meier estimate of the empirical hazard $\hat{\lambda}_m$ (dashed line), $m = 1, 2, 3, \ldots, M$ and $M = 28$.

twelve months (July 1980 to June 1981), the 1980 sophomores were in their junior year. During this period, the dropout hazard climbs continuously and reaches the highest point in June 1981. For the last eleven months of the sampling period (July 1981 to May 1982), the 1980 sophomores were in their senior year. Contrary to what we have observed in the previous years, the dropout hazard decreases during this last period.[13] One explanation for this duration dependence pattern is that during the sophomore and junior years, the school curriculum becomes more and more challenging over time, and this results in a gradual increase in the dropout hazard. In contrast, after students enter their senior year, the future benefits associated with graduation from high school become more and more attractive, and this leads to a decrease in the dropout hazard.

Next I discuss the coefficient estimates associated with the individual and school level covariates and the time-varying covariate. I list the posterior mean $E(\beta \mid Data)$, standard deviation $Std(\beta \mid Data)$, and probability of being positive $P(\beta > 0 \mid Data)$ of each coefficient in

[13]It is somewhat surprising to observe that the dropout hazards associated with the summer months, including the sixth (July 1980), seventh (August 1980), eighteenth (July 1981) and nineteenth (August 1981) months of the sampling period, are not exactly zero, although they are quite small and close to zero. A close look at the data set suggests that indeed a small proportion of the high school dropouts reported that they left high school during these summer months. Since in general schools are not in session during summer months, one may interpret that these students made their decisions on dropping out of high school during summer months and chose not to enroll in school when new academic years started.

TABLE 2 The posterior mean $E(\beta \mid Data)$, standard deviation $Std(\beta \mid Data)$, and probability of being positive $P(\beta > 0 \mid Data)$ of each coefficient, and the effect of each control variable on the dropout hazard $E(\Delta\%\lambda_m^* \mid \Delta x, Data)$

Explanatory variables[a]	$E(\beta \mid D)$[b]	$Std(\beta \mid D)$	$P(\beta > 0 \mid D)$	$E(\Delta\%\lambda_m^* \mid \Delta x, D)$[c]
Dropout eligibility	0.758	0.0838	1	114
Base year test score	−0.708	0.0362	0	−50.7
Female	−0.188	0.0518	0	−17.1
Minority	−0.379	0.0672	0	−31.4
Father's education	−0.0498	0.00853	0	−4.85
Mother's education	−0.0662	0.00937	0	−6.4
Family income ($10,000)	−0.0592	0.0297	0.0211	−5.71
Number of siblings	0.12	0.0158	1	12.7
Suburban	−0.23	0.0814	0.001	−20.3
Rural	−0.207	0.103	0.0119	−18.3
Pupil–teacher ratio	0.013	0.0072	0.965	1.31
Books per pupil	−0.00491	0.0031	0.0581	−0.49
% Teachers M.A./Ph.D.	−0.00123	0.00167	0.238	−0.123
% Teachers 10+ years	−0.000265	0.00143	0.419	−0.0264
District expend./pupil	−0.0864	0.0709	0.115	−8.05
Average fam. inc. ($10,000)	−0.139	0.0733	0.0362	−12.7
Average test score	−0.175	0.0392	0	−16
County employ. growth	0.0139	0.00621	0.976	1.4
Variance parameters				
Individual level (r^{-1})	0.798	0.115	1	
School level (u^{-1})	0.249	0.0415	1	
State level (v^{-1})	0.0541	0.0306	1	

[a]The descriptive statistics of the data are reported in Table 1. To reduce the number of missing observations, I also include in the model a set of dummy variables indicating whether an observation is missing for each of the following covariates: base year cognitive test score, father's education, mother's education, family income, number of siblings, pupil–teacher ratio, books per pupil, percentage of teachers with an M.A. or Ph.D. degree, percentage of teachers staying at school for 10 years or more, and district expenditure per pupil. I do not report the coefficient estimates associated with these dummy variables to save space.

[b]D denotes the data.

[c]$E(\Delta\%\lambda_m^* \mid \Delta x, Data)$ denotes the percentage change (%) in the dropout hazard resulting from a one-unit increase in the explanatory variable x, holding other covariates constant.

Table 2.[14] In many situations, one can interpret the posterior mean of a coefficient $E(\beta \mid Data)$ roughly as the marginal effect of the associated covariate on the dropout hazard, i.e., the percentage change in the dropout hazard due to a one-unit increase in the covariate, holding other covariates constant. This rule of thumb often serves as a good approximation when the marginal effect is relatively small and close to zero.

To see this, note that with a one-unit increase in a particular covariate x, the dropout hazard changes from λ_m^* to $\exp(\beta)\lambda_m^*$, where λ_m^* denotes the

[14]Since I adopt a Bayesian analysis framework in this paper, it is possible for me to calculate the posterior probability that a particular coefficient is positive. This posterior quantity of interest is analogous to the statistical significance often reported in most non-Bayesian analyses.

dropout hazard of a representative individual in month m and β represents the coefficient associated with the covariate x. The percentage change in the dropout hazard is hence captured by the quantity $[\exp(\beta)\lambda_m^* - \lambda_m^*]/\lambda_m^* = \exp(\beta) - 1$, and according to the Taylor expansion, $\exp(\beta) - 1 \approx \beta$ when β is close to zero. This suggests that using β to approximate the marginal effect of a covariate cannot be very accurate if β is substantially different from zero. For example, from Table 2, the coefficient estimate associated with the time-varying covariate of dropout eligibility is 0.76. This coefficient estimate suggests that the marginal effect of the dropout eligibility is $\exp(0.76) - 1 = 1.14$, which is substantially different from 0.76. To this end, I also calculate the exact marginal effect associated with each covariate, $E(\Delta\%\lambda_m^* \mid \Delta x, Data)$, and list these marginal effects in the last column of Table 2.

According to Table 2, the coefficient estimates in general have their expected signs and are statistically significant. As previously discussed, after students become eligible to drop out of high school under the state compulsory schooling laws, their dropout hazard increases by 114 percent. In addition, a one-standard-deviation increase in the base year cognitive test score reduces the dropout hazard by 51 percent. A one-year increase in the father's education decreases the hazard by 4.9 percent, and a one-year increase in the mother's education reduces the hazard by 6.4 percent. A ten-thousand-dollars increase in family income decreases the dropout hazard by 5.7 percent. Finally, having one more sibling increases the hazard by 13 percent.

An issue of substantial interest to the education policy makers that is commonly studied in the schooling literature is whether school quality, often measured in class size, district expenditure per pupil, etc., can improve various student outcomes. There is a huge debate in the literature regarding the effectiveness of school resources on students' outcomes (Betts, 1995, 1996; Card and Krueger, 1992; Figlio, 1999; Grogger, 1996a,b; Hanushek et al., 1996, 1998; Heckman et al., 1996; Hoxby, 1998, 2000). So far, much has been done to examine the effectiveness of different school quality variables on students' labor market outcomes, such as wages and unemployment, and their academic performance, such as students' scores on standard subject tests. However, little research has been done on the possible effects of school resources on students' decisions on dropping out of high school.

It is possible that school quality variables have a bigger influence on students' dropout behavior than on other outcomes such as wages and test scores. This paper examines the impacts of school resources on students' dropout decisions. From Table 2, a one-unit increase in the school pupil–teacher ratio raises the student's dropout hazard by 1.3 percent. Having one more book per student in school's library reduces the dropout hazard by 0.5 percent. A one-thousand-dollars increase in the district expenditure per pupil decreases the hazard by 8.1 percent. These results suggest that school resources are effective in preventing students from dropping out of high school.

One common concern in the school quality literature is the difficulty of disentangling the impacts of the school quality variables from the effects of the community characteristics associated with the school. In particular, high-quality schools are often related to wealthy communities and good social environment. Without an appropriate control for each student's family background and the community characteristics associated with each school, the estimated school quality effects may be confounded with the effects of family and community characteristics. To take into account this concern, I include in the model an array of individual and family characteristics that are discussed before, such as the base year cognitive test score, parental education, family income, and family size.

I also calculate the average family income of the students in a school and the average base year cognitive test score of the students in the school. I include these variables in the model to control for the impacts of community characteristics. Finally, a student may have a higher chance of dropping out of high school if there are more job opportunities in the local area. Thus I also control for the effect of local employment opportunity by including in the specification the county level employment growth rate from 1980 to 1982. The results from Table 2 suggest that a ten-thousand-dollars increase in the average family income reduces the dropout hazard by 13 percent. A one-standard-deviation increase in the average base year cognitive test score of the students in the school decreases the hazard by 16 percent. Finally, a one-percentage-point increase in the county level employment growth rate from 1980 to 1982 increases the hazard by 1.4 percent.

A well-known finding in the hazard analysis literature is that failure to account for the observed and unobserved heterogeneity in the hazard results in downward biases in the duration dependence estimates (Heckman, 1991; Heckman and Singer, 1984a,b). Hence it is essential to include in the model specification not only the observed individual and school level covariates but also unobserved heterogeneity in the hazard at various levels of the model. From Table 2, the variance parameter estimates of the unobserved heterogeneity in the hazard are 0.8, 0.25, and 0.054 at the individual, school, and state levels, respectively. These estimates suggest that substantial variation in the dropout hazard exists at different levels of the model.

To take a further look into this problem, I calculate the random effect in the dropout hazard associated with each individual, school, and state. In Figure 2, I plot the tenth, fiftieth, and ninetieth percentiles and the outliers of the simulated posterior draws of the log of each state's random effect in the hazard. Note that a normalization assumption

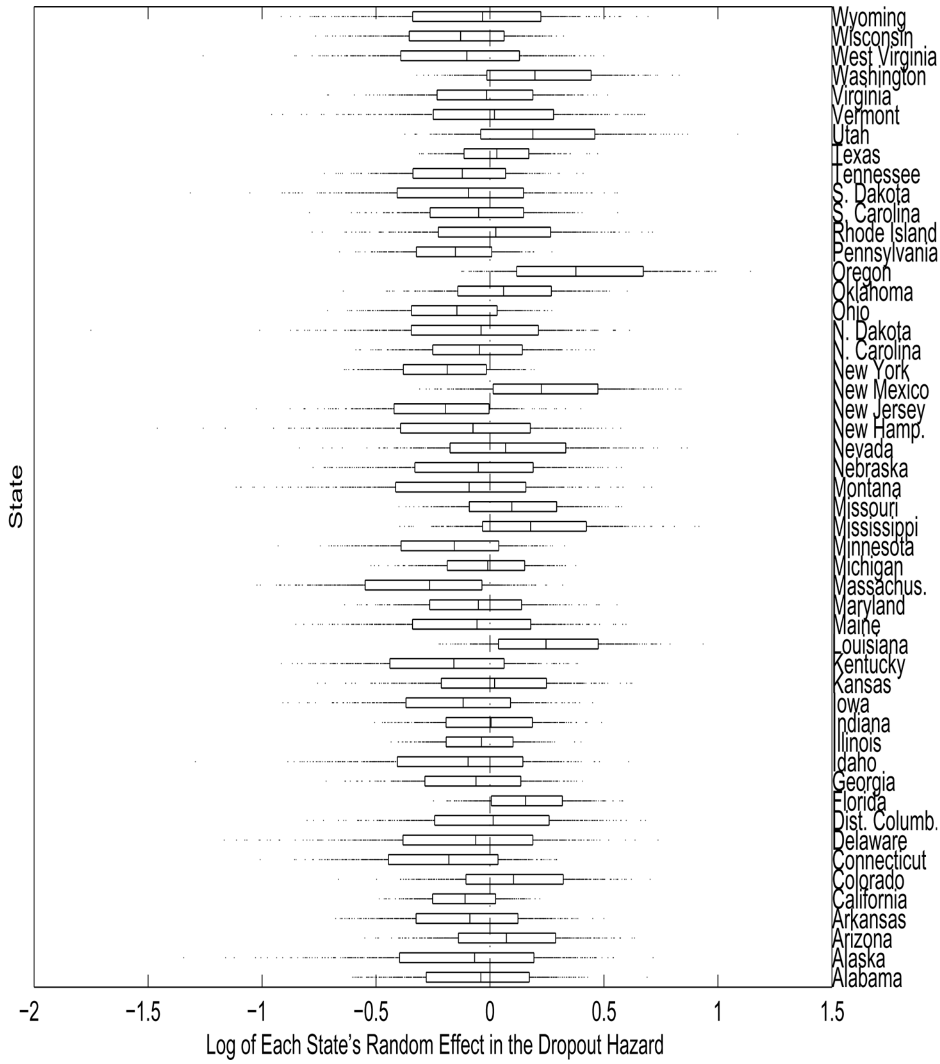

FIGURE 2 The tenth (left bar), fiftieth (middle bar), and ninetieth (right bar) percentiles and the outliers (whiskers) of the simulated posterior draws of the log of each state's random effect in the dropout hazard, $\log(\eta_s)$, $s = 1, 2, \ldots, S$.

adopted in this paper that is common in the hazard analysis literature is that random effects, such as η_s, ξ_{sc}, and α_{sci} which are specified in the model, are centered around one. Because these random effects are all positive numbers, it is natural to take the log of each state's random effect and compare it with the value of zero. From Figure 2, the states of Massachusetts, New Jersey, and New York are associated with lower dropout hazards as the logs of these states' random effects are smaller than zero with a posterior probability of 90 percent or more. In contrast, the states of Florida, Louisiana, New Mexico, and Oregon are connected with higher chances of dropping out of high school as the logs of these states' random effects are bigger than zero with a posterior probability of 90 percent or more. This sharp contrast again demonstrates the substantial variability in the dropout hazard across the entire nation.

4.1. Downward and Upward Biases in the Duration Dependence Estimates

A well-established claim in the duration analysis literature is that failure to account for the observed and unobserved heterogeneity in the hazard results in downward biases in the duration dependence estimates (Heckman, 1991; Heckman and Singer, 1984a,b). To verify this claim using the application studied in this paper, I estimate an alternative model that does not account for both time-invariant and time-varying heterogeneity in the dropout hazard. So far, the maintained model assumption is that the dropout hazard of individual i from school c in state s and in month m is $\lambda_{scim} = \exp(x_{scim}\beta)\eta_s\xi_{sc}\alpha_{sci}\lambda_m$. This specification, denoted as model (1) here, takes into account observed and unobserved, and time-invariant and time-variant heterogeneity in the dropout hazard. An alternative assumption of the dropout hazard is $\lambda_{scim} = \lambda_m$, and this is called model (2). Model (2) is intended to capture the change in the baseline dropout hazard over time, by specifying piecewise constant baseline hazard or month-specific dropout hazard. However, this approach does not capture the heterogeneity in the dropout hazard between different individuals.

I obtain the duration dependence estimates from model (2) and list them in the second column of Table 3. In particular, to characterize the duration dependence pattern of the dropout hazard, I calculate the posterior mean of the percentage change in each month's dropout hazard from the previous month, $\mathrm{E}[(\lambda_m - \lambda_{m-1})/\lambda_{m-1}\,|\,Data] \cdot 100$, for $m = 2, 3, \ldots, M$ and $M = 28$. For comparison, I also obtain the corresponding duration dependence estimates from model (1), the full model considered in this paper. I list these results in the first column of Table 3, calculate the differences in the duration dependence estimates between models (2) and (1), and list the differences in the fifth column of the table. These differences should reflect the directions of potential biases resulting from the failure to account for the heterogeneity in the dropout hazard. If there are downward biases in the duration dependence estimates, we should expect that the duration dependence estimates from model (2) will be all smaller than their corresponding estimates from model (1), or that the numbers in column five will all be negative. Contrary to my prior expectation, however, the biases of the duration dependence estimates calculated in column 5 are not all negative. In fact, most biases in the earlier months of the sampling period are positive. This does not seem to

TABLE 3 Posterior mean of the percentage change in each month's dropout hazard from the previous month, $m = 2, 3, \ldots, M$ and $M = 28$, from four different models

	Percentage change in each month's dropout hazard from the previous month[a]								
	Model				Difference between two models				
m	(1)[b]	(2)[c]	(3)[d]	(4)[e]	(2)–(1)	(3)–(1)	(2)–(3)	(4)–(1)	(2)–(4)
2	25.9	29.2	23.6	30.7	3.23	−2.37	5.6	4.75	−1.52
3	−6.43	−4.04	−11.9	−3.21	2.39	−5.5	7.89	3.22	−0.826
4	150	152	144	157	2.5	−5.68	8.18	6.92	−4.42
5	−5.37	−4.87	−8.18	−2.71	0.499	−2.82	3.31	2.66	−2.16
6	−92.5	−92.3	−92.6	−92.3	0.161	−0.136	0.297	0.193	−0.0317
7	246	255	245	257	9.04	−1.21	10.2	11	−1.95
8	116	121	114	123	4.9	−2.26	7.16	6.91	−2.01
9	0.685	2.29	−0.846	3.35	1.61	−1.53	3.14	2.67	−1.06
10	15.9	17.6	14.5	18.8	1.75	−1.38	3.13	2.91	−1.16
11	−24	−23.2	−24.7	−22.6	0.761	−0.704	1.47	1.38	−0.616
12	107	107	105	110	0.59	−1.82	2.41	3.62	−3.03
13	−28.9	−29	−29.6	−28.2	−0.0774	−0.618	0.541	0.737	−0.815
14	22.7	21.9	21.1	23.5	−0.735	−1.59	0.851	0.846	−1.58
15	5.57	5.63	3.86	6.99	0.0622	−1.71	1.77	1.42	−1.36
16	61.5	59.3	58.8	62.3	−2.22	−2.72	0.494	0.767	−2.99
17	48.5	45.4	44.4	49.4	−3.12	−4.14	1.02	0.832	−3.95
18	−92.8	−92.8	−92.8	−92.7	−0.0256	−0.0677	0.0421	0.0907	−0.116
19	182	184	182	185	1.52	−0.218	1.73	2.9	−1.38
20	221	220	216	222	−1.01	−4.24	3.24	1.69	−2.69
21	6.61	5.07	4.79	7.36	−1.54	−1.82	0.278	0.755	−2.3
22	2.66	1.48	0.787	3.52	−1.18	−1.88	0.695	0.853	−2.03
23	−44.7	−45.2	−45.5	−44.3	−0.467	−0.788	0.32	0.385	−0.852
24	34	33	32.4	34.8	−1.06	−1.57	0.509	0.8	−1.86
25	−29.8	−30	−30.4	−29.3	−0.181	−0.562	0.382	0.52	−0.701
26	4.75	4.21	3.75	5.2	−0.537	−0.991	0.455	0.456	−0.992
27	−64.1	−64.1	−64.4	−63.8	−0.0111	−0.228	0.217	0.323	−0.334
28	−39.2	−38.8	−39.2	−38.5	0.355	−0.0176	0.373	0.665	−0.31

[a]The estimate is the posterior mean of the percentage change (%) in the dropout hazard in month m from the previous month, $\mathrm{E}[(\lambda_m - \lambda_{m-1})/\lambda_{m-1} \mid Data] \cdot 100$, $m = 2, 3, \ldots, M$ and $M = 28$.

[b]Model (1) is the full model considered in the paper.

[c]Compared with model (1), model (2) does not account for both time-invariant and time-variant heterogeneity in the dropout hazard.

[d]Compared with model (1), model (3) does not account for time-invariant heterogeneity in the dropout hazard but accounts for the time-variant heterogeneity.

[e]Compared with model (1), model (4) does not account for time-variant heterogeneity in the dropout hazard but accounts for the time-invariant heterogeneity.

be consistent with the well-established claim that the failure to account for the heterogeneity in the hazard results in downward biases in the duration dependence estimates.

To take a deeper look into the issue, I choose to differentiate further between two different types of heterogeneity in the hazard. Specifically, I distinguish between time-invariant heterogeneity and time-varying heterogeneity. Thus, I estimate another model, called model (3), which, as compared with model (1), does not account for time-invariant heterogeneity in the hazard but accounts for time-varying heterogeneity. Thus in model (3), the dropout hazard associated with individual i from school c in state s and month m is $\lambda_{scim} = \exp(q_{scim}\gamma)\lambda_m$, where q_{scim} denotes the time-varying covariate of dropout eligibility and γ denotes the coefficient associated with the covariate. In other words, model (3) does not take into consideration the individual, school and state level covariates and random effects. However, model (3) still accounts for the time-varying heterogeneity in the dropout hazard, i.e., the time-varying covariate of dropout eligibility.

Estimates of the duration dependence pattern of model (3) are listed in the third column of Table 3. In addition, differences between the duration dependence estimates from model (3) and those from the previous two models, model (1) and model (2), are listed in the sixth and seventh columns, respectively. In particular, the differences between the duration dependence estimates from model (3) and model (1) should illustrate the directions of potential biases resulting from omitting only the time-invariant heterogeneity from the model. Interestingly, these differences are all negative and are consistent with the common intuition that failure to model the heterogeneity in the dropout hazard leads to downward biases in the duration dependence estimates. Moreover, a comparison between the duration dependence estimates from models (2) and (3) reveals a type of bias in the duration dependence estimates that has not been widely studied in the literature. Specifically, since model (2) does not account for both time-invariant and time-varying heterogeneity in the hazard, and model (3) only accounts for time-varying heterogeneity, the difference between the two models is that model (2) further omits the time-varying heterogeneity. Surprisingly, results show that omitting the time-varying heterogeneity from the model, in this application, leads to upward biases in the duration dependence estimates.

To verify this observation, I estimate yet another model, model (4), which, as compared with model (1), does not account for time-varying heterogeneity in the hazard but accounts for time-invariant heterogeneity. Formally, in model (4), the dropout hazard associated with individual i from school c in state s and in month m is $\lambda_{scim} = \exp(z_{sci}\theta)\eta_s\xi_{sc}\alpha_{sci}\lambda_m$, where z_{sci} denotes the time-invariant individual and school level covariates and θ denotes the vector of coefficients associated with these covariates. As compared with model (1), model (4) only excludes the time-varying covariate of dropout eligibility q_{scim}. I estimate the duration dependence pattern from model (4) and list the results in the fourth column of Table 3. The differences between the duration dependence pattern estimated from model (4) and those from models (1) and (2) are listed in the last two columns. As with what has been discussed before, the difference between models (1) and (4) is the absence of time-varying heterogeneity. Again, the results suggest that the exclusion of time-varying heterogeneity, in this application, leads to upward biases in the duration dependence estimates. On the other hand, the difference between models (2) and (4) is the further omission of the time-invariant heterogeneity. As with the previous results, the absence of time-invariant heterogeneity results in the well-known and more commonly observed downward biases in the duration dependence estimates.

These findings confirm that it is necessary to distinguish between the two types of heterogeneity in the hazard when we study the duration dependence pattern, namely the time-invariant heterogeneity and the time-varying heterogeneity. One may ask about the causes of the upward biases that we observe in this application. Before giving an explanation, it is worthwhile to reexamine briefly the causes of the more commonly observed downward biases in the duration dependence estimates owing to the failure to account for the time-invariant heterogeneity in the hazard. Note that the dropout hazard is roughly a student's conditional probability of dropping out of high school time t given that she has remained in school until t. Define the risk set as the group of students who have stayed until t. Since the dropout hazard is calculated by conditioning on the probability of survival, only those students who drop out at t and the group of students who have stayed until t (risk set) matter in the calculations of the dropout hazard. It is natural that students with higher dropout hazards (high-risk students) are more likely to drop out than students with lower dropout hazards (low-risk students). Therefore, on average, high-risk students drop out of high school earlier than low-risk students. This suggests that the proportion of high-risk students in the risk set decreases over time, and in contrast, the proportion of low-risk students in the risk set increases over time. Without an explicit modeling of the heterogeneity in the dropout hazard between different individuals (time-invariant heterogeneity), one would naively think that the dropout hazard decreases over time. In fact the truth is that the risk set is composed of a smaller proportion of high-risk students and a bigger proportion of low-risk students over time.

The explanation for the upward biases owing to the failure to account for the time-varying heterogeneity in the hazard turns out to be similar. To gain some insights, note that the time-varying heterogeneity in the hazard, in this paper, is the time-varying covariate of dropout eligibility. As students grow older, they are more likely to be eligible to drop out of high school under the state compulsory schooling laws. Hence the proportion of students in the risk set who are eligible to drop out of high school increases over time and the proportion of students in the risk set who are not eligible decreases over time. Moreover, the estimation results show that students who are eligible to drop out are associated with a higher likelihood of dropping out. Combined together, these two factors suggest that the risk set is composed of a bigger proportion of high-risk students and a smaller proportion of low-risk students over time, because more students become eligible to drop out as they grow older. As discussed before, without capturing in an explicit way the time-varying dropout eligibility (time-varying heterogeneity), one may be tempted to believe that the dropout hazard increases over time. Nevertheless, the truth is that the proportion of students in the risk set who are eligible to drop out increases over time and these students are associated with higher dropout hazard.

Finally, note that the existence of upward biases in the duration dependence estimates owing to the absence of time-varying heterogeneity from the model is not merely a theoretical concern. In the application considered in this paper, the magnitude of the upward biases is also quantitatively important and warrants particular attention. In fact, these upward biases are substantial and comparable to the more commonly noticed downward biases owing to the omission of time-invariant heterogeneity. Note again the comparison between the duration dependence estimates from model (2) and those from model (1). From column five of Table 3, it needs to be emphasized that, in the earlier months of the sampling period, the upward biases that are less commonly studied in the literature are even bigger than the more commonly observed downward biases.

5. CONCLUSION

In this paper, I examine the dropout behavior of high school students and study the timing of high school dropout decisions using data from High School and Beyond. I set up a Bayesian proportional hazard analysis framework that incorporates simultaneously the specification of piecewise constant baseline hazard, the time-varying covariate of dropout eligibility, and individual, school and state level random effects in the dropout hazard. I find that students who are eligible to drop out of high school under state schooling laws are 114 percent more likely to drop out than those who are not eligible. In addition, a one-unit increase in the school pupil–teacher ratio raises the student's dropout hazard by 1.3 percent. A one-thousand-dollars increase in the district expenditure per pupil decreases the hazard by 8.1 percent. An interesting finding that comes along with the empirical results is that the absence of modeling the time-varying heterogeneity in the hazard, in this paper, results in upward biases in the duration dependence estimates. Moreover, these upward biases are comparable in magnitude to the well-known downward biases in the duration dependence estimates owing to failure to account for the time-invariant heterogeneity in the hazard.

ACKNOWLEDGMENTS

I appreciate helpful comments and suggestions from Marigee Bacolod, David Brownstone, Linda Cohen, Dale Poirier, Sarah Senesky, Padhraic Smyth, Justin Tobias, and Jeremy Verlinda. I am also very grateful to the two anonymous referees, the Associated Editor in charge, and the Editor Esfandiar Maasoumi for their insightful and constructive comments, which have helped to improve the paper significantly. All errors are of course mine.

REFERENCES

Abbring, J. H., Van den Berg, G. J. (2007). The unobserved heterogeneity distribution in duration analysis. *Biometrika* 94(1):87–99.

Angrist, J. D., Krueger, A. B. (1991). Does compulsory school attendance affect schooling and earnings? *Quarterly Journal of Economics* 106(4):979–1014.

Bane, M. J., Ellwood, D. (1983). The dynamics of dependence: the routes to self-sufficiency. Report prepared for the Office of Planning and Evaluation, U.S. Department of Health and Human Services.

Betts, J. R. (1995). Does school quality matter? Evidence from the National Longitudinal Survey of Youth. *Review of Economics and Statistics* 77(2):231–250.

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

Blakemore, A. E., Low, S. A. (1984). The high-school dropout decision and its wage consequences. *Economics of Education Review* 3(2):111–119.

Bolstad, W. M., Manda, S. O. (2001). Investigating child mortality in Malawi using family and community random effects: a Bayesian analysis. *Journal of the American Statistical Association* 96:12–19.

Bray, J. W., Zarkin, G. A., Ringwalt, C., Qi, J. (2000). The relationship between marijuana initiation and dropping out of high school. *Health Economics* 9(1):9–18.

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):1–22.

Campolieti, M. (2003). On the estimation of hazard models with flexible baseline hazards and nonparametric unobserved heterogeneity. *Economics Bulletin* 3(24):1–10.

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

Chib, S., Greenberg, E. (1995). Understanding the Metropolis–Hastings algorithm. *American Statistician* 49(4):327–335.

Clayton, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* 65(1):141–151.

Clayton, D. G. (1991). A Monte Carlo method for Bayesian inference in frailty models. *Biometrics* 47(2):467–485.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society, Series B* 34(2):187–220.

DesJardins, S. L. Ahlburg, D. A., McCall, B. P. (1999). An event history model of student departure. *Economics of Education Review* 18(3):375–390.

Elbers, C., Ridder, G. (1982). True and spurious duration dependence: the identifiability of the proportional hazard model. *Review of Economic Studies* 49(3):403–409.

Figlio, D. N. (1999). Functional form and the estimated effects of school resources. *Economics of Education Review* 18(2):241–252.

Garfinkel, I., McLanahan, S. S. (1986). *Single Mothers and Their Children: A New American Dilemma.* Washington, DC: Urban Institute Press.

Gamerman, D. (1997). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference.* Chapman and Hall.

Ganderton, P. T. (1992). The effect of subsidies in kind on the choice of college. *Journal of Public Economics* 48:269–292.

Gelman, A., Carlin, J. B., Stern, H. S., Rubin, D. B. (1995). *Bayesian Data Analysis.* Chapman and Hall.

Gilks, W. R., Richardson, S., Spiegelhalter, D. J. (1996). *Markov Chain Monte Carlo in Practice.* Chapman and Hall.

Gørgens, T., Horowitz, J. L. (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.

Guo, G., Rodriguez, G. (1992). Estimating a multivariate proportional hazards model for clustered data using the EM algorithm, with an application to child survival in Guatemala. *Journal of the American Statistical Association* 87:969–976.

Han, A., Hausman, J. A. (1990). Flexible parametric estimation of duration and competing risk models. *Journal of Applied Econometrics* 5(1):1–28.

Hanushek, E. A., Taylor, L. L. (1990). Alternative assessments of the performance of schools: measurement of state variation in achievement. *Journal of Human Resources* 25(2):179–201.

Hanushek, E. A., Rivkin, S. G., Taylor, L. L. (1996). Aggregation and the estimated effects of school resources. *Review of Economics and Statistics* 78(4):611–627.

Hanushek, E. A., Kain, J. F., Rivkin, S. G., (1998). Teachers, schools and academic achievement. NBER working paper No. 6691.

Heckman, J. J. (1991). Identifying the hand of past: distinguishing State dependence from heterogeneity. *American Economic Review* 81(2):75–79.

Heckman, J. J., Singer, B. (1984a). A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica* 52(2):271–320.

Heckman, J. J., Singer, B. (1984b). The identifiability of the proportional hazard model. *Review of Economic Studies* 51(2):231–241.

Heckman, J. J., Layne–Farrar, A., Todd, P. (1996). Human capital pricing equations with an application to estimating the effect of schooling quality on earnings. *Review of Economics and Statistics* 78(4):562–610.

Holford, T. R. (1976). Life tables with concomitant information. *Biometrics* 32(3):587–597.

Horowitz, J. L. (1999). Semiparametric estimation of a proportional hazard model with unobserved heterogeneity. *Econometrica* 67(5):1001–1028.

Hoxby, C. M. (1998). The effects of class size and composition on student achievement: new evidence from natural population variation. NBER working paper No. 6869.

Hoxby, C. M. (2000). Peer effects in the classroom: learning from gender and race variation. NBER working paper No. 7867.

Hsing, Y. (1995). Impacts of welfare participation, school dropouts, and policy force on illegal drug arrests. *International Journal of Social Economics* 22(8):16–21.

Kalbfleisch, J. D. (1978). Non-parametric Bayesian analysis of survival time data. *Journal of the Royal Statistical Society, Series B,* 40(2):214–221.

Kaplan, E. L., Meier, P., (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* 53:457–481.

Kiefer, N. M. (1988). Economic duration data and hazard functions. *Journal of Economic Literature* 26(2):646–679.

Lancaster, T. (1979). Econometric methods for the duration of unemployment. *Econometrica* 47(4):939–956.

Lancaster, T. (1990). *The Econometric Analysis of Transition Data.* Cambridge University Press.

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.

Li, M. (2006). High school completion and future youth unemployment: new evidence from high school and beyond. *Journal of Applied Econometrics*, forthcoming.

Li, M., Poirier, D. J., Tobias, J. L. (2004). Do dropouts suffer from dropping out? Estimation and prediction of outcome gains in generalized selection models. *Journal of Applied Econometrics* 19:203–225.

McCaul, E. J., Donaldson, G. A., Coladarci, T., Davis, W. E. (1992). Consequences of dropping out of school: findings from high school and beyond. *Journal of Educational Research* 85(4):198–207.

Meyer, B. D. (1990). Unemployment insurance and unemployment spells. *Econometrica* 58(4):757–782.

Nickell, S. (1979). Estimating the probability of leaving unemployment. *Econometrica* 47(5):1249–1266.

Oakes, D. (1982). A model for association in bivariate survival data. *Journal of the Royal Statistical Society, Series B* 44(3):414–422.

Olsen, R. J., Farkas, G. (1989). Endogenous covariates in duration models and the effect of adolescent childbirth on schooling. *Journal of Human Resources* 24(1):39–53.

Poirier, D. J. (1995). *Intermediate Statistics and Econometrics*. Cambridge: MIT Press.

Rivkin, S. G. (1991). Schooling and Employment in the 1980's: Who Succeeds? Ph.D. dissertation, UCLA Department of Economics.

Ruggiero, M. (1994). Bayesian semiparametric estimation of proportional hazards models. *Journal of Econometrics* 62:277–300.

Sastry, N. (1997). A nested frailty model for survival data, with an application to the study of child survival in northeast Brazil. *Journal of the American Statistical Association* 92:426–435.

Stern, D., Paik, I.-W., Catterall, J. S., Nakata, Y.-F. (1989). Labor market experience of teenagers with and without high school diplomas. *Economics of Education Review* 8(3):233–246.

Van den Berg, G. J. (2001). Duration models: specification, identification and multiple durations. In: Heckman, J. J., Leamer, E., eds. *Handbook of Econometrics*. Vol. 5. Amsterdam: North-Holland.

Vaupel, J. W., Manton, K. G., Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* 16(3):439–454.