Estimating the Returns to Schooling: Implications from a Dynamic Discrete Choice Model†

Suqin Ge
Virginia Tech

November 2, 2012

Abstract

This paper assesses the applicability of a dynamic discrete choice model in accounting for the observed ordinary least squares (OLS) and instrumental variable (IV) estimates of the Mincer equation parameter on returns to schooling. A dynamic model of schooling and employment choices is estimated and used to simulate educational attainment, employment history, and wages. Estimations of the Mincer wage equation using simulated data appear to validate the model. Ability selection is found to be the major source of bias in the OLS estimates of schooling returns. Although the IV estimates lie within the support of true returns to schooling if a strong and strictly exogenous instrument is used and if dynamic employment selection is controlled, these conditions may be easily violated in practice.

Keywords: returns to schooling, instrumental variable (IV), ordinary least squares (OLS), dynamic discrete choice model

JEL code: I2, C1, C3.

†I would like to thank Philippe Belley, Zvi Eckstein, the editor and two anonymous reviewers for their comments and suggestions. Contact information: Ge, Department of Economics, Virginia Tech, Email: ges@vt.edu.

1 Introduction

The goal of an enormous empirical literature is to estimate the (average) rate of returns to schooling. Knowledge of the causal effect of school attainment on labor market rewards is crucial to drawing policy implications. Much of this research uses standard OLS estimates of variants of the human capital earnings function known as the "Mincer equation" (Mincer 1974):

$$\ln y = \beta_0 + \beta_1 S + \beta_2 X + \beta_3 X^2 + \varepsilon,$$

where y denotes some measure of earnings, S denotes the years of schooling, X stands for the years of experience, and ε is the wage residual.[1] The coefficient on schooling measures the (percentage) effect of incremental increases in schooling on earnings and is interpreted as the rate of return to schooling.[2] However, OLS estimates of returns to schooling are potentially biased because they cannot disentangle the effect of education on earnings from unobserved personal traits (e.g., innate skills or ability) correlated with schooling. Individuals with a higher level of skills will be more likely to obtain additional schooling. Therefore, OLS estimates of the schooling coefficient are upward-biased estimates of the true returns to schooling.

The search for methods for the estimation of the schooling coefficient has been ongoing for several decades. Different researchers use varying identification assumptions to correct the potential bias in OLS estimates. The literature is voluminous but can be categorized into several groups. The first approach includes explicit measures of ability in the wage regression.[3] However, the available test scores are, at best, proxies for the ability that is rewarded in the labor market. The second approach is largely focused on looking for valid instrumental variables (IV) that exploit the natural variation in factors affecting schooling decision.[4] Alternatively, data on twins (or siblings) are used to eliminate omitted-ability bias on the assumption that much of this ability is common across twins and can therefore

[1] Heckman, Lochner, and Todd (2006) challenge Mincer's functional form assumptions as well as the validity of using schooling coefficient as the measure of returns to schooling. Belzil and Hansen (2002) emphasize non-linearity in schooling. All the theoretical analyses do not rely on the specific functional form, although in the empirical study I follow Mincer's functional form assumptions, as is common in the literature.

[2] This interpretation is based on the model of human capital production in competitive market first proposed by Ben-Porath (1967). In wage equations derived from equilibrium search models (Burdett and Mortensen 1998), the schooling coefficient recovers composites of skill production parameters and other structural parameters, such as those of the job search technology (Wolpin 2003).

[3] For example, the Knowledge of the World of Work test score (Griliches 1977) and test scores from the Armed Services Vocational Aptitude Battery (Blackburn and Neumark 1995).

[4] Examples of instruments include Vietnam draft lottery (Angrist and Krueger 1992), date of birth (Angrist and Krueger 1991), minimum school leaving age (Harmon and Walker 1995), college proximity (Card 1993), college tuition (Kane and Rouse 1995), school subsidy (Lemieux and Card 2001), and school building program (Duflo 2001). Card (1999) provides an excellent survey.

be differenced out.[5]

OLS estimates of the average rate of return to schooling are somewhat stable across data sets from a given time period in the US.[6] With controls for test scores or family background, or other proxies for innate ability, estimates of the schooling coefficient are lower, indicating an upward ability bias. Although the IV approach has been developed to correct the upward bias in OLS estimates, the resulting estimates of the schooling coefficient using IVs are mostly as large or larger than the corresponding OLS estimates (Card 2001).

In this paper, I formulate and implement a standard dynamic discrete choice model of endogenous education and employment following Keane and Wolpin (1997) to investigate whether observed OLS and IV estimates of schooling returns in the Mincer equation can be reproduced in such models. To illustrate, a stylized two-period model of schooling and employment choices is considered first, and its analytical solution is derived. OLS and IV estimators are obtained under alternative assumptions on individual endowment, preferences, and behavior. These estimates are compared with the population average returns to schooling, and the sources of biases are discussed. In a generic specification of the model with ability heterogeneity, both OLS and IV estimates of schooling coefficients in the Mincer wage equation may be greater than or less than the true average returns to schooling, and their relative magnitude is indeterminate. Therefore, the dynamic choice model is adequately flexible to account for the observed estimates of schooling returns.

To assess quantitatively the model performance in accounting for observed schooling returns (i.e., OLS and IV estimates), I formulate and then estimate a dynamic choice model of human capital accumulation both in school and on the job. The model considers heterogeneous individuals characterized by different returns to schooling and utilities of attending school. Self-selection is controlled in the behavior model by allowing for unobserved types, and the dynamic decision process is solved for each type. Hence, the model implements a correction for selection biases. The model is estimated using a panel of white females taken from the National Longitudinal Survey of Youth 1979 (NLSY79). The estimates of population average return to schooling are in accordance with those from the literature on life-cycle education and employment choices.[7]

Individual choices on educational attainment and work experience are simulated from the estimated dynamic choice model along with wages. Mincer wage equations are estimated by OLS using the simulated data. Although wages are generated for a population with

[5] This approach is exemplified in Ashenfelter and Krueger (1994), Ashenfelter and Rouse (1998), among others. Rosenzweig and Wolpin (2000) provide insightful comments to this line of research.

[6] Increase in college premium in the US since the 1980s is well documented (e.g., Katz and Murphy 1992). Returns to education also vary across countries (e.g., Psacharopoulos 1994).

[7] See Belzil (2007) for an excellent survey.

relatively low rates of return to schooling (in the range of 3% to 6%), the OLS estimates are significantly higher (approximately 10%). The addition of a large number of ability measures in the wage regression reduces the schooling coefficient, indicating an upward ability bias. However, even if individual skill type is known and controlled, individual heterogeneity in the values of school and leisure will affect schooling and work decisions in a systematic manner that biases the schooling return estimates.

In addition, I investigate the properties of standard IV estimators by employing two IVs that are widely used in the literature, namely, the presence of a local college and a college subsidy program. Model simulation with instruments (interventions) mimics the scenario of a controlled experiment. Mincer wage equations are estimated using the IV technique, and the results are compared with OLS estimates and the population average. I find that an unbiased estimate of weighted average return to schooling may be generated by using the IV approach, but the requirements are stringent. First, the instrument must be strongly correlated with education outcome. Second, the instrument has to be strictly exogenous. Finally, dynamic employment selection has to be controlled. When these three conditions are satisfied, I find that the IV estimator is bounded by the maximal and minimal returns to schooling in the population. If any of the conditions is violated, an IV estimate may lie outside the support of the distribution of true returns and may be greater than the corresponding OLS estimate. In particular, IV estimates are very sensitive to the variations in schooling induced by the instrument. A small correlation between the instrument and the unobserved heterogeneity may result in a large bias in the estimated schooling coefficient. Overall, the estimates of schooling returns using OLS and IV methods are in accordance with the empirically estimated schooling returns using various data sets and also conform to the theoretical results derived from the simple model.

This paper is related to the literature on the discrepancy between OLS and IV estimates of schooling returns. The empirical analysis in Ashenfelter and Krueger (1994) suggests that the measurement error problem is important and that the OLS estimate is, in fact, biased downward. However, this study does not consider potential endogeneity bias in the schooling coefficient. Card (2001) postulates that returns to schooling vary in the population. The fact that IV estimates are larger than OLS estimates suggests that the return for the marginal person is greater than that for the average person. The same fact can be generated because OLS and IV estimates are different weighted averages of marginal treatment effects (Carneiro and Heckman 2003). Belzil and Hansen (2007) argue that an instrument that aims at increasing school attendance will not only affect the marginal person who would not have attended school, but also the continuation probabilities of those who have already attended school. Even if the marginal person has low returns to schooling, the IV estimates could be higher than the corresponding OLS estimates if the latter group has very high returns to schooling. Bound, Jaeger and Baker (1995) show that even a weak correlation between the instrument and the unobserved heterogeneity that affects wages can lead to a large bias in the IV estimator of the schooling coefficient by investigating the study of Angrist and Krueger (1991). However, the literature suggests no consensus on the features of the underlying data structure that produces the observed OLS and IV estimates of schooling returns or on why both estimates may lie outside the support of the true schooling return distribution.

The remainder of this paper is organized as follows: Section 2 presents a simple behavioral model, derives the OLS and IV estimators under alternative assumptions, and then discusses their properties. Section 3 briefly reviews the empirical model and the estimation result. Section 4 applies OLS and IV procedures on simulated data and then analyzes the implications. Section 5 concludes.

2 An Illustrative Model

2.1 A Two-period Model

In this section, I lay out a canonical model of schooling and employment decisions. Each person lives for two periods in the model and is endowed with ability μ, where μ is a draw from distribution $F_\mu(\cdot)$ defined on a finite set Ω. In the first period, an individual either attends school and pays a fixed cost cs, or works in the labor market. In the second period, the individual can no longer attend school and decides whether to work or stay at home. Following Becker (1975) and Mincer (1974), individuals face different levels of earnings associated with alternative schooling choices, work experience, and innate ability. Earnings depend on years of schooling (S), years of experience (X), innate ability (μ), and idiosyncratic productivity shocks (ε); they are denoted by $y(S, X, \mu, \varepsilon)$. Both schooling and experience enhance productivity; thus, $\partial y/\partial S > 0$ and $\partial y/\partial X > 0$. The earnings function is assumed to take the form of

$$\ln y(S, X, \mu, \varepsilon) = g_{S,\mu}(S, X, \mu) + \varepsilon. \tag{1}$$

The contemporaneous utility is assumed to be linear in consumption (c_t) and leisure (v_t). The value of leisure, v, is also assumed to depend on education, experience, and ability. Each individual solves the following optimization problem:

$$\max_{s_1 \in \{0,1\}, h_2 \in \{0,1\}} E\{c_1 + \beta[c_2 + v(S_2, X_2, \mu)(1 - h_2)]\} \tag{2}$$

$$\begin{aligned} s.t. \quad c_1 + cs \cdot s_1 &= (1 - s_1) y(S_1, X_1, \mu, \varepsilon_1) &(3) \\ c_2 &= h_2 y(S_2, X_2, \mu, \varepsilon_2) &(4) \end{aligned}$$

where s_1 equals 1 if school attendance is chosen and 0 otherwise, h_2 equals 1 if employment is chosen and 0 otherwise, and β is the discount rate. The expectation is taken over the distribution of the earnings. Note that, for each individual, the innate ability μ is persistent, but ε_1, ε_2 are *i.i.d.* shocks drawn from distribution $F_\varepsilon(\cdot)$.

The model is solved backwards. At $t = 2$, the alternative-specific value function conditional on individual education, experience, and ability can be written as follows:

$$\begin{aligned} V_2(h_2 &= 1 | S_2, X_2, \mu) = y(S_2, X_2, \mu, \varepsilon_2), \\ V_2(h_2 &= 0 | S_2, X_2, \mu) = v(S_2, X_2, \mu). \end{aligned}$$

The individual works if and only if

$$\varepsilon_2 \geq \varepsilon_2^*(S_2, X_2, \mu) = \ln v(S_2, X_2, \mu) - g_{S,\mu}(S_2, X_2, \mu). \tag{5}$$

If $\partial \ln v / \partial S > \partial g / \partial S$, the income effect dominates. As an individual becomes more educated, she values leisure more; therefore, she is less likely to work. If $\partial \ln v / \partial S < \partial g / \partial S$, the substitution effect dominates. As an individual becomes more educated, her opportunity cost of leisure is higher; therefore, she is more likely to work.

At $t = 1$, the present value of attending school is

$$\begin{aligned} V_1[(s_1 = 1, h_1 = 0)] &= -cs + \beta E \max[V_2(h_2 = 1), V_2(h_2 = 0) | S_2 = 1, X_2 = 0, \mu], \\ &= -cs + \beta \left[\int_{-\infty}^{\varepsilon_2^*(1,0,\mu)} v(1, 0, \mu) dF_\varepsilon(\varepsilon) + \int_{\varepsilon_2^*(1,0,\mu)}^{\infty} y(1, 0, \mu, \varepsilon) dF_\varepsilon(\varepsilon) \right]. \end{aligned}$$

The present value of working is

$$\begin{aligned} V_1[(s_1 = 0, h_1 = 1)] &= y(0, 0, \mu, \varepsilon_1) + \beta E \max[V_2(h_2 = 1), V_2(h_2 = 0) | S_2 = 0, X_2 = 1, \mu] \\ &= y(0, 0, \mu, \varepsilon_1) + \beta \left[\int_{-\infty}^{\varepsilon_2^*(0,1,\mu)} v(0, 1, \mu) dF_\varepsilon(\varepsilon) + \int_{\varepsilon_2^*(0,1,\mu)}^{\infty} y(0, 1, \mu, \varepsilon) dF_\varepsilon(\varepsilon) \right]. \end{aligned}$$

An individual attends school if and only if

$$V_1[(s_1 = 1, h_1 = 0)] \geq V_1[(s_1 = 0, h_1 = 1)].$$

The employment decision is characterized by the following cut-off rule: an individual works (and does not attend school) if and only if

$$\varepsilon_1 \geq \varepsilon_1^*(\mu). \tag{6}$$

Clearly, schooling is negatively correlated with the productivity shocks. Schooling is also correlated with ability, but the sign of correlation $\partial\varepsilon_1^*(\mu)/\partial\mu$ is indeterminate without further restrictions.[8]

2.2 Returns to Schooling

Returns to schooling are defined as the average return to $S = 1$ in the population. Conditional on experience X, the average return b equals

$$\begin{aligned} & E_{\mu,\varepsilon}\left[\ln y\left(1, X, \mu, \varepsilon\right) - \ln y\left(0, X, \mu, \varepsilon\right)\right] \\ = & \int\int \left[\ln y\left(1, X, \mu, \varepsilon\right) - \ln y\left(0, X, \mu, \varepsilon\right)\right] f_\mu\left(\mu\right) f_\varepsilon\left(\varepsilon\right) d\mu d\varepsilon, \end{aligned} \tag{7}$$

where f_μ and f_ε are density functions of μ and ε.

Although the two-period model is simple, it illustrates the self-selection in the observed earnings. At $t = 1$, $y(0, 0, \mu, \varepsilon_1)$ is observed if and only if $\varepsilon_1 \geq \varepsilon_1^*(\mu)$. At $t = 2$, $y\left(0, 1, \mu, \varepsilon_2\right)$ is observed if and only if $\varepsilon_1 \geq \varepsilon_1^*(\mu)$ and $\varepsilon_2 \geq \varepsilon_2^*(0, 1, \mu)$; $y\left(1, 0, \mu, \varepsilon_2\right)$ is observed if and only if $\varepsilon_1 < \varepsilon_1^*(\mu)$ and $\varepsilon_2 \geq \varepsilon_2^*(1, 0, \mu)$.

When cross-sectional observed wages are used in both periods, the OLS estimator calculates the difference between the average earnings of the more educated $(S = 1)$ and the average earnings of the less educated $(S = 0)$. Let Ω_1 be the set of individuals that attend school and $\Omega_0 = \Omega\backslash\Omega_1$ be the set of individuals that do not attend school. Conditional on

[8] At $\varepsilon_1^*\left(\mu\right)$, an individual with ability μ is indifferent between attending school and working.

$$\begin{aligned} & -cs + \beta\left[\int_{-\infty}^{\varepsilon_2^*(1,0,\mu)} v\left(1, 0, \mu\right) dF(\varepsilon) + \int_{\varepsilon_2^*(1,0,\mu)}^{\infty} y\left(1, 0, \mu, \varepsilon\right) dF(\varepsilon)\right] \\ = \; & \exp[g_{S,\mu}(0, 0, \mu) + \varepsilon_1^*\left(\mu\right)] + \beta\left[\int_{-\infty}^{\varepsilon_2^*(0,1,\mu)} v\left(0, 1, \mu\right) dF(\varepsilon) + \int_{\varepsilon_2^*(0,1,\mu)}^{\infty} y\left(0, 1, \mu, \varepsilon\right) dF(\varepsilon)\right] \end{aligned}$$

Taking the total derivative yields $\frac{\partial\varepsilon_1^*(\mu)}{\partial\mu}$. As μ increases, both the opportunity cost of schooling and future return will increase. Therefore, the effect of μ on school enrollment is theoretically indeterminate. Empirical evidence shows that its effect on opportunity cost is typically below its effect on future returns, and thus a positive correlation between ability and school enrollment is observed.

$X = 0$, the OLS estimate of returns to schooling is[9]

$$\begin{aligned} b_{OLS} &= E_{\mu,\varepsilon}\left[\ln y\left(1,0,\mu,\varepsilon\right) - \ln y\left(0,0,\mu,\varepsilon\right)\right] \\ &= \int_{\Omega_1}\int_{\varepsilon_2^*(1,0,\mu)}^{\infty} \ln y\left(1,0,\mu,\varepsilon\right) f_\varepsilon\left(\varepsilon\right) d\varepsilon f_\mu\left(\mu\right) d\mu \\ &\quad - \int_{\Omega_0}\int_{\varepsilon_1^*(\mu)}^{\infty} \ln y\left(0,0,\mu,\varepsilon\right) f_\varepsilon\left(\varepsilon\right) d\varepsilon f_\mu\left(\mu\right) d\mu. \end{aligned} \tag{8}$$

To demonstrate what the IV method estimates, I consider a policy instrument. Compulsory schooling is a widely used instrument following Angrist and Krueger (1991). Consider a compulsory schooling law that forces a randomly selected half of all individuals to attend school. Those that are not subject to the law are used as the control group as their schooling decision is not affected. The other half subject to the law become the treatment group. The IV estimator identifies the local average treatment effect ($LATE$) introduced in Imbens and Angrist (1994), i.e., the average returns of these individuals who would not have attended school without the enforcement of compulsory schooling law.[10] Therefore,

$$\begin{aligned} b_{IV} &= \int_{\Omega_0}\int_{\varepsilon_2^*(1,0,\mu)}^{\infty} \ln y\left(1,0,\mu,\varepsilon\right) f_\varepsilon\left(\varepsilon\right) d\varepsilon f_\mu\left(\mu\right) d\mu \\ &\quad - \int_{\Omega_0}\int_{\varepsilon_1^*(\mu)}^{\infty} \ln y\left(0,0,\mu,\varepsilon\right) f_\varepsilon\left(\varepsilon\right) d\varepsilon f_\mu\left(\mu\right) d\mu. \end{aligned} \tag{9}$$

Equations (8) and (9) show that OLS and IV are the weighted average of returns to education for different sub-populations. As individuals vary in endowment (innate ability) and luck (idiosyncratic shock), they self-select themselves into different schooling levels and employment status. The IV estimate may be greater than the OLS estimate even without measurement error in observed schooling.

Next, I simplify the general framework to discuss the properties of the OLS and IV estimators under alternative assumptions. In particular, I assume that ability differs only in two types, which are denoted by μ_1 and μ_0, where $\mu_1 > \mu_0$, and let π be the proportion of type μ_1.

[9] Strictly speaking, in the model the OLS estimator conditional on $X = 0$ (i.e., controlling for experience) is the difference between the second period wages of those who do attend school and the first period wages of those who do not. However, a typical cross-sectional sample usually has individuals from different cohorts, thus the OLS estimator conditional on experience can compare the observed wages of those who do attend school with the wages of those who do not from the same time period.

[10] As emphasized by Heckman and Urzúa (2010), IV is a weighted average of the effects for individuals induced into a state from different margins. In this simple case, IV induces behavior change only from one margin.

Case 1: Exogenous employment decision

Assume that individuals always work when they are not in school. Thus, the only decision is whether to attend school in the first period. There is no uncertainty in earnings, and earnings only depend on schooling and ability. A high-ability individual is more productive at all schooling levels and their corresponding occupations, i.e., $\ln y(S,\mu) = g(S,\mu)$ and $\partial g/\partial\mu > 0$. The value of choosing to attend school is

$$V(S=1) = -cs + \beta y(1,\mu),$$

and the value of not attending school is

$$V(S=0) = y(0,\mu) + \beta y(0,\mu).$$

The decision rule is to attend school if $V(S=1) \geq V(S=0)$. If

$$\beta\frac{\partial y(1,\mu)}{\partial\mu} - (1+\beta)\frac{\partial y(0,\mu)}{\partial\mu} > 0,$$

then individuals with ability at or above some cut-off value of ability μ^* would attend school and those below it would not. Assume that $\mu_1 > \mu^* > \mu_0$. The optimization of schooling choice then creates a positive correlation between ability and completed schooling. All high-ability individuals will attend school, and all low-ability ones will not. Based on observed wages, the OLS estimate of the return to schooling calculates the earnings differences between the two schooling groups:

$$\begin{aligned} b_{OLS} &= E_\mu[\ln y(1,\mu)\,|S=1] - E_\mu[\ln y(0,\mu)\,|S=0] \\ &= E_\mu[\ln y(1,\mu)\,|\mu \geq \mu^*] - E_\mu[\ln y(0,\mu)\,|\mu < \mu^*] \\ &= \ln y(1,\mu_1) - \ln y(0,\mu_0). \end{aligned}$$

The "true" return for high-ability individuals is $\ln y(1,\mu_1) - \ln y(0,\mu_1)$ and that for low-ability individuals is $\ln y(1,\mu_0) - \ln y(0,\mu_0)$. The average returns to schooling in the population are

$$\begin{aligned} b &= E_\mu[\ln y(1,\mu) - \ln y(0,\mu)] \\ &= \pi[\ln y(1,\mu_1) - \ln y(0,\mu_1)] + (1-\pi)[\ln y(1,\mu_0) - \ln y(0,\mu_0)]. \end{aligned}$$

OLS estimate is unbiased if and only if ability is homogeneous, and therefore there is no selection into school based on ability. The bias of OLS estimate is easy to calculate and is determined by

$$b_{OLS} - b = (1-\pi)\left[\ln y(1,\mu_1) - \ln y(1,\mu_0)\right] + \pi\left[\ln y(0,\mu_1) - \ln y(0,\mu_0)\right].$$

As $\frac{\partial y(1,\mu)}{\partial \mu} > 0$ and $\frac{\partial y(0,\mu)}{\partial \mu} > 0$, $b_{OLS} > b$. This difference is known as the "ability bias." Note that even when the returns to schooling are identical in both ability types, i.e., $\ln y(1,\mu_1) - \ln y(0,\mu_1) = \ln y(1,\mu_0) - \ln y(0,\mu_0)$, OLS estimate is still upward biased.

Instrumental variables are used to obtain an estimate of returns to schooling without ability bias. Consider a compulsory schooling law that forces a randomly selected half of individuals to attend school. The Wald estimate of returns to schooling then compares the earnings of the control group and the treatment group, and it is determined by

$$\begin{aligned} b_{IV} &= E_\mu\left[\ln y(1,\mu)\,|\mu < \mu^*\right] - E_\mu\left[\ln y(0,\mu)\,|\mu < \mu^*\right] \\ &= \ln y(1,\mu_0) - \ln y(0,\mu_0). \end{aligned}$$

Therefore, the IV estimator identifies returns to schooling for the low ability type.

If returns to schooling are identical for all ability groups, IV estimate is unbiased and consistent. If returns to schooling are heterogeneous in the population, then IV estimate is biased. In particular, when the average return to schooling is higher for the type affected by the instrument

$$\ln y(1,\mu_0) - \ln y(0,\mu_0) > \ln y(1,\mu_1) - \ln y(0,\mu_1),$$

then IV estimate is upward biased. This is Card's (2001) hypothesis on why IV estimates of the returns to schooling are typically as large or larger than the corresponding OLS estimates. He suggests that if low-ability individuals choose a low schooling level because of the higher than average costs of schooling, rather than because of the lower than average returns to schooling, IV estimates will overstate the average marginal return to schooling in the population by identifying the "local average treatment effect." However, in this example, although IV estimate is upward biased, it remains smaller than OLS estimate. To be consistent with the empirical results, OLS estimate has to be downward biased for reasons beyond the "ability bias."[11]

Case 2: Endogenous employment decision and homogeneous ability

Schooling is negatively correlated with productivity shock when labor supply is endogenous. Those who receive high wages are more likely to leave school earlier. To study the effect of endogenous employment separately, I assume that the population is homogeneous in

[11] Measurement error in schooling is a classical source of downward bias (Griliches 1977).

ability for now, and I suppress μ in functions of earnings such that $\ln y(S, X, \varepsilon) = g(S, X) + \varepsilon$. The decision rule then becomes: at $t = 1$, an individual attends school if and only if $\varepsilon_1 < \varepsilon_1^*$ and at $t = 2$, an individual works if and only if $\varepsilon_2 \geq \varepsilon_2^*(S, X)$. The average returns to schooling conditional on X equal

$$\begin{aligned} b(X) &= E_\varepsilon [\ln y(1, X, \varepsilon) - \ln y(0, X, \varepsilon)] \\ &= \int [\ln y(1, X, \varepsilon) - \ln y(0, X, \varepsilon)] f_\varepsilon(\varepsilon) d\varepsilon. \end{aligned}$$

OLS estimate conditional on $X = 0$ identifies

$$b_{OLS} = \int_{\varepsilon_2^*(1,0)}^{\infty} \ln y(1, 0, \varepsilon) f_\varepsilon(\varepsilon) d\varepsilon - \int_{\varepsilon_1^*}^{\infty} \ln y(0, 0, \varepsilon) f_\varepsilon(\varepsilon) d\varepsilon.$$

Owing to employment selection, OLS estimates are biased. If the magnitude of sample selection is smaller in the second period, the OLS estimates are lower than the true returns. For example, consider an extreme case in which there is no sample selection in the second period, i.e., $\varepsilon_2^*(1, 0) = -\infty$. It is trivial that $b_{OLS}(X = 0) < b(X = 0)$.

Again, if a compulsory schooling law forces a randomly selected half to attend school in the first period, then IV estimates are determined by

$$b_{IV} = \int_{\varepsilon_2^*(1,0)}^{\infty} \ln y(1, 0, \varepsilon) f_\varepsilon(\varepsilon) d\varepsilon - \int_{\varepsilon_1^*}^{\infty} \ln y(0, 0, \varepsilon) f_\varepsilon(\varepsilon) d\varepsilon.$$

As ε_1 and ε_2 are *i.i.d.*, $b_{OLS} = b_{IV}$. When the population is homogeneous and productivity shocks are serially uncorrelated, OLS and IV estimates are the same, and they are likely to be downward biased.

Case 3: Endogenous employment decision and heterogeneous ability

Next, I consider a more realistic case similar to the model presented in Equations (2)–(4), where employment decision is endogenous and individual ability differs in two types, μ_1 and μ_0, $\mu_1 > \mu_0$. Population size is normalized to one, and let π be the proportion of type μ_1. The decision rule is such that at $t = 1$, an individual attends school if and only if $\varepsilon_1 < \varepsilon_1^*(\mu)$ as defined in Equation (6), and at $t = 2$, an individual works if and only if $\varepsilon_2 \geq \varepsilon_2^*(S, X, \mu)$ as defined in Equation (5). The average returns to schooling conditional on $X = 0$ equal

$$\begin{aligned} b &= \pi \int [\ln y(1, 0, \mu_1, \varepsilon) - \ln y(0, 0, \mu_1, \varepsilon)] f_\varepsilon(\varepsilon) d\varepsilon \\ &\quad + (1 - \pi) \int [\ln y(1, 0, \mu_0, \varepsilon) - \ln y(0, 0, \mu_0, \varepsilon)] f_\varepsilon(\varepsilon) d\varepsilon. \end{aligned}$$

The choice model predicts that $\pi F\left(\varepsilon_1^*(\mu_1)\right)$ of type μ_1 and $(1-\pi)\,F\left(\varepsilon_1^*(\mu_0)\right)$ of type μ_0 will attend school in the first period. Therefore, OLS estimates of returns to schooling conditional on $X=0$ are determined by

$$\begin{aligned} b_{OLS} &= \pi\gamma_1 \int_{\varepsilon_2^*(1,0,\mu_1)}^{\infty} \ln y\left(1,0,\mu_1,\varepsilon\right) f_{\varepsilon}\left(\varepsilon\right) d\varepsilon + (1-\pi)\,\gamma_0 \int_{\varepsilon_2^*(1,0,\mu_0)}^{\infty} \ln y\left(1,0,\mu_0,\varepsilon\right) f_{\varepsilon}\left(\varepsilon\right) d\varepsilon \\ &\quad -\pi(1-\gamma_1) \int_{\varepsilon_1^*(\mu_1)}^{\infty} \ln y\left(0,0,\mu_1,\varepsilon\right) f_{\varepsilon}\left(\varepsilon\right) d\varepsilon - (1-\pi)\left(1-\gamma_0\right) \int_{\varepsilon_1^*(\mu_0)}^{\infty} \ln y\left(0,0,\mu_0,\varepsilon\right) f_{\varepsilon}\left(\varepsilon\right) d\varepsilon. \end{aligned}$$

where $\gamma_1 = F\left(\varepsilon_1^*(\mu_1)\right)$ and $\gamma_0 = F\left(\varepsilon_1^*(\mu_0)\right)$. Hence,

$$\begin{aligned} & b_{OLS} - b \\ =\; & \pi(\gamma_1 - 1) \int_{\varepsilon_2^*(1,0,\mu_1)}^{\infty} \ln y\left(1,0,\mu_1,\varepsilon\right) f_{\varepsilon}\left(\varepsilon\right) d\varepsilon + (1-\pi)\left(\gamma_0 - 1\right) \int_{\varepsilon_2^*(1,0,\mu_0)}^{\infty} \ln y\left(1,0,\mu_0,\varepsilon\right) f_{\varepsilon}\left(\varepsilon\right) d\varepsilon \\ & +\pi\gamma_1 \int_{\varepsilon_1^*(\mu_1)}^{\infty} \ln y\left(0,0,\mu_1,\varepsilon\right) f_{\varepsilon}\left(\varepsilon\right) d\varepsilon + (1-\pi)\,\gamma_0 \int_{\varepsilon_1^*(\mu_0)}^{\infty} \ln y\left(0,0,\mu_0,\varepsilon\right) f_{\varepsilon}\left(\varepsilon\right) d\varepsilon. \end{aligned}$$

Similarly a compulsory schooling law will send a random half of those who would have worked in the first period to school. Therefore, IV procedure identifies

$$\begin{aligned} b_{IV} &= \pi(1-\gamma_1) \int_{\varepsilon_2^*(1,0,\mu_1)}^{\infty} \ln y\left(1,0,\mu_1,\varepsilon\right) f_{\varepsilon}\left(\varepsilon\right) d\varepsilon + (1-\pi)\left(1-\gamma_0\right) \int_{\varepsilon_2^*(1,0,\mu_0)}^{\infty} \ln y\left(1,0,\mu_0,\varepsilon\right) f_{\varepsilon}\left(\varepsilon\right) d\varepsilon \\ &\quad -\pi(1-\gamma_1) \int_{\varepsilon_1^*(\mu_1)}^{\infty} \ln y\left(0,0,\mu_1,\varepsilon\right) f_{\varepsilon}\left(\varepsilon\right) d\varepsilon - (1-\pi)\left(1-\gamma_0\right) \int_{\varepsilon_1^*(\mu_0)}^{\infty} \ln y\left(0,0,\mu_0,\varepsilon\right) f_{\varepsilon}\left(\varepsilon\right) d\varepsilon. \end{aligned}$$

The difference between IV estimate and OLS estimate is

$$\begin{aligned} & b_{IV} - b_{OLS} \\ =\; & \pi(1-2\gamma_1) \int_{\varepsilon_2^*(1,0,\mu_1)}^{\infty} \ln y\left(1,0,\mu_1,\varepsilon\right) f_{\varepsilon}\left(\varepsilon\right) d\varepsilon + (1-\pi)\left(1-2\gamma_0\right) \int_{\varepsilon_2^*(1,0,\mu_0)}^{\infty} \ln y\left(1,0,\mu_0,\varepsilon\right) f_{\varepsilon}\left(\varepsilon\right) d\varepsilon. \end{aligned}$$

The sign of the differences between the true return, OLS, and IV estimates depends on the relative size of the skill types (π), and selections in schooling choice (ε_1^*) and employment choice (ε_2^*), which both depend on individual heterogeneity μ.

Table 1 summarizes the properties of OLS and IV estimates of returns to schooling under various assumptions. The result presented in this table is based on a one-time discrete choice of school attendance and on using compulsory schooling as an instrument, which forces a randomly selected half to attend school.

As discussed above, OLS and IV procedures generate consistent estimates of returns to schooling under very strict assumptions. A dynamic model of endogenous schooling and employment choices for heterogenous individuals is capable of generating OLS and IV estimates of schooling returns in accordance with those observed in the data. That is, OLS estimates are likely to be upward biased, and IV estimates may be greater than OLS estimates.

3 The Empirical Model

In this section, I present an empirical model in which individuals make school attendance and employment decisions simultaneously. The model is an extended version of the dynamic discrete choice model discussed in the previous section. I structually estimate the model to recover the underlying preference parameters, as well as the ability and earnings distributions.

3.1 The Model

The previous discussion demonstrates that sample selection based on employment choice may be an important source of bias when returns to schooling are estimated in a reduced-form earnings regression. To assess better the effect of work experience on estimates of returns to schooling, I consider the life-cycle choices of women, who have larger variations in employment choices than men.

Consider a model in which young women make joint decisions on school attendance and work.[12] Each year, a woman with a high school degree decides whether to attend college and whether to work. In total, there are four mutually exclusive and exhaustive alternatives. Let s_t and h_t be the indicators for school attendance and employment, respectively. Each alternative will be $(s_t, h_t) \in J = \{(s_t, h_t) : s_t \in \{0, 1\}, h_t \in \{0, 1\}\}$. The contemporaneous utility $U_t(s_t, h_t)$ associated with choice (s_t, h_t) is given by

$$\begin{aligned} U_t(c_t, s_t, h_t) &= (\alpha_1 + \alpha_2 s_t + \alpha_3 h_t)c_t \\ &\quad + v_1 s_t + v_2(1 - h_t) + v_3 s_t(1 - s_{t-1}) + v_4 s_t h_t + \epsilon_t^{(s,h)}. \end{aligned} \tag{10}$$

The utility function is assumed to be linear in consumption c_t. The marginal utility of consumption depends on college attendance and employment as captured by the parameters α_2 and α_3, respectively. Parameters v_1 and v_2 evaluate the net utility of attending school and the net utility of not working, respectively. Parameter v_3 captures the adjustment cost of returning to school, and v_4 captures the additional (dis)utility of attending school while working.

[12] The decision model closely follows Keane and Wolpin's (1997) model on career decisions of young men and is a simplified version of Ge (2011).

Finally, $\epsilon_t^{(s,h)}$s are alternative-specific random components representing random variations in the individual's preference for school and work. They are known to the individual in period t but are unknown before t.

The choice decision is subject to the budget constraint given by

$$c_t + c_S \cdot s_t = y_t h_t. \tag{11}$$

c_S is the direct cost of schooling. The direct cost of one year in college equals cs. A college degree is assumed to be completed in four years. When a woman attends graduate school, she pays an additional tuition cost cg, that is, $c_S = cs + cg$. y_t denotes the annual earnings of the woman. The budget constraint is assumed to be satisfied period by period.[13]

An unemployed woman receives job offers with probability p_0 every year. As in Eckstein and Wolpin (1989), potential annual earnings are obtained by multiplying hourly wage by 2000 hours, that is, $y_t = w_t \cdot 2000$. Essentially, each woman is assumed to be deciding about full-time work, and the wage rate is assumed to be independent of hours worked. For the purpose of this work, I specify the earnings function in detail. Hourly wage offer is assumed to be dependent on prior education and work experience, as measured by cumulative years of schooling S_t, cumulative years of experience X_t, and an idiosyncratic shock. Thus, the wage function is given by

$$\ln w_t = \beta_0 + \beta_1 S_t + \beta_2 X_t + \beta_3 X_t^2 + \epsilon_{wt}.$$

The schooling coefficient β_1 measures the earnings return to each additional year of school. The quadratic term in work experience is meant to capture the depreciation of human capital, such that wage is hump-shaped over the life cycle. The productivity shock ϵ_{wt} is normally distributed with a mean zero and standard deviation σ_w. A measurement error in observed wages is allowed, such that $\ln w^o = \ln w + u$, where w^o is the observed wage, w is the true wage, and the error term is normally distributed: $u \sim N(0, \sigma_u^2)$. At time t, the woman observes the wage rate (hereby earnings) and then decides whether to work. Before time t, she does not observe ϵ_{wt}, but she knows how wage and earnings evolve, as well as the distribution of ϵ_{wt}. She is also aware of the expected earnings gains from education in the labor market, as denoted by β_1.

The model considered above corresponds to the decision problem of a representative woman. However, young women differ in numerous aspects, such as their family backgrounds

[13] The saving decision is simplified from the model because limited evidence exists to support that borrowing constraints play an important role in educational attainment (Cameron and Heckman 1998, 2001; Cameron and Taber 2004).

as measured by parental education levels, number of siblings, and family income; as well as their cognitive backgrounds, as measured by Armed Forces Qualification Test (AFQT) scores. The abilities and preferences of individual women are also likely to vary in unobserved ways (e.g., motivation, perseverance, or ambition) that are both persistent and correlated with observed traits. These characteristics may affect a young woman's college and employment decisions, as well as her earnings.

Assume that there exist $k = 1, 2, \cdots, K$ different skill types. Denote the *ex ante* probability that a woman i is of type k by π_i^k. Let π_i^k depend on her observed initial traits, including mother's schooling S_i^m, father's schooling S_i^f, number of siblings N_i^{sib}, household structure (whether she lives with both parents) at age 14 HH_i, net family income Y_i^0, AFQT score $AFQT_i$, and age at high school graduation AGE_i^0, in the form of a multinomial logit. For $k = 2, \cdots, K$,

$$\pi_i^k = \frac{\exp\left[\begin{array}{c} \lambda_0^k + \lambda_1^k S_i^m + \lambda_2^k S_i^f + \lambda_3^k N_i^{sib} + \lambda_4^k HH_i \\ +\lambda_5^k Y_i^0 + \lambda_6^k AFQT_i + \lambda_7^k AGE_i^0 \end{array}\right]}{1 + \sum_{l=2}^K \exp\left[\begin{array}{c} \lambda_0^l + \lambda_1^l S_i^m + \lambda_2^l S_i^f + \lambda_3^l N_i^{sib} + \lambda_4^l HH_i \\ +\lambda_5^l Y_i^0 + \lambda_6^l AFQT_i + \lambda_7^l AGE_i^0 \end{array}\right]}, \tag{12}$$

and normalize π_i^1 as $1 - \sum_{k=2}^K \pi_i^k$.[14]

Women of different skill types have distinct preferences for school and non-employment (the v's in the utility function), as well as different earning returns to schooling (β_1). Therefore, these parameters are type specific and potentially correlated with observed characteristics. For skill type k, the wage offer when working after college is determined by

$$\ln w_{it} = \beta_0 + \beta_1^k S_{it} + \beta_2 X_{it} + \beta_3 X_{it}^2 + \epsilon_{iwt}. \tag{13}$$

Therefore, the earnings function has random coefficients because of individual heterogeneity.

The costs and benefits of choices on school and employment are also affected by a number of individual-specific exogenous factors. In particular, the direct cost of college for individual i at date t is specified as

$$cs_{it} = \gamma_0 + \gamma_1 Col_{it} + u_{it}. \tag{14}$$

The variable Col_{it} is a dummy for the presence of any college in individual i's county of residence. Card (1993) and a number of subsequent studies show that the existence of a local college reduces the cost of college. Therefore, the coefficient γ_1 is expected to be

[14]The unobserved types may vary in multidimensional skills and preferences such as innate ability, motivation, perseverance, and tastes for school, and they have no natural ordering. Therefore a multinomial logit model for types is chosen over an ordered logit model.

negative. The constant γ_0 represents the cost of college in a county without a local college. The error term u_{it} is $i.i.d.$ idiosyncratic shocks, which can be absorbed into the utility shocks associated with school attendance.

The objective of individual i is to maximize the expected present discounted value of utility over a finite horizon from the first year after high school graduation to a known terminal time T, that is,

$$\max_{\{c_{it},s_{it},h_{it}\}} E\left[\sum_{t=1}^{T} \beta^{t-1} U_{it}(c_{it}, s_{it}, h_{it}|\Psi_{it})\right],$$

where $\beta > 0$ is the woman's subjective discount factor, and Ψ_{it} is the state space at time t. The state space consists of all factors, known to the person, that affect current utilities or the probability distribution of any of the future utilities. Choice of the optimal sequence of control variables $\{c_{it}, s_{it}, h_{it}\}$ for $t = 1, \cdots, T$ maximizes the expected present value given the current realization of the state space. The model can be solved backwards numerically.

To solve the optimization problem, the value function $V_{it}(\Psi_{it})$ is defined as the maximal value of the individual i's optimization problem at t

$$V_{it}(\Psi_{it}) = \max_{\{c_{it},s_{it},h_{it}\}} E\left[\sum_{\tau=t}^{T_i} \beta^{\tau-t} U(c_{i\tau}, s_{i\tau}, h_{i\tau}|\Psi_{it})\right]. \tag{15}$$

The value function can be written as the maximum over alternative-specific value functions $V_{it}(\Psi_{it}) = \max_{(s_t,h_t)\in J}\{V_{it}^{(s,h)}(\Psi_{it})\}$, which obeys the Bellman equation

$$V_{it}^{(s,h)}(\Psi_{it}) = U_{it}\left(c_t, s_t, h_t\right) + \beta E[V_{it+1}(\Psi_{it+1})|\Psi_{it},\ \ (s_t, h_t)\ \text{ is chosen at } t]. \tag{16}$$

The alternative-specific value function assumes that future choices are optimally made for any given current decision.

The solution to the model can be characterized by sequential cut-off rules. In this multiple-period model with multiple choices, the cut-off values do not have analytical forms, but the model can be solved backwards, and the cut-off values can be simulated numerically. Similar to Keane and Wolpin (2001) and Eckstein and Wolpin (1999), the backward recursion starts at a computationally convenient terminal period, $T_i = T^*$.[15] During the first $T^* - 1$ periods, for each individual i, the model is solved explicitly. At the terminal period

[15] In the empirical estimation, the terminal period is set to 10, such that $T^* = 10$. The model was solved explicitly for 10 years for all individuals.

T^*,

$$V_{iT^*}^{(s,h)}(\Psi_{iT^*}) = U_{iT^*}(c_{T^*}, s_{T^*}, h_{T^*}) + \beta E[V_{iT^*+1}(\Psi_{iT^*+1})|\Psi_{iT^*}, (s_{T^*}, h_{T^*}) \text{ is chosen at } T^*]. \tag{17}$$

Similar to the rest of the model, prior random shocks to wages or preferences only affect decisions through state variables, including years of schooling and experience. Therefore, a polynomial form of the state variables at the terminal period was used to estimate the terminal value function[16], that is,

$$V_{iT^*+1}(\Psi_{iT^*+1}) = \delta_1 S_{iT^*+1} + \delta_2 X_{iT^*+1} + \delta_3 X_{iT^*+1}^2. \tag{18}$$

The parameters of this terminal condition are estimated along with the structural parameters of the model.

Using the end condition and assuming a known distribution of ϵ_{it}, each individual's optimization problem was solved recursively from the final period T^*. Solving the dynamic programming problem requires high-dimensional integrations for computing the "$E\max$ function" at each point of the state space. As discussed in Keane and Wolpin (1994), Monte Carlo integrations were used to evaluate the integrals.

3.2 Data and Estimation Method

The micro data are taken from the 1979 to 1998 waves of the NLSY79. The empirical analysis is based on a sample of 487 females who graduated from high school between 1980 and 1983, with a total of 4,770 person-year observations. Table 2 shows the proportion of women who choose each of the four alternatives and their average wages for 10 years after high school graduation. Among all women in the sample, 48.5% attended college in the first year after high school graduation. The proportion of full-time college attendees decreases annually throughout the first three years. After the fourth year, a discrete drop is observed, corresponding to typical college graduation. The labor force participation rate exhibits the well-known hump shape. It increases from 43% to approximately 80% in the first six years and then becomes flat and declines slightly.[17] The proportion of working students hovers at approximately 11% in the first four years and falls to approximately 6% after seven years,

[16] In an early specification search, a skill-type specific constant was included in the terminal condition. However, the estimated values were not significant from zero; therefore, the constant was dropped from Equation (18).

[17] NLSY79 work history records weekly hours worked for each week since the beginning of 1978. Annual hours worked is based on accumulating weekly hours worked over a year. A woman in the model is defined as employed (full-time) if her working hours are reported during at least 26 weeks of the year, and annual hours worked are at least 1000 hours.

reflecting the fact that some women return to school. Observed real hourly wage increases from \$6.17 to \$12.70 over the 10 years for those who are employed. All wages are measured in 2000 dollars. Detailed family and cognitive background variables, such as parental education and AFQT scores, are constructed for the selected sample. These background variables are highly correlated with school outcomes.

Table 3 presents the average year-to-year transition rates. The row percentage is the percentage of transition from origin to destination choices, and the column percentage is the percentage in a particular destination choice that started from each origin. The transition matrix provides evidence on strong state dependence, as indicated by the large percentages on the diagonal. An individual who is in school full time (not working) in one year stays in school full time the next year 58% of the time (row percentage), whereas 76% of those in school full time in any year came from school the previous year (column percentage). Considerable state dependence is also observed in employment. Approximately 86% of employed workers not in school in one year remained employed the following year. Table 3 also reveals the considerable immobility in the no school and no work option, but the working at school option is less persistent as people move out of school over time.

The model is estimated by the simulated maximum likelihood. At any time t, denote the vector of outcomes as $O_t = \{(s_t, h_t), w_t^o\}$. The likelihood function for a sample of I individuals from period $t = 1, 2, ..., T^*$ is given by

$$\Pi_{i=1}^{I} \Pr(O_{i1}, O_{i2}, ..., O_{iT^*} | \Psi_{i0}),$$

where Ψ_{i0} is the initial state space. The joint serial independence among the shocks implies that the likelihood function can be written as the product of within-period outcome probabilities.

The solution to the individual's optimization problem provides the within-period choice probabilities. To illustrate the computation of the likelihood, let us consider a specific outcome at some period. Suppose a woman who chooses not to attend school ($s = 0$) but opts to work full time instead ($h = 1$) reports receiving a wage w_t^o in period t. Further, assume that the individual enters the period being unemployed and having state space Ψ_t. The probability of this outcome is

$$\begin{aligned} & \Pr[(0,1), w_t^o | \Psi_t] \\ = \; & p_0 \Pr[V_t^{(0,1)} = \max_{j \in J} V_t^j | w_t, \Psi_t] \Pr(w_t, w_t^o | \Psi_t). \end{aligned} \tag{19}$$

This probability has three components: the first term on the right-hand side is the probability of receiving a job offer p_0; the second term is the choice probability of not attending school and accepting the job offer; and the last term on the right-hand side of (19) is the probability of observing the woman's wage w_t^o. The choice probability involves the calculation of multivariate integrals, similar to general multinomial choice problems. I calculate the joint probability of choosing $(s_t = 0, h_t = 1)$ conditional on the true wage by a smoothed simulator following Eckstein and Wolpin (1999). For each of $n = 1, 2, ..., N$ draws of the error vector, the $\epsilon' s$, a smoothed simulator of the probability that $(0, 1)$ is chosen, is given by the kernel

$$\exp[\frac{V_{tn}^{(0,1)} - \max_{j\in J}(V_{tn}^j)}{\tau}]/\sum_i \exp[\frac{V_{tn}^i - \max_{j\in J}(V_{tn}^j)}{\tau}],$$

with τ as the smoothing parameter, which is set to 500. The integral is then the average of the kernel over the N draws. The probabilities of observing a reported wage w_t^o for the woman are the joint density of the observed and true wages. The probabilities of other outcomes are calculated in a similar manner.

For the selected sample indexed by $i = 1, \cdots, I$, I observe each individual's family and cognitive background, the presence of any college at her county of residence Col_i, schooling, employment status every year (s_{it}, h_{it}), and wages if employed (w_{it}^o) for $t = 1, \cdots, T^*$. I assume that the parameters describing the initial preferences, ability, and market skills are related to the measured family and cognitive background. As previously discussed, there are K discrete types in total, and each type is described by a vector of parameters. The likelihood function for individual i in this case is a finite mixture of the type-specific likelihoods, that is,

$$L_i(\theta) = \prod_{t=1}^{T^*} \sum_{k=1}^{K} \pi_i^k \Pr[(s_{it}, h_{it}), w_{it}^o | \Psi_{it}],$$

where the skill-type probability π_i^k is determined by Equation (12). The sample log-likelihood function is

$$\log L(\theta) = \sum_{i=1}^{I} \log L_i(\theta).$$

The resulting estimate of θ, $\widehat{\theta}$ satisfies

$$\sqrt{N}(\widehat{\theta} - \theta_0) \to N(0, E[s_i(\theta_0)s_i(\theta_0)']^{-1}),$$

where $s_i(\theta_0) = \partial L_i(\theta_0)/\partial\theta'$, and θ_0 is the true value of θ.

3.3 Estimation Results

The model is fit with three skill types ($K = 3$). I present the estimates of the earnings equation and preference heterogeneity in Table 4.[18] The population appears to be very heterogeneous. The rate of return to schooling is the lowest for the first type, followed by type two, whereas type three has the highest return to schooling. Each additional year of schooling increases wages by 2.9%, 4.3%, and 5.9%, respectively, for each type. The estimated utility values of school indicate significant preference heterogeneity among skill types. According to the rank order of the values, type three has the highest value of school, followed by type two, with type one as the lowest, independent of working status. Type three has the highest value of non-employment, and type two has the highest cost to return to school. According to estimated logit parameters in type probabilities (not shown), higher parental education, fewer siblings, living with both parents at age 14, higher family income, good AFQT score, and graduation from high school at an early age imply a higher probability of being skill type two relative to type one. Parental schooling, family income, and AFQT score also have a positive (but lesser) effect on the probability of being skill type three.

Given the estimated parameters, I calculate the predicted proportions of women who choose each alternative in every year after high school. Figure 1 shows the fit of the model to the choice proportions. Each of the profiles implied by the estimated model has approximately the right shape and matches the levels of the data closely. Table 5 presents the predicted average year-to-year transitions rates, compared with the actual percentages of transition from origin to destination choices as shown in Table 3. The model can match transitions reasonably well. The data demonstrates much persistence in each state; the model recovers persistence in the state of full-time school attendance (not working) and in the state of full-time employment (not attending school) but understates the persistence in nonemployment when women are not in school. Figure 2 demonstrates further the fitting of wage distribution. The model-predicted wage moments follow the data closely overall, but there is a rise in wages between the 5th and 8th year that is not well predicted by the model. This discrepancy indicates that a simple log wage equation cannot capture all the year-to-year wage dynamics over the lifetime.

The empirical model is a more realistic version of the two-period discrete choice model discussed in the previous section. Several features of the model are important to our discus-

[18] Given that the cost of college (cs) enters the model linearly with the value of schooling, γ_0 is set to be 7,515 in the estimation based on the estimates from the National Center for Education Statistics from 1980 to 1988 (NCES 1990, 285, Table 291). I did not attempt to estimate the discount factor β because it is unlikely to be well identified separately from the terminal value function. Therefore it is set to 0.96, corresponding to an annual interest rate of approximately 4%. The full set of parameter estimates are available upon request from the author.

sion on the observed estimates of returns to schooling.

First, heterogeneity is explicitly modeled. Individuals differ in family and cognitive backgrounds, which are observed. The model also allows the ability and preference of individuals, which are both persistent and correlated with observed traits, to vary in unobserved ways (e.g., motivation, perseverance, or ambition). Self-selection is controlled in the behavior model by allowing for unobserved types in skills, and the dynamic decision process is solved for each type. Hence, the model implements a correction for the selection bias.

Second, returns to schooling vary in the population and have relatively tight bound between 3% and 6%. Once the distribution of returns to schooling is known, properties of various estimators can be evaluated.

Third, an endogenous employment decision generates a selected sample of observed wages, and this employment selection depends on work experience. If there is a high wage realization at time t, and an individual chooses to work, then at time $t+1$, her experience is higher, and she is more likely to work. The cutoff wage realization for her to choose to work could be lower.

4 Simulations

4.1 Simulated Data and Sample Statistics

Based on within-sample goodness of fit, the dynamic discrete choice model can be concluded to be a good approximation of how individuals make schooling and employment decisions. In the remainder of the paper, I will treat this estimated behavioral model as the true underlying data-generating process and simulate individual lifetime choices and their labor market outcomes. The joint decision on schooling and employment determine the realized (observed) wages.

A total of 100,000 individuals' schooling and employment choices are simulated for 10 years based on the estimated model. I first resample (with replacement) individuals' initial traits, including mother's schooling, father's schooling, number of siblings, household structure, net family income, AFQT score, age at high school graduation, and existence of a local college, from the NLSY79 empirical sample used to estimate the model. The dynamic programming problem is then solved for each individual, and their choices on schooling and employment are simulated. I also simulate observed wages (for those who are employed) and potential wages (for both employed and non-employed individuals). The true wage equation is determined by Equation (13). I attempt to recover the returns to schooling parameters presented in Table 4.

Table 6 presents the descriptive statistics of the simulated sample. These descriptive statistics are nearly identical to those of the NLSY79 sample used to estimate the choice model. Mother's average school attainment is 12.3 years, and father's average schooling is slightly higher at 12.6 years but is more dispersed. Individuals in the sample have close to three siblings, and 86% of them lived with both parents at age 14. Net family income is above 65,000 in 2000 dollars. On average, the women graduated from high school at the age of 17.9, and they continued for more than two years of college after graduating from high school and had close to six years of total work experience. The average observed hourly wage is approximately 9.70 dollars, whereas the potential hourly wage is below the observed wage at 9.38, as predicted by the economic theory.

4.2 OLS Estimates

Table 7 reports the OLS estimates of the wage equation using simulated wages. I also report the White–Huber standard errors that consider the correlation at the individual level over time.[19] In the first two columns, I use observed wages for those employed that take into account measurement errors. In the benchmark specification in column 1, the schooling coefficient is approximately 10.3%. As more background variables are added as controls in column 2, the schooling coefficient decreases slightly to 10.1%. This result indicates upward ability bias, given that background variables are likely proxy for innate ability. To investigate the potential bias attributable to measurement error in wages, the dependent variable used in columns 3 and 4 is the simulated true wage. The estimated coefficients based on true wages are almost identical to those based on observed wages, but they are more precisely estimated. Overall measurement error in wages does not seem to play a significant role in the wage equation estimates. In the last two columns, the schooling coefficient drops to 9.8% and 9.1% when I run the same regressions by using all potential (true) wages simulated from the model, for all individuals both employed and non-employed. This result suggests that employment selection may be important. Compared with the true parameter values in Table 4, the coefficients on years of schooling are upward biased in all OLS regressions.

To investigate the effects of ability selection in OLS estimates, Table 8 compares the true wage equation parameters with the OLS estimates controlling for skill types. In particular, I estimate the wage Equation (13) using simulated observed wages and all potential wages, while each individual's skill type is assumed to be known. Schooling coefficient is set to

[19] The estimates reported in Table 7 are based on a pooled repeated cross-sectional sample. Alternatively, I have selected randomly one wage per individual to form a cross-section of wages. The coefficient estimates in the wage equation are very similar to those reported in Table 7 but are less precise, given that the number of observation is considerably smaller.

be type-specific, whereas all other parameters in the wage equation are independent of skill type. Estimated coefficients on years of schooling for all types are significantly lower than the estimates based on the pooled sample shown in column (1) of Table 7. When observed wages are used as the dependent variable, the schooling coefficients are downward biased for all skill types. However, when all potential wages are used in the regression, as the last column of Table 8 shows, the point estimates of wage equation parameters are very close to the true values.

Estimates in Table 8 reveal that ability selection is the main driving force behind the upward bias in OLS estimates of schooling coefficient, but endogenous employment selection is also a non-negligible source of bias. I simulate wages for homogeneous samples to investigate further the sources of bias aside from ability selection. The discrete choice model is estimated on three skill types. I start with simulating 100,000 individuals for 10 years for skill type one by set model parameters to type one's specific estimates. I then repeat the procedure for type two and type three. Within each set of simulated data, all individuals have the same ability and preferences.

Table 9 presents OLS estimates on these homogeneous samples, where both observed wages and all potential wages are again used as dependent variables. When the endogenous employment decision is not considered, estimates of schooling coefficients using observed wages are biased for all three homogeneous samples. For the type one individuals who spend less time in school and who work longer, the selection effect attributable to endogenous employment is relatively small; for the types two and three who spend more time in the non-employment state, the biases are considerably larger. The differences in schooling and labor market outcomes are determined by the ability and preference heterogeneity demonstrated in Table 4. The economic theory described in Section 2 predicts that if the sample is homogeneous and employment decision is exogenous, OLS estimates are consistent and unbiased. Consistent with this prediction, estimates using all potential wages return unbiased estimates of schooling returns for each skill type in Table 9.

4.3 IV Estimates

The conventional IV approach estimates the following two-equation system describing the determination of earnings and years of schooling:

$$\begin{aligned} \ln w_{it} &= \mathbf{X}_{it}\boldsymbol{\delta}+\beta_1 S_{it} + \varepsilon_{it}, \\ S_{it} &= \mathbf{Z}_{it}\boldsymbol{\alpha} + u_{it}, \end{aligned}$$

where $\mathbf{X}$ and $\mathbf{Z}$ are the vectors of observed attributes, with $\mathbf{X}$ usually including experience, experience square, and other control variables. The IV approach identifies β_1 only if two conditions are satisfied. First, some of the variables in vector $\mathbf{Z}$ are not contained in $\mathbf{X}$, and they are strongly correlated with schooling outcome. Second, an IV that is not correlated with the error term in the wage equation, that is, $Cov\left(\mathbf{Z}_{it}, \varepsilon_{it}\right) = 0$, must exist. In what follows, two different instruments used in the literature are considered. Individual choices and wages are simulated, in which people are randomly assigned to a control group and a treatment group given each of the instruments. OLS and IV estimates from the reduced-form wage equations are presented and compared.

Example one: Local college

The first set of instruments for education that I explore is based on the existence of a local college, following the idea of Card (1993). As specified in Equation (14), the existence of a local college provides a source of variation in education outcome by shifting the cost of school in the underlying data-generating process. IV identification relies on the inclusion of a dummy variable for the presence of a college in the county of residence in the set of variables $\mathbf{Z}$. The dummy variable is set to one if a local college exists. Table 10 summarizes the estimates using the conventional OLS approach and using IVs based on local college. Column 1 presents the OLS estimate of the returns to schooling, which is upward biased and similar to results in Table 7. In columns 2 to 9, I compare estimates using simulated data from four different instrument variable designs based on local college. In addition, the upper panel A uses observed wage data generated by the decision model, and the lower panel B uses all potential wage data for both employed and non-employed individuals.

I will start with the simulated observed wages. First, I simulate a textbook-style strong IV in a sample of 100,000 individuals, among which a randomly selected half is assumed to live in a county with a local college, with the rest living in a county without a local college. This instrument is correlated with education outcome through Equation (14) in the schooling decision process, but is not correlated with the error term in the wage equation by the random design. Therefore, this instrument provides exogenous variation to recover the effect of schooling on earnings. Column 2 shows the coefficient of an indicator for local college in the first-stage regression for years of schooling. As expected, the existence of a local college has a positive and significant effect on schooling outcome. Column 3 reports the second-stage estimate of return to schooling to be 4.9%, which is considerably lower than the OLS estimate in column 1 and lies within the support of the underlying true rates of returns between 2.9% and 5.9%.

In the NLSY79 sample, approximately 88% of the individuals live in a county with a two-year or four-year college. Next, I simulate an instrument based on local college that mimics the actual data available. Specifically, I simulate another sample of 100,000 individuals, and let a randomly selected 88% of them live in a county with a local college. Given the random assignment, the second instrument based on local college is also not correlated with wage errors, but is still correlated with schooling outcome. However, this instrument does not induce as much schooling variation as the first instrument. Columns 4 and 5 present two-stage least squares ($2SLS$) results. Although in the first stage, local college still has a significantly positive effect on schooling outcome, the estimated 12.2% return to schooling is upward biased and larger than the OLS estimate. This instrument that is based on local college appears to be a weak instrument.[20]

The existence of a local college is, thus far, rendered completely exogenous by the random experimental design. However, households with different socioeconomic backgrounds choose to live in different neighborhoods and counties. For instance, in the NLSY79 sample, 91% of the households in the top quartile of income live in a county with a local college, but for the bottom quartile, only 85% live near a college. All households with a college graduate mother live near a local college, whereas the number drops to 80% for households with a high school dropout mother. Given that family background is correlated with individual ability as in Equation (12), whether a person lives in a county with a local college is also likely to be correlated with her ability. To investigate the effects of these correlations, I simulate a third instrument based on local college. In particular, when simulating the decisions for 100,000 individuals, I allow the proportion living near a local college to be 48%, 50%, and 52% for skill types one, two, and three, respectively. The $2SLS$ results based on simulated data are presented in columns 6 and 7 of Table 10. When the exogeneity assumption of the instrument is violated, the estimated return to schooling becomes even more biased than the OLS estimate. In the last two columns of Table 10, I investigate whether the bias in the IV estimator can be corrected by controlling for the observed family and cognitive background. After adding the background variables, the explanatory power of local college on schooling outcome is weakened, but the estimated return to schooling in the second stage is almost identical to that in the regression without background variables.

Panel B presents regression results using all potential wages from model simulations. Similar to observations from Table 7, the bias in OLS estimates tends to be smaller when employment selection is controlled. When a strong instrument is used, the estimated return lies within the support of the true returns and is more precisely estimated (column 3). If the instrument is weak or correlated with the unobservables (column 5, 7, and 9), estimates using all potential wages are not necessarily less biased than those using observed wages

[20] Stock and Yogo (2002) define an instrument as weak if the bias of the IV estimator exceeds that of OLS by a certain threshold, for example, 10%.

because the employment selection may also be correlated with the unobservables.

Example two: School subsidy program

The second set of instruments I consider is based on a college subsidy program. A school subsidy will increase the incentive for people to attend and continue college. IV identification relies on the inclusion of a dummy variable for school subsidy in the set of variables $\mathbf{Z}$. Table 11 summarizes the estimates using the conventional OLS approach and using IVs based on college subsidy. Column 1 shows the OLS estimate of return to schooling, which is again upward biased. In columns 2 to 9, I compare estimates using simulated data from four different instrument variable designs based on school subsidy.

I first simulate schooling, employment, and wages for a sample of 100,000 individuals, in which half is kept as the control group, and the other randomly selected half is exposed to a 50% tuition subsidy program. By experimental design, the school subsidy program is correlated with the education outcome but not with the error term in wage. Column 2 shows the first-stage result. As expected, school subsidy has a positive and significant effect on educational attainment. The IV estimate of return to schooling is 2.6%, as column 3 presents. Given that the lower bound of the true return is 2.9%, this value is a downward biased estimate of the true return.

To investigate the sensitivity of the IV estimates to the size of the school subsidy, I simulate another 100,000 individuals, in which a randomly selected half is eligible for a 10% reduction in tuition. Columns 4 and 5 present the $2SLS$ results. In this case, school subsidy has a smaller, yet still significantly positive effect on schooling outcome. The estimated schooling coefficient, however, becomes upward biased and larger than the OLS estimate, similar to the weak instrument case presented in Table 10.

A school subsidy program is generally either need-based or merit-based. In both cases, eligibility is correlated with family background and unobserved ability. Therefore, in the third simulation of an instrument based on school subsidy, I let the proportion eligible for a 50% tuition subsidy to be 48%, 50%, and 52% for skill types one, two, and three, respectively. As shown in column 7, when the exogeneity assumption of the instrument is violated, the estimated returns to schooling become upward biased and slightly greater than the OLS estimate. The bias in the IV estimator remains even after the observed background variables are controlled in column 9.

The regression results in Panel B confirm the findings in Table 10. The bias in OLS estimates is smaller when all potential wages are used (column 1). Column 3 shows that when a strong instrument is used and employment selection is controlled, the IV estimate lies within the support of the true returns. On the contrary, if the instrument is weak or correlated with the unobservables (column 5, 7, and 9), estimates using all potential wages are still biased.

To summarize, the discrete choice model makes explicit the dynamic selection process that determines wages, education outcome, and work experience. The results indicate that the model is capable of accounting for the relatively low returns to schooling with high observed estimates from OLS and IV estimates. The population appears to be heterogeneous in returns to schooling (schooling coefficient) and preferences for school and for work (utility parameters). Each dimension of heterogeneity plays a role in generating the observed estimates of schooling coefficient. Ability selection is the major source of upward bias in the OLS estimates of return to schooling, whereas preferences for school and for work may determine the sign of bias for a sample with homogeneous ability. Furthermore, the dynamic employment selection is not innocuous in estimating the return to schooling.

An unbiased estimate of weighted average return to schooling may be generated by using the IV approach, but the requirements are stringent. First, the instrument must be strongly correlated with education outcome. Second, the instrument cannot have any correlation with wage errors. Finally, dynamic employment selection needs to be controlled. When all these three conditions are satisfied, I find that the IV estimator is bounded by the maximal and minimal returns to schooling in the population. If any of the conditions is violated, an IV estimate may lie outside the support of the distribution of true returns and may be greater than corresponding OLS estimate. In particular, IV estimates are very sensitive to the variations in schooling induced by the instrument. A small correlation between the instrument and the unobserved heterogeneity may result in a large bias in the estimated schooling coefficient, which cannot be easily mitigated by the addition of more control variables.

5 Concluding Remarks

This paper aims to investigate the applicability of a dynamic discrete choice model of schooling and employment in accounting for the observed OLS and IV estimates of returns to schooling in a log earnings function. I estimate a dynamic discrete choice model of endogenous schooling and employment decisions and then attempt to use simulated data to recover the (known) parameters on schooling returns. I find relatively low returns from the structural model, but OLS and IV estimates can be considerably higher. Both theoretically and empirically, I show that the dynamic model can reproduce the observed estimates of schooling returns. Analysis based on simulated data reveals that ability selection is the major source of bias in the OLS estimates of schooling returns. Although a well-designed IV estimator lies between the maximal and minimal returns to schooling in the population, the estimates are sensitive to a weak instrument, the correlation between instrument and wage errors, and to employment selection, which are common features of non-experimental data.

Aside from instrument exogeneity, this study finds that the effects of dynamic employment decision and instrument relevance are far from innocuous in estimating schooling returns.[21] Recent developments in the microeconometric theory of treatment effects and marginal returns to policies and its applications in estimating returns to schooling (Heckman and Vytlacil 2001, 2005; Carneiro, Heckman and Vytlacil 2010, 2011) are primarily set within a static environment with a strong instrument. Extension to a dynamic framework and the consideration of weak identification are important topics for future research.

[21] Using a calibrated model, Belzil and Hansen (2008) also find that the consideration of employment selection is important in the estimation of schooling returns.

References

Angrist, Joshua D., and Alan B. Krueger. 1991. Does compulsory schooling attendance affect schooling and earnings. *Quarterly Journal of Economics* 106: 979–1014.

Angrist, Joshua D., and Alan B. Krueger. (1992). Estimating the payoff to schooling using the Vietnam-era draft lottery. Working Paper no. 4067, National Bureau of Economic Research, Cambridge, MA.

Ashenfelter, Orley, and Alan B. Krueger. 1994. Estimates of the economic return to schooling from a new sample of twins. *American Economic Review* 84: 1157–1173.

Ashenfelter, Orley, and Cecilia E. Rouse. 1998. Income, schooling and ability: Evidence from a new sample of identical twins. *Quarterly Journal of Economics* 113: 253–284.

Becker, Gary S. 1975. *Human capital.* 2nd ed. Chicago: University of Chicago Press.

Belzil, Christian. 2007. The return to schooling in structural dynamic models: A survey. *European Economic Review* 51: 1059–1105.

Belzil, Christian, and Jörgen Hansen. 2002. Unobserved ability and the return to schooling. *Econometrica* 70: 575–591.

Belzil, Christian, and Jörgen Hansen. 2007. A structural analysis of the correlated random coefficient wage regression model. *Journal of Econometrics* 140: 827–848.

Belzil, Christian, and Jörgen Hansen. 2008. Calibration and IV estimation of a wage outcome equation in a dynamic environment. IZA discussion paper No. 3528.

Ben-Porath, Yoram. 1967. The production of human capital and the life cycle of earnings. *Journal of Political Economy* 75: 352–365.

Blackburn, McKinley L. and David Neumark. 1995. Are OLS estimates of the return to schooling biased downward? Another look. *Review of Economics and Statistics* 77: 217–230.

Bound, John, David A. Jaeger, and Regina M. Baker. 1995. Problems With Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable Is Weak. *Journal of the American Statistical Association* 90: 443–450.

Burdett, Kenneth and Dale T. Mortensen. 1998. Wage differentials, employer size, and unemployment. *International Economic Review* 39: 257–273.

Cameron, Stephen V. and James J. Heckman. 1998. Life cycle schooling and dynamic selection bias: Models and evidence for five cohorts of American males. *Journal of Political Economy* 106: 262–311.

Cameron, Stephen V. and James J. Heckman. 2001. The dynamics of educational attainment for black, Hispanic, and white males. *Journal of Political Economy* 109: 455–499.

Cameron, Stephen V. and Christopher Taber. 2004. Estimation of educational borrowing constraints using returns to schooling. *Journal of Political Economy* 112: 132–182.

Card, David. 1993. Using geographic variation in college proximity to estimate the return to schooling. Working Paper no. 4483, National Bureau of Economic Research, Cambridge, MA.

Card, David. 1999. The causal effect of education on earnings. In O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics*, Vol. 3, Chapter 30, Amsterdam: Elsevier Science.

Card, David. 2001. Estimating the return to schooling: Progress on some persistent econometric problems. *Econometrica* 69: 1127–1160.

Carneiro, Pedro, and James J. Heckman. 2003. Empirical estimates of the returns to schooling. Working paper. University of Chicago.

Carneiro, Pedro, James J. Heckman, and Edward Vytlacil. 2010. Evaluating marginal policy changes and the average effect of treatment for individuals at the margin. *Econometrica* 78: 377–394.

Carneiro, Pedro, James J. Heckman, and Edward Vytlacil. 2011. Estimating marginal returns to education. *American Economic Review* 101: 2754–2781.

Duflo, Esther. 2001. Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment. *American Economic Review* 91: 795–813.

Eckstein, Zvi, and Osnat Lifshitz. 2011. Dynamic female labor supply. *Econometrica*, 79: 1675–1726.

Eckstein, Zvi, and Kenneth I. Wolpin. 1989. Dynamic labour force participation of married women and endogenous work experience. *Review of Economic Studies* 56: 375–390.

Eckstein, Zvi, and Kenneth I. Wolpin. 1999. Why youths drop out of high school: The impact of preferences, opportunities, and abilities. *Econometrica* 67: 1295–1339.

Ge, Suqin. 2011. Women's college decisions: How much does marriage matter? *Journal of Labor Economics* 29: 773–818.

Griliches, Zvi. 1977. Estimating the returns to schooling: Some econometric problems. *Econometrica* 45: 1–22.

Harmon, Colm, and Ian Walker. 1995. Estimates of the Economic Return to Schooling for the United Kingdom. *American Economic Review* 85: 1278-1286.

Heckman, James J., Lance Lochner, and Petra E. Todd. 2006. Earnings functions, rates of return, and treatment effects: The Mincer equation and beyond. In E. Hanushek and F. Welch (eds.), *Handbook of the Economics of Education*, Vol. 1, Chapter 7, Amsterdam: Elsevier Science.

Heckman, James J., and Sergio Urzúa. 2010. Comparing IV with structural models: What simple IV can and cannot identify. *Journal of Econometrics* 156: 27–37.

Heckman, James J., and Edward J. Vytlacil. 2001. Policy-relevant treatment effects. *American Economic Review* 91: 107–111.

Heckman, James J., and Edward J. Vytlacil. 2005. Structural equations, treatment effects and econometric policy evaluation. *Econometrica* 73: 669–738.

Imbens, Guido W., and Joshua D. Angrist. 1994. Identification and estimation of local average treatment effects. *Econometrica* 62: 467–475.

Kane, Thomas J. and Cecilia E. Rouse. 1995. Labor market returns to two- and four-year colleges. *American Economic Review* 85: 600–614.

Katz, Lawrence, and Kevin M. Murphy. 1992. Changes in relative wages, 1963-1987: Supply and demand factors. *Quarterly Journal of Economics* 107: 35-78.

Keane, Michael P., and Kenneth I. Wolpin. 1994. The solution and estimation of discrete choice dynamic programming models by simulation and interpolation: Monte Carlo evidence. *Review of Economics and Statistics* 74: 648–672.

Keane, Michael P., and Kenneth I. Wolpin. 1997. The career decisions of young men. *Journal of Political Economy* 105: 473–522.

Keane, Michael P., and Kenneth I. Wolpin. 2001. The effect of parental transfers and borrowing constraints on educational attainment. *International Economic Review* 42: 1051-1103.

Lemieux, Thomas, and David Card. 2001. Education, earnings, and the 'Canadian G.I. Bill.' *Canadian Journal of Economics* 34: 313–344.

Mincer, Jacob. 1974. *Schooling, experience, and earnings.* New York: National Bureau of Economic Research.

Psacharopoulos, George. 1994. Returns to investment in education: A global update. *World Development* 22: 1325–1343.

Rosenzweig, Mark R. and Kenneth I. Wolpin. 2000. "Natural" natural experiments in economics. *Journal of Economic Literature* 38: 827–874.

Stock, James H. and Motohiro Yogo. 2002. Testing for Weak Instruments in Linear IV Regression. NBER technical working paper 284.

Van Der Klaauw, Wilbert. 1996. Female labour supply and marital status decisions: A life-cycle model. *Review of Economic Studies* 63: 199–235.

Wolpin, Kenneth I. 2003. Wage equations and education policy. In M. Dewatripont, L.P. Hansen and S.J. Turnovsky (eds.), *Advances in Economics and Econometrics.* Cambridge University Press.

Table 1

Comparison of OLS and IV Estimates Under Different Assumptions

	OLS	*IV*
Exogenous employment decision		
Homogeneous ability	$b_{OLS} = b$	
Heterogeneous ability, constant return	$b_{OLS} > b$	$b_{IV} = b, b_{IV} < b_{OLS}$
Heterogeneous ability and heterogeneous return	$b_{OLS} > b$	$b_{IV} \gtreqqless b, b_{IV} < b_{OLS}$
Endogenous employment decision		
Homogeneous ability	$b_{OLS} < b$	$b_{IV} < b, b_{IV} = b_{OLS}$
Heterogeneous ability	$b_{OLS} \gtreqqless b$	$b_{IV} \gtreqqless b, b_{IV} \gtreqqless b_{OLS}$

Note. b: average returns to schooling; b_{OLS}: *OLS* estimates; b_{IV}: *IV* estimates.

Table 2

Choice Proportions and Average Wages by Years After High School

Year	No. Obs	Not work		Work		Average wage
		No School	School	No School	School	
1	(487)	18.7	37.8	32.9	10.7	6.17
2	(486)	13.6	32.1	42.2	12.1	6.81
3	(485)	14.6	28.7	46.0	10.7	6.93
4	(481)	13.7	22.5	52.4	11.4	7.59
5	(478)	12.6	8.2	69.7	9.6	8.01
6	(475)	14.7	5.5	73.1	6.7	10.01
7	(472)	17.6	3.8	72.5	6.1	11.38
8	(470)	17.2	3.2	73.2	6.4	12.66
9	(469)	19.4	3.4	71.2	6.0	12.65
10	(467)	21.6	2.8	69.6	6.0	12.70

Table 3

Transition Matrix

	Choice (t)			
Choice ($t-1$)	No school, no work	School, no work	No school, work	School, work
No school, no work				
Row %	60.6	5.4	33.3	0.7
Column %	58.6	6.8	8.2	1.4
School, no work				
Row %	6.7	57.8	20.2	15.3
Column %	6.8	76.2	5.2	29.8
No school, work				
Row %	9.0	1.2	85.9	3.9
Column %	33.0	5.9	80.5	27.3
School, work				
Row %	2.9	15.4	42.8	38.9
Column %	1.6	11.1	6.1	41.5

Table 4
Estimates of the Wage Equation and Preference Heterogeneity

	Parameter	Estimate	(Standard Error)
Women's Post-School Wage Function			
Years of schooling skill Type One	β_1^1	0.029	(0.0026)
Years of schooling skill Type Two	β_1^2	0.043	(0.0021)
Years of schooling skill Type Three	β_1^3	0.059	(0.0020)
Years of experience	β_2	0.091	(0.0072)
Experience squared	β_3	-0.002	(0.0008)
Constant	β_0	1.367	(0.0313)
True error standard deviation	σ_w	0.370	(0.0018)
Measurement error standard deviation	σ_u	0.199	(0.0010)
Consumption Value of College			
skill Type One	v_1^1	-1.415e+5	(8.949e+3)
skill Type Two	v_1^2	8.396e+4	(4.477e+3)
skill Type Three	v_1^3	9.326e+4	(5.008e+3)
Value of Non-employment			
skill Type One	v_2^1	-1.842e+4	(3.437e+3)
skill Type Two	v_2^2	3.206e+3	(3.257e+3)
skill Type Three	v_2^3	1.606e+4	(3.673e+3)
Cost of returning school			
skill Type One	v_3^1	-5.100e+4	(2.279e+4)
skill Type Two	v_3^2	-1.254e+5	(4.993e+3)
skill Type Three	v_3^3	-2.081e+4	(8.679e+2)
Consumption Value of College			
Work, skill Type One	v_4^1	-3.036e+4	(1.268e+4)
Work, skill Type Two	v_4^2	-7.275e+4	(6.673e+3)
Work, skill Type Three	v_4^3	-1.250e+4	(6.394e+3)

Table 5

Fit of the Transitions Rates

From\To	No school, no work	School, no work	No school, work	School, work
No school, no work	29.1 (60.6)	8.2 (5.4)	57.9 (33.3)	4.8 (0.7)
School, no work	12.1 (6.7)	52.0 (57.8)	19.7 (20.2)	16.2 (15.3)
No school, work	14.1 (9.0)	1.1 (1.2)	81.8 (85.9)	3.0 (3.9)
School, work	12.0 (2.9)	25.4 (15.4)	33.4 (42.8)	29.2 (38.9)

Note. Data moments are in parentheses.

Table 6

Descriptive Statistics of the Simulated Sample

Variable	Mean	Std. Dev.
Family and cognitive background		
Mother's schooling	12.29	2.02
Father's schooling	12.60	2.96
Number of siblings	2.78	1.78
Proportion living with both parents at 14	0.86	0.35
Net family income	65,482	32,981
AFQT percentile score	54.05	23.81
Age at high school graduation	17.89	0.43
Schooling and labor market outcome		
Highest grade completed (HGC)	14.23	2.21
Years of experience	6.09	1.96
Observed hourly wage	9.70	5.19
Potential hourly wage	9.38	4.62

Table 7

OLS Wage Equation Estimates

Independent Variables	Observed wages		True wages		All potential wages	
	(1)	(2)	(3)	(4)	(5)	(6)
Years of Schooling (β_1)	0.103	0.101	0.103	0.101	0.096	0.091
	(3.4e-4)	(4.3e-4)	(3.1e-4)	(3.9e-4)	(2.7e-4)	(3.2e-4)
Years of Experience (β_2)	0.084	0.085	0.085	0.085	0.071	0.074
	(7.7e-4)	(7.7e-4)	(6.9e-4)	(6.8e-4)	(5.0e-4)	(5.1e-4)
Experience^2 (β_3)	-0.001	-0.001	-0.001	-0.001	-2.8e-4	-5.0e-4
	(9.6e-5)	(9.6e-5)	(8.5e-5)	(8.5e-5)	(6.7e-5)	(6.8e-5)
$AFQT$ Score		3.5e-4		3.4e-4		6.6e-4
		(3.2e-5)		(2.9e-5)		(2.5e-5)
Mother's Schooling		4.4e-4		3.5e-4		0.001
		(3.9e-4)		(3.6e-4)		(3.0e-4)
Father's Schooling		0.001		0.001		0.002
		(2.6e-4)		(2.4e-4)		(2.1e-4)
Family Income		0.001		0.001		0.001
(in 10 thousands)		(3.4e-4)		(3.0e-4)		(2.9e-4)
Number of Siblings		4.9e-5		-6.6e-5		1.5e-6
		(2.2e-4)		(2.0e-4)		(1.7e-4)
Living with Parents		-0.030		-0.030		-0.038
		(0.002)		(0.002)		(0.002)
Constant	0.515	0.531	0.515	0.533	0.656	0.672
	(0.005)	(0.006)	(0.004)	(0.006)	(0.003)	(0.005)
No. of observations	604,858	604,858	604,858	604,858	1,000,000	1,000,000
R-squared	0.248	0.249	0.295	0.295	0.272	0.274

Table 8

OLS Wage Equation Estimates: Control for Types

Independent Variables	True value	Estimates using	
		Observed wages	All potential wages
Years of Schooling			
type one	0.029	0.016	0.028
		(0.001)	(4.2e-4)
type two	0.043	0.034	0.042
		(0.001)	(3.3e-4)
type three	0.059	0.055	0.059
		(0.001)	(3.4e-4)
Years of Experience	0.091	0.088	0.092
		(0.001)	(5.0e-4)
Experience^2	-0.002	-0.002	-0.002
		(9.4e-5)	(6.6e-5)
Constant	1.367	1.536	1.373
		(0.010)	(0.005)

Table 9

OLS Wage Equation Estimates on Homogeneous Samples

	Type One Wage		Type Two Wage		Type Three Wage	
	Observed	All	Observed	All	Observed	All
Years of Schooling	0.031	0.029	0.037	0.043	0.050	0.059
	(0.002)	(0.002)	(5.0e-4)	(2.7e-4)	(5.9e-4)	(3.0e-4)
Years of Experience	0.091	0.091	0.077	0.091	0.093	0.091
	(6.2e-4)	(4.7e-4)	(0.001)	(6.2e-4)	(0.001)	(5.9e-4)
Experience^2	-0.002	-0.002	-1.3e-4	-0.002	-0.002	-0.002
	(7.4e-5)	(5.8e-5)	(1.6e-4)	(1.1e-4)	(1.4e-4)	(7.5e-5)
Constant	1.347	1.370	1.506	1.365	1.607	1.365
	(0.028)	(0.023)	(0.008)	(0.004)	(0.008)	(0.004)
No. of observations	815,685	1,000,000	463,884	1,000,000	341,019	1,000,000
R-squared	0.177	0.223	0.108	0.212	0.232	0.329

Note. Regressions are based on the simulation of 100,000 individuals of each type for 10 years.

TABLE 10

IV ESTIMATES: LOCAL COLLEGE

	OLS	IV		Weak IV		Correlated IV		Correlated IV+control	
	$\ln wage$	$School$	$\ln wage$	$School$	$\ln wage$	$School$	$\ln wage$	$School$	$\ln wage$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
A. Use observed wages									
local college		0.034		0.046		0.107		0.088	
		(0.005)		(0.007)		(0.005)		(0.004)	
years of schooling	0.104		0.049		0.122		0.119		0.120
	(3.5e-4)		(0.040)		(0.041)		(0.012)		(0.014)
B. Use all potential wages									
local college		0.047		0.057		0.110		0.089	
		(0.003)		(0.005)		(0.003)		(0.003)	
years of schooling	0.096		0.052		0.073		0.123		0.123
	(2.7e-4)		(0.024)		(0.027)		(0.009)		(0.011)

Table 11

IV Estimates: School Subsidy

	OLS	*IV*		Weak *IV*		Correlated *IV*		Correlated *IV*+control	
	ln *wage*	*School*	ln *wage*	*School*	ln *wage*	*School*	ln *wage*	*School*	ln *wage*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
A. Use observed wages									
school subsidy		0.042		0.013		0.117		0.102	
		(0.005)		(0.005)		(0.005)		(0.004)	
years of schooling	0.103		0.026		0.136		0.104		0.103
	(3.4e-4)		(0.036)		(0.099)		(0.011)		(0.012)
B. Use all potential wages									
school subsidy		0.067		0.015		0.133		0.112	
		(0.004)		(0.003)		(0.004)		(0.003)	
years of schooling	0.095		0.050		0.072		0.109		0.108
	(2.6e-4)		(0.017)		(0.068)		(0.007)		(0.009)

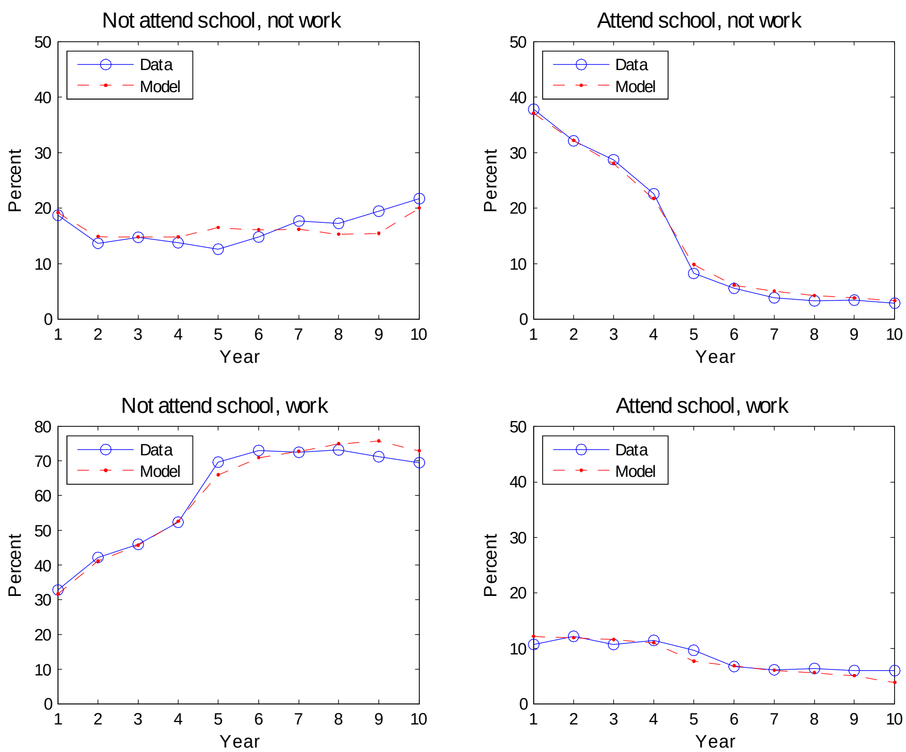

Figure 1: Fit of Choice Proportions

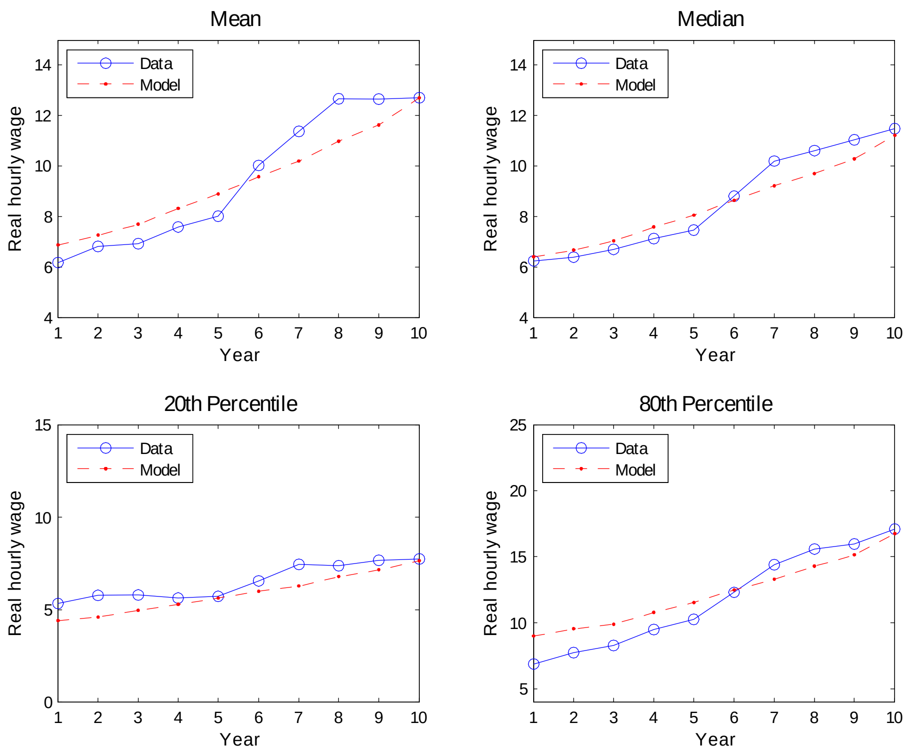

Figure 2: Actual and Predicted Wages