Returns to Education and Occupation Choices

Jinwen Xu *

Department of Economics, Concordia University

November 1, 2014

Abstract

This paper examines how returns to education are related to occupation choices. Specifically, I investigate the returns to attending a two-year college and a four-year college and how these returns to education differ from a blue-collar occupation to a white-collar occupation. To address the endogenous education and occupation choices, I use a finite mixture model. I show how the finite mixture model can be nonparametrically identified by using test scores and variations in wages across occupations over time. Using data taken from the National Longitudinal Survey of Youth (NLSY) 1979, I estimate a parametrically specified model and find that returns to education are occupation specific. Specifically, a two-year college attendance enhances blue-collar wages by 24% and white-collar wages by 17% while a four-year college attendance increases blue-collar wages by 23% and white-collar wages by 30%.

Keywords: Returns to education, Education choice, Occupation choice, Comparative advantage

JEL Classification: J2, J3

*Address: 1455 de Maisonneuve Blvd. West, H 1155, Montreal, Quebec, Canada H3G 1M8; E-mail: jinwen.j.x@gmail.com

1 Introduction

The association between education and earnings is perhaps the most well-documented and studied subject in social science. Much recent work by economists has investigated the extent to which this correlation is causal in nature (see Card, 2001, for a recent survey of this literature, also see Heckman et al., 2006a). However, how returns to education are related to the choice of occupations has received less attention.

It is plausible that returns to education are occupation specific. Returns to education are found to be lower in secondary sector occupations (Blaug, 1985, Dickens and Lang, 1985) and in occupations which do not require the education that one obtains(Duncan and Homan, 1981, Sicherman, 1991). In this paper, I examine how returns to attending a two-year college and a four-year college differ from those of a blue-collar occupation to those of a white-collar occupation, using the National Longitudinal Survey of Youth (NLSY) 1979. Intuitively, the wage premium for high school graduates attending a two-year college may be higher in a blue-collar occupation such as that of a machinist than it may be in a white-collar occupation such as that of a manager while the wage premium for high school graduates attending a four-year college may be higher in a white-collar occupation than it may be in a blue-collar occupation.

The main complication of estimating the occupation-specific returns to education comes from the endogenous education and occupation choice. As in Roys model (Roy, 1951), individuals are endowed with different abilities to work in a blue-collar occupation or a white-collar occupation. They tend to work in the occupation in which they have a comparative advantage. Moreover, occupation abilities can also influence the education choice. For example, individuals who know that they are more likely to work in a white-collar occupation are more likely to attend a four-year college, which would increase the white-collar wages they would earn more so than attending a two-year college would. In addition, individuals vary in their education psychic costs. Those with lower education psychic costs may obtain more education than may those with higher education psychic costs (Willis and Rosen, 1979; Willis, 1986; Carneiro et al., 2003). While the occupation abilities and the education psychic costs are known to individuals making education and occupation decisions, these abilities and costs are unobserved by the econometrician. In the presence of self-selection in both education and occupation, the Ordinary Least Squares (OLS) estimates of occupation-specific returns to education are biased. One traditional way of dealing with the endogeneity issue in the returns to education literature is to use compelling instruments for education such as institutional rules or natural experiments (see Card, 2001, for a survey of papers using IV approach in this literature). However, the standard IV approach is hard to implement here because it is difficult to find good instruments for both education and occupation choices.

I address the issue of endogeneity in education and occupation by explicitly modelling the sequential education and occupation choices. The unobserved occupation abilities and education psychic costs are specified with a flexible multinomial distribution in a finite mixture model. Departing from previous papers that use a finite mixture model to tackle the endogeneity issue in the education literature, I achieve nonparametric identification of the finite mixture model without imposing parametric assumptions on the joint distribution of wages, education, and occupation choices. Based on Kasahara and Shimotsu (2009) and Kasahara and Shimotsu (2012), I rigorously show how to nonparametrically identify the occupation abilities using the variations in wages across occupations over time. Since the information from the panel data alone is not enough to identify the unobserved education psychic costs, I bring in additional data. Specifically, I use scores from four tests (math skills, verbal skills, coding speed, and mechanic comprehension) conducted by the Armed Force Vocational Aptitude Battery(ASVAB), together with the Rotter Locus of Control test score and the Rosenberg Self-Esteem Scale. I show that conditional on occupation abilities and education psychic costs the education psychic costs can be nonparametrically identified under the assumption that the test scores do not directly affect wages, education, or occupation choices. My identification strategy allows the unobserved occupation abilities and education psychic costs to be freely correlated. Carneiro et al. (2003), Hansen et al. (2004), Heckman et al. (2006b), and Cunha and Heckman (2008) also use test scores to identify their mixture model. However, they assume the unobserved variables to be mutually independent. Cunha et al. (2010) relax the strong independence assumption, but their identification replies on the assumption that distributions are bounded complete. My identification strategy does not require this strong rank condition.

While I show that the finite mixture model can be nonparametrically identified, estimating the high-dimensional model nonparametrically is nearly impossible given the relatively small sample size of NLSY 1979. Therefore, I impose some parametric forms in wage, education, occupation, and test scores to facilitate the estimation. I find that attendance of a two-year college enhances blue-collar wages by 24% and white-collar wages by 17%. Therefore, attendance of a two-year college helps accumulate more blue-collar skills than it does white-collar skills. The reverse holds true for attendance of a four-year college, which increases blue-collar wages by 23% and white-collar wages by 30%.

This paper is the first to quantify the occupation-specific returns to attending a two-year and a four-year college. Although many papers have estimated returns to a two-year college and a four-year college (Kane and Rouse, 1995; Grubb, 1997; Light and Strayer, 2004; Marcotte et al., 2005), these papers assumed that returns to education are homogeneous across occupations. The occupation-specific returns to education suggest that analyzing the potential impact of an education policy, such as tuition subsidy, requires consideration of individuals possible occupation choices when these individuals have finished school because returns to education depend on their subsequent occupation choices.

Moreover, this paper helps us understand the choice made between attendance of a two-year and that of a four-year college. I find that individuals make their post-secondary education choices based on both occupation abilities and education psychic costs. The idea that individuals invest in education based on their occupation abilities was first raised in a seminal paper by Willis and Rosen (1979). Willis and Rosen studied the choice made by high school graduates between entering the labour market or attending a college; they suggest that individuals who are more suitable to the college labour market are more likely to attend a college. Keane and Wolpin (1997) extend Willis and Rosen (1979) by taking into account the sequential choices of education and occupation. They studied how individuals with different occupation abilities make year-by-year decisions as to whether to further their education. My paper departs from that of Keane and Wolpin by bringing additional data, the test scores, to achieve nonparametric identification of the education psychic costs. I find that the education psychic costs play an important role in post-secondary decisions. This is consistent with the findings in Carneiro et al. (2003). Carneiro et al. (2003) extend the model of Willis and Rosen (1979) to account for the education psychic costs and use the ASVAB scores to identify the education psychic costs. They find that individuals decide whether to attend college or not taking into account education psychic costs. In addition, I show that without considering the selection based on the unobserved education psychic costs, the returns to attending a two-year college are biased upward.

The rest of the paper proceeds as follows. Section 2 describes the data. Section 3 discusses the empirical specifications. Section 4 shows the nonparametric identification of the finite mixture model. Section 5 reports the empirical results, and Section 6 concludes.

2 Data

This paper uses data taken from the NLSY79. The NLSY79 is a U.S. national survey of 12686 young men and women who were 14-22 years old in 1979. It consists of a core random sample of civilian youths, a supplemental sample of minority and economically disadvantaged youths and a sample of youths in the military. The analysis is based on the 2439 male respondents in the core random sample. The individuals were interviewed annually through 1994 and are currently interviewed on a biennial basis. I use the observations from 1979 to 1994.

The NLSY79 collects information on individuals' education attainment and the type of post-secondary education individuals in which were enrolled. I assign individuals to three educational categories: high school graduates, two-year college attendants and four-year college attendants.[1] High school graduates are those who are reported to have completed at least 12 years of education and have never attended either a two-year college or a four-year college. Two-year college attendants are those who are reported to have enrolled in a two-year college and have never attended a four-year college. Four-year college attendants are the ones who are reported to have enrolled in a four-year college. I distinguish two-year college and four-year college education because a two-year college provides more technical and vocational programs while a four-year college offers more academic and professional programs.

The NLSY79 asks individuals about their occupations and the associated hourly payment in each survey year. I assign individuals to a blue-collar occupation and a white-collar occupation[2] according to the occupation they work the most during the survey year based on one-digit census codes. Blue-collar occupations are (1) craftsmen, foremen, and kindred; (2)operatives and kindred; (3) laborers, except farm; (4) farm laborers and foremen; and (5) service workers. White-collar occupations are (1) professional, technical, and kindred; (2) managers, officials, and proprietors; (3) sales workers; (4) formers and farm managers; and (5) clerical and kindred.

One advantage of the NLSY79 is that many of the respondents were in school when they were first interviewed. Therefore, information about their first jobs are available. Such information about initial conditions is especially useful because it is important to take into account the persistent shocks in wages and occupation choices as pointed out by Hoffmann (2011).

To identify the individual unobserved occupation abilities and education psychic costs,

[1] Although the main analysis of this paper is based on these three education groups, I examine the occupation-specific returns to a bachelor's degree because usually college graduates earn more than college dropouts (Jaeger and Page, 1996). However, I do not investigate the occupation-specific returns to an associate degree because the sample size of associate degree earners are too small to give any reasonable estimates.

[2] Although a finer aggregation is possible, I focus on two occupation categories to emphasize the importance of the role of occupational choices in returns to education.

I use ASVAB, which was administrated in 1979, to construct four test scores: math skill, verbal skill, coding speed, and mechanic comprehension. The higher scores indicates higher skills. In addition, I use the Rotter Locus of Control Scale, which was administered in 1979, and the Rosenberg Self-Esteem Scale ,which was administered in 1980. The Rotter Locus of Control Scale measures whether individuals believe that events in their life derive primarily from their own actions. It is normalized to the case that a higher score indicates higher degree of control individuals feel they possess over their life. The Rosenberg Self-Esteem Scale measures perceptions of self worth. A higher score indicates higher self-esteem.[3]

Table 1 presents the sample summary statistics by the three education groups: high school graduates, two-year college attendants, and four-year college attendants.[4] The sample consists of 934 individuals, of which 34% are high school graduates, 17.5% are two-year college attendants, and 48.5% are 4-year college attendants. On average, the high school graduates complete 11.9 years of schooling. The two-year college attendants finish 13.1 years of school. The complete years of schooling of the two-year attendants suggests that a large fraction of the two-year attendants do not graduate[5]. The four-year college attendants complete 15.7 years of schooling, which suggests that a large proportion of the four-year college attendants obtain a bachelor's degree[6]. The comparison of the fraction of individuals working in a white-collar occupation as their first jobs[7] across the three education groups suggests that the probability of the initial job in a white-collar occupation increases with education: around 11% of the high school graduates, 26.4% of the two-year college attendants, and 64% of the four-year college attendants work initially in a white-collar occupation. The average wages associated with the first jobs as presented in table 1 suggest that the higher the education, the higher the wages: on average, the high school graduates earn $9.88, the two-year college attendants earn $10.80,

[3]All measures are standardize to mean zero and variance one.

[4]Table A1 gives the summary statistics for two-year college dropouts, those with an associate degree, four-year college dropouts, and those with a bachelor's degree.

[5]Around 75% of the two-year attendants do not have an associated degree.

[6]Around 70% of the four-year college attendants obtain a bachelor's degree.

[7]I look at individuals' first jobs to get rid of the impact of work experience on the probability of working in a white-collar occupation.

and the four-year college attendants earn \$13.10. Further, I look at the blue-collar wages and white-collar wages associated with individuals' first jobs. For those who initially work in a white-collar occupation, I find that higher education are associated with higher wages: the high school graduates earn around \$8.41, the two-year attendants earn around \$10.47, and the four-year attendants earn around \$14.80. However, the relationship between wages and education is different for those who initially work in a blue-collar occupation: the two-year college attendants earn the most among the three education groups, and the high school graduates and the four-year college attendants earn almost the same. The average blue-collar wages of the high school graduates are \$10.06, those of the two-year college attendants are \$10.91, and those of the four-year college attendants are \$10.09. Table 1 also shows that the three education groups have quite different family background. Individuals whose parents have more education, who have fewer siblings, grew up in a two-parent family, and live in an urban area at age 14 tend to obtain more education.

Table 2 presents the average test scores of the six tests across education groups and occupation groups. Table 2a shows that individuals who initially work in a white-collar occupation perform better than those who initially work in a blue-collar occupation in all the six test scores, and therefore, the six test scores may be informative about individuals' occupation abilities. Table 2b shows the six test scores increase with education. Further, table 2c and table 2d show that the six test scores increase with education when conditional on initial occupations. The positive correlation between education attainment and the test scores suggest that the six test scores may be informative about individuals' education psychic costs.

3 Empirical Specification

In this section I specify the wage regression in which returns to education are occupation-specific, explicitly model how individuals make their subsequent postsecondary education choices and occupation choices based on their unobserved occupation abilities and education psychic costs, and present the test scores regression specification, which is essential for the identification of the finite mixture model.

To control for the selection in education and occupation, I specify the joint distribution of occupation abilities and education psychic costs by a multinomial distribution in a finite mixture model. A finite mixture model assumes that the overall population consists of M types of people. Each type shares the same occupation abilities and education psychic costs, and different types are different in occupation abilities and/or education psychic costs. I assume that the unobserved types affect the intercepts of the wage regression, the test scores regression, and the expectations on utility in the choice of postsecondary education and occupations. The superscript m in the following equations represents the mth type-specific parameters. The finite mixture model is discussed in more detail in Section 3.2.

3.1 A Model for Postsecondary Education, Occupation Choice and Wages

3.1.1 The Model for Wages

Different from the conventional Mincer-type wage specification, I allow returns to attending a two-year and a four-year college to depend on the occupation choice. The log wage, W_{it}, for individual i at time t is as follows,

$$W_{it} = \alpha^m_{W,1} + \alpha^m_{W,2} O_{it} + \beta_1 2\mathrm{YR}_i + \beta_2 4\mathrm{YR}_i + \beta_3 2\mathrm{YR}_i O_{it} + \beta_4 4\mathrm{YR}_i O_{it} + X'_{it}\beta_5 + \varepsilon_{W,it}, \quad (1)$$

where $2\mathrm{YR}_i$ is a dummy variable which equals 1 if individual i is a two-year college attendant, $4\mathrm{YR}_i$ is a dummy variable which equals 1 if individual i is a four-year college attendant, and O_{it} is a dummy variable which equals 1 if individual i works in a white-collar occupation at time t. The occupation-specific work experience and its squared terms are collected into X_{it}. Since different occupations reward the occupation-specific work experience differently, X_{it} also includes the interaction terms of the occupation-specific work experience and the occupation dummy variable O_{it}, and the interaction terms of the occupation-specific work experience squared and O_{it}.

The returns to attending a two-year college and a four-year college in a blue-collar occupation are represented by β_1 and β_2 respectively, and the returns to attending a two-year college and a four-year college in a white-collar occupation are denoted by $\beta_1 + \beta_3$ and $\beta_2 + \beta_4$ respectively. Since a two-year college focuses on technical and vocational programs while a four-year college provides academic and vocational programs, we would expect the returns to attending a two-year college to be higher in a blue-collar occupation than a white-collar occupation, i.e. $\beta_3 < 0$, and the returns to attending a four-year college to be higher in a white-collar occupation than a blue-collar occupation, i.e. $\beta_4 > 0$.

The relationship between wages and innate occupation abilities are captured by $\alpha^m_{W,1}$ and $\alpha^m_{W,2}$, which are specific to type m. A large value of $\alpha^m_{W,1}$ means type m has a high blue-collar ability, and a large value of $\alpha^m_{W,1} + \alpha^m_{W,2}$ implies type m has a high white-collar ability. In other words, if type m has a comparative advantage in a white-collar occupation than a blue-collar occupation, we would expect $\alpha^m_{W,2} > 0$.

I assume that productivity shocks $\varepsilon_{W,it}$ follow a first-order Markov process[8]:

$$\varepsilon_{W,it} = \rho \varepsilon_{W,it-1} + \zeta_{it},$$

where $\varepsilon_{W,i1} \overset{iid}{\sim} N(0, \sigma_{W,1})$ and $\zeta_{it} | \varepsilon_{W,it-1} \overset{iid}{\sim} N(0, \sigma_{W,2})$.

[8] I assume the same productivity shock for a blue-collar and a white-collar occupation. It is because occupation choice in current period is influenced by current wages in blue- and white-collar occupations. The current occupation-specific wages depend on the blue- and white-collar productivity shocks in the last period. However, we only observe the wage associated with the last period occupation an individual worked in. The wage associated with the other occupation is unobserved. So if we consider occupation-specific productivity shocks, we have to integrate out the unobserved productivity shock associated the other occupation. This significantly increase the computation burden.

3.1.2 The Model for Occupation Choices

In each period, individuals choose to work in either in a blue-collar or a white-collar occupation to maximize life-time income. Let $I_{O,it}$ denote the latent utility associated with a white-collar occupation relative to a blue-collar occupation at time t:

$$I_{O,it} = \alpha_O^m + \lambda_1 2\text{YR}_i + \lambda_2 4\text{YR}_i + \lambda_3 O_{it-1} + X_{it}'\lambda_4 + \varepsilon_{O,it}, \tag{2}$$

where $\varepsilon_{O,it} \overset{iid}{\sim} N(0,1)$. Since the latent utility, I_{it}, depends on wages, all the regressors in Equation (1) are included. In addition, the occupation choice at time $t-1$ may affect the occupation choice at time t because job switching costs may prevent individuals from moving from one occupation to another. Such a relationship between the occupation choices at time $t-1$ and time t are captured by the dummy variable, O_{it-1}, which equals to 1 if the job at time $t-1$ is a white-collar occupation.

The type-specific intercept, α_O^m, reflects that the latent utility, $I_{O,it}$, depends on occupation abilities. In other words, holding everything else the same, an individual with a comparative advantage in a white-collar occupation is more likely to work in a white-collar occupation than an individual with a comparative advantage in a blue-collar occupation. As illustrated in figure 1, occupation abilities drive both wages and occupation choices. Therefore, the occupation choice in Equation (1) is endogenous and OLS estimates of occupation-specific returns to education are biased in general.

3.1.3 The Model for The Education Choice

A high school graduate faces three options: attending a two-year college, attending a four-year college, and entering the labour market without pursuing more education. She makes the postsecondary education decision to maximize the life-time utility. Let $I_{S,ij}$ [9]represent the net benefit associated with education level j ($j \in \{1,2,3\}$) relative to the benefit associated

[9] $I_{S,i1}$ is normalized to 0.

with education level 1:

$$I_{S,ij} = \begin{cases} \varepsilon_{S,ij} & \text{if } j = 1 \\ \alpha^m_{S,j} + Z'_{S,i}\delta_j + \varepsilon_{S,ij} & \text{if } j = 2, 3 \end{cases} \tag{3}$$

where $\{\varepsilon_{S,ij}\}_{j=1}^{3}$ are mutually independent and follows type I extreme value distribution while $Z_{S,i}$ includes family background variables. The intercept, $\alpha^m_S = (\alpha_{S,2}, \alpha_{S,3})'$, is different across types. The reasons are twofold. First, future occupations and wages depend on occupation abilities. For instance, an individual with a comparative advantage in a white-collar occupation would expect herself to be more likely to work in a white-collar occupation and tend to attend a four-year college because a four-year college helps accumulating more white-collar skills than blue-collar skills. Second, education psychic costs also play an important role. For example, an individual with a comparative advantage in a white-collar occupation may choose to attend a two-year college rather than a four-year college if her psychic costs to attend a four-year college are high, although a four-year college enhances white-collar skills more than a two-year college.

As shown in figure 1, occupation abilities are related to both wages and the postsecondary education choice. Education psychic costs, which affect the education choice, may be correlated with occupation abilities and cause the correlation between wages and the education choices as well. Hence, the education dummy variables are endogenous in Equation (1) and the OLS estimates of the occupation-specific returns to education are biased in general.

To sum up, the education dummy variables and the occupation choice in Equation (1) are endogenous because the unobserved types connects wages, occupation choices, and the postsecondary education choice. I address the endogeneity issue using a finite mixture model in which the distribution of types are specified by a flexible multinomial distribution. Since same types of individuals have the same occupation abilities and education psychic costs, the variations in education and occupation choices within type, holding the observables constant, are purely random. Once the finite mixture model is nonparametrically identified, we can get unbiased and consistent estimates of the occupation-specific returns to attending a two-year and a four-year college.

3.1.4 The Model for The Six Test Scores

As I will discuss in details in Section 4, To achieve nonparametric identification of the finite mixture model, I bring in additional information. Specifically, I use four test scores conducted from ASVAB. They are tests for math skills, verbal skills, coding speed, and mechanic comprehension. I also use the Rotter Locus of Control, and the Rosenberg Self-Esteem Scale.

In the following specification for test scores, I take into account the possibility that the test scores are influenced by the education level at the date of the tests. Since the tests were administered to all respondents in the sample in year 1979 and 1980, when they were between 14 and 22 years of age and many had finished their schooling, the tests may not be fully informative about the occupation abilities and education psychic costs (Hansen et al., 2004;Heckman et al., 2006b). Let $Q_{i,r}$ denote the test score in test r:

$$Q_{ir} = \alpha^m_{Q,r} + \theta_{r,1} 2\mathrm{YR}_{ir} + \theta_{r,2} 4\mathrm{YR}_{ir} + Z'_{i,r}\theta_{r,3} + \varepsilon_{Q,ir}, \ \ for \ r = 1, \dots, 6, \tag{4}$$

where $2\mathrm{YR}_{ir}$ is a dummy variable, which equals 1 if individual i was a two-year college attendant at the time test r was administrated, and $4\mathrm{YR}_{ir}$ is a dummy variable, which equals 1 if individual i was a four-year college attendant at the time test r was administrated. Other observables, which influence the test score r, such as family background variables and the age when test r was administrated, are collected in $Z_{i,r}$.

The intercept $\alpha^m_{Q,r}$ is subpopulation-specific, because the test scores reflect the occupation abilities and education psychic costs. For example, mechanic comprehension is important to a blue-collar occupation. The Rotter Locus of Control which measures people's belief in their ability to control life may be important to a management job. Math and verbal skills can reflect education psychic costs.

I assume that the test scores are mutually independent conditional on occupation abilities, education psychic costs, and the observables, i.e. $\varepsilon_{Q,ir} \perp\!\!\!\perp \varepsilon_{Q,ir'}$ for $r \neq r'$ and $\varepsilon_{Q,ir} \sim N(0, \sigma_{Q,r})$ for $r \in \{1, \ldots, 6\}$. Further, I assume that the test scores do not *directly* affect wages, occupation and education choices once conditional on occupation abilities, education psychic costs, and the observables, i.e. $\varepsilon_{Q,ir} \perp\!\!\!\perp \varepsilon_{W,it}$, $\varepsilon_{Q,ir} \perp\!\!\!\perp \varepsilon_{O,it}$, and $\varepsilon_{Q,ir} \perp\!\!\!\perp \varepsilon_{S,ij}$. These two assumptions are the key for the nonparametric identification of the finite mixture model, which will be discussed in Section 4.

3.2 A Finite Mixture Model

In the finite mixture model, the conditional joint distribution of wages $\{W_{it}\}_{t=1}^{T}$, occupations $\{O_{it}\}_{t=1}^{T}$, education S_i, and tests $\{Q_{ir}\}_{r=1}^{6}$ in the overall population is a weighted average of type-specific conditional joint distribution. The weight π^m is the proportion of type m. Formally,

$$
\begin{aligned}
& f(\{W_{it}, O_{it}\}_{t=1}^{T}, S_i, \{Q_{ir}\}_{r=1}^{6} | \{X_{it}\}_{t=1}^{T}, Z_{S,i}, \{Z_{ir}\}_{r=1}^{6}) \qquad (5) \\
= & \sum_{m=1}^{M} \pi^m f^m(\{W_{it}, O_{it}\}_{t=1}^{T}, S_i, \{Q_{ir}\}_{r=1}^{6} | \{X_{it}\}_{t=1}^{T}, Z_{S,i}, \{Z_{ir}\}_{r=1}^{6}),
\end{aligned}
$$

where $\{X_{it}\}_{t=1}^{T}$, $Z_{S,i}$, and $\{Z_{ir}\}_{r=1}^{6}$ are observables in Equation (1), Equation (2), Equation (3), and Equation (4). With the assumptions (i) test scores do not directly affect wages, occupations, and education conditional on type, (ii) the error terms in test scores are mutually independent, (iii) the error terms in wage follows a first order Markov process, (iv) the occupation choice is only affected by the previous occupation, not the whole occupation history, and (v) the regressors and the error terms in Equation (1), Equation (2), Equation (3), and Equation (4) are independent, I simplify the type-specific conditional joint distribution of wages, occupations, education, and test scores, and express the population conditional joint distribution as follows[10]:

$$
\begin{aligned}
& f(\{W_{it}, O_{it}\}_{t=1}^{T}, S_i, \{Q_{ir}\}_{r=1}^{6} | \{X_{it}\}_{t=1}^{T}, Z_{S,i}, \{Z_{ir}\}_{r=1}^{6}) \\
= & \sum_{m=1}^{M} \pi^m f^m(W_{i1}|O_{i1}, S_i) \prod_{t=2}^{T} f^m(W_{it}|O_{it}, S_i, X_{it}, W_{it-1}, O_{it-1}, X_{it-1}) \\
& \times f^m(O_{i1}|S_i) \prod_{t=2}^{T} f^m(O_{it}|O_{it-1}, S_i, X_{it}) f^m(S_i|Z_{S,i}) \prod_{r=1}^{6} f^m(Q_{ir}|Z_{ir}).
\end{aligned} \tag{6}
$$

In Section 4, I rigorously show how this finite mixture model is nonparametrically identified, i.e. how to recover the unknowns, which are on the right hand side of Equation (6), from the observed restriction, which is on the left hand side of Equation (6). Many papers that use a finite mixture model in the returns to education literature do not show the nonparametric identification. In other words, their finite mixture models may rely on restrictive parametric assumptions, which can lead to biased estimates of occupation-specific returns.

Once the nonparametric identification of the finite mixture model is established, I use the Maximum Likelihood Estimator (MLE) to estimate the occupation-specific returns to attending a two-year and a four-year college. Although the finite mixture model can be nonparametrically identified, estimating a high dimensional nonparametric statistical model requires very heavy computation and is nearly impossible given the relatively small sample size of NLSY 1979. Therefore, I estimate a statistical model with parametric assumptions in Section 3.1. Let $Y_i = (\{W_{it}, O_{it}, X_{it}\}_{t=1}^{T}, S_i, Z_i, \{Q_{ir}, Z_{ir}\}_{r=1}^{6})$. The log-likelihood contribution for a particular individual is as follow:

$$
\begin{aligned}
& L(Y_i; \alpha_W, \alpha_O, \alpha_S, \alpha_R, \beta, \lambda, \delta, \theta, \sigma_W, \sigma_Q, \rho) \\
= & log(\sum_{m=1}^{M} \pi^m \mathcal{L}_W^m(Y_i; \alpha_W, \beta, \sigma_W, \rho) \mathcal{L}_O^m(Y_i; \alpha_O, \lambda) \mathcal{L}_S^m(Y_i; \alpha_S, \delta) \mathcal{L}_Q^m(Y_i; \alpha_Q, \theta, \sigma_Q)).
\end{aligned} \tag{7}
$$

where $\alpha_W = \{\alpha_{W,1}^m, \alpha_{W,2}^m\}_{m=1}^{M}$, $\alpha_O = \{\alpha_O^m\}_{m=1}^{M}$, $\alpha_S = \{\{\alpha_{S,j}\}_{j=2}^{3}\}_{m=1}^{M}$, $\alpha_R = \{\{\alpha_{Q,r}\}_{r=1}^{6}\}_{m=1}^{M}$,

[10]Please refer to Appendix B for more details about how these assumptions simplify the type-specific conditional joint distribution of wages, occupations, education, and test scores.

$\beta = \{\beta_1, \beta_2, \beta_3, \beta_4, \beta_5\}$, $\lambda = \{\lambda_1, \lambda_2, \lambda_3, \lambda_4\}$, $\sigma_W = \{\sigma_{W,1}, \sigma_{W,2}\}$, and $\sigma_Q = \{\sigma_{Q,r}\}_{r=1}^{6}$.

Note the likelihood contribution of a particular individual who belongs to subpopulation m consists of four pieces:

$\mathcal{L}_W^m(Y_i; \alpha_W, \beta, \sigma_W, \rho)$–the likelihood contribution of wages;

$\mathcal{L}_O^m(Y_i; \alpha_O, \lambda)$–the likelihood contribution of occupation;

$\mathcal{L}_S^m(Y_i; \alpha_S, \delta)$–the likelihood contribution of education;

$\mathcal{L}_Q^m(Y_i; \alpha_Q, \theta, \sigma_Q)$–the likelihood contribution of test scores.

The detailed expressions for each of the four likelihood contributions are collected in Appendix C.

4 Nonparametric Identification of The Finite Mixture Model

In this section, I discuss the nonparametric identification of the finite mixture model using the results in Kasahara and Shimotsu (2009) and Kasahara and Shimotsu (2012). Nonparametric identification means that the proportion of types, and type-specific joint distributions of wages, occupations, education, and test scores, which is unknown, can be recovered from the observed empirical population joint distribution of wages, occupations, education, and test scores. I use two sources of information to achieve the nonparametric identification: variations of wages across occupations over time and test scores. I show that the wage history is helpful to identify the occupation abilities. Yet, test scores are essential to identify the education psychic costs.

4.1 Nonparametric Identification of the Occupation Abilities

I use variations in wages across occupations over time to identify the occupation abilities. Intuitively, individuals with a comparative advantage in a white-collar occupation may have high white-collar wages, and hence, the fraction of individuals with high white-collar wages can be informative about the fraction of individuals with a comparative advantage in a white-collar occupation. In other words, the fractions of individuals with high white-collar wages over time impose restrictions on the unknowns type probabilities and type-specific distributions.

Three elements are the important determinants of identification: (1) the time-dimension of panel data, (2) the variation in the occupation-specific work experience, and (3) the heterogeneity in wages and occupational choices of individuals with different occupation abilities conditional on the occupation-specific work experience. The number of observed restrictions depend on the first two elements. The third element says that variations in wages are informative about the occupation abilities.

Let's start with a simple case in which wage and occupation distribution functions are stationary and there is no serial correlation.

Proposition 4.1. *Suppose Assumption 1 and Assumption 2 hold. With $T \geq 3$, π^m, $f^m(S_i|Z_i)$, $f^m(W_{i1}|O_{i1}, S_i)$,$f^m(O_{i1}|S_i)$, $f^m(W_{it}|O_{it}, S_i, X_{it}, W_{it-1}, O_{it-1}, X_{it-1})$, and $f^m(O_{it}|S_i, X_{it}, O_{it-1})$ for $t \geq 2$ can be identified up to M types.*

The assumptions and the proof of Proposition 4.1 are collected in Appendix D. The number of types M that can be identified depends on the number of values $\{X_{it}\}_{t=1}^{T}$ can take and its changes over time. The key insight is that each different value of $\{X_{it}\}_{t=1}^{T}$ imposes different restrictions on the type probabilities and type-specific distributions.

The assumption that current wage and occupational choice are not influenced by the lagged values is restrictive. The productivity shocks in the wage equation can be serially correlated and occupation in the last period can affect the occupation searching cost in the current period. The next proposition relaxes this strong assumption by allowing current wage and occupation depend on those in the last period.

Proposition 4.2. *Suppose Assumption 3 and Assumption 4 hold, and assume* $T \geq 6$*. Then* π^m*,* $f^m(S_i|Z_i)$*,* $f^m(W_{i1}|O_{i1}, S_i)$*,* $f^m(O_{i1}|S_i)$*,* $f^m(W_{it}|O_{it}, S_i, X_{it}, W_{it-1}, O_{it-1}, X_{it-1})$*, and* $f^m(O_{it}|S_i, X_{it}, O_{it-1})$ *for* $t \geq 2$ *can be nonparametrically identified up to* M *types.*

The assumptions and the proof of Proposition 4.2 are in Appendix D. If there is longer dependence in either wage or occupational choice, a longer panel is required. For example, suppose current wage is affected by wage two periods before, then at least 9-period observations are needed for identification.

The education psychic costs cannot be nonparametrically identified with panel data. The reason is that postsecondary education is one-period choice in my model, so there is no information over time that can distinguish the unobserved noises and unobserved education psychic costs in the education equations. Although Keane and Wolpin (1997) consider year-by-year schooling decisions, education is not an option for each time period due to the fact that the probability of going back to school after working is very low. Therefore, education psychic costs are not nonparametrically identified in Keane and Wolpin (1997).

4.2 Nonparametric Identification of the Education Psychic Costs

In order to nonparametrically identify the education psychic costs, I use six test scores. They are math, verbal, coding, mechanical tests in ASVAB, the Rotter Locus of Control, and the Rosenberg Self-Esteem Scale. The nonparametric identification using test scores are intuitive. For example, individuals with low education psychic costs may have good math and verbal test scores. So the fraction of individuals with good test scores in math and verbal is informative about the fraction of individuals with low education psychic costs.

Assume that the test scores do not directly affect postsecondary choices conditional on type and some observables. In other words, test scores do not affect postsecondary education application and admission once type and other observables are known. In addition, assume that there are three test scores which are independent from each other conditional on type and some observables. These two assumptions lead to the nonparametric identification of education psychic costs.

Proposition 4.3. *With access to three test scores* (Q_1, Q_2, Q_3)*, Suppose Assumption 5 holds. Then* π^m*,* $f^m(S_i|Z_i)$*, and* $\{f^m(Q_{ir}|S_{ir}, Z_{ir})\}_{r=1}^3$ *can be nonparametrically identified up to M types.*

Assumption 5, and the proof of Proposition 4.3 are collected in Appendix D.

The nonparametric identification of the education psychic costs using test scores does not require the six test scores to be perfect proxies. In other words, the nonparametric identification does not need the assumption that education does not affect test scores as it does when test scores are used as proxies. Such assumption is restrictive because some respondents already finished schooling when the test were administrated. The finding that education does influence the ASVAB scores, Rotter Locus of Control, and Rosenberg Self-Esteem Scale in Heckman et al. (2006b) further show the importance to relax such assumption.

Not only can test scores identify the education psychic costs, they can identify the occupation abilities as well. For instance, the Rotter Locus of Control which measures people's belief in their ability to control life may be important to a management job. So the fraction of individuals who perform well in this test is informative about the fraction of individuals with a comparative advantage in a white-collar occupation. It implies that the nonparametric identification of the finite mixture model can be achieve without the panel data, although the additional information from the variations in wages across occupations over time are helpful to increase the efficiency. Different from previous papers such as Keane and Wolpin (1997), Belzil and Hansen (2002), and Belzil and Hansen (2007), which reply on a long panel data to identify a finite mixture model, test scores allow me to apply a finite mixture model to data with limited periods of observations.

Assume that the test scores do not directly affect wages, education and occupation choices conditional on type and some observables. Further, assume that there are three test scores which are independent from each other conditional on type and some observables. These two assumptions lead to the nonparametric identification of both occupation abilities and education psychic costs.

Proposition 4.4. *With access to three test scores* (Q_1, Q_2, Q_3)*, Suppose Assumption 6 holds. Then* π^m*,* $f^m(S_i|Z_i)$*,* $\{f^m(Q_{ir}|S_{ir}, Z_{ir})\}_{r=1}^2$*,* $f^m(W_{i1}|O_{i1}, S_i)$*,* $f^m(O_{i1}|S_i)$*,* $f^m(W_{it}|O_{it}, S_i, X_{it}, W_{it-1}, O_{it-1}, X_{it-1})$*, and* $f^m(O_{it}|S_i, X_{it}, O_{it-1})$ *for* $t \geq 2$ *can be nonparametrically identified up to* M *types.*

Assumption 6 and the proof of Proposition 4.4 are collected in Appendix D.

The exclusion condition that test scores do not directly affect wage, education and occupation choices conditional on type and some observables is the key asumption to nonparametrically identify occupation abilities and education psychic costs using test scores. Intuitively, the exclusion of test scores from occupation and wage means that once employers know an individual's type, addition information about test scores would not influence the their decision on hiring and salary. The exclusion of test scores from education means that test scores do not affect postsecondary education application and admission once type is known. The assumption that test scores are exclusive from wage, education, and occupation is different from the exclusion condition in the IV approach and Heckman's two-step. While the exclusion variable in the IV approach and Heckman's two-step must not be correlated with the unobserved type, the exclusion variable in the finite mixture model has to be correlated with the unobserved type.

5 Empirical Results

EM algorithm (Dempster et al., 1977) is applied in this paper to facilitate the computation of finding the maximum likelihood estimates. As is well known, direct maximization of the likelihood function based on Newton-Raphson type algorithm is difficult for a finite mixture model because of the possibility of many local maxima. EM algorithm is a method for finding maximum likelihood estimates by iterating an expectation (E) step and a maximization (M) step. In E step, the expectation of the log-likelihood is calculated given the estimated the type proportions and parameters in the previous iteration. In M step, parameters are updated by maximizing the expected log-likelihood found in E step. The details about the E step and M step are discussed in Appendix E. Each iteration increases the value of log-likelihood and it stops when convergence is achieved. The corresponding estimates upon convergence are either a local maximum or saddle points. EM algorithm is found to be sensitive to initial parameters. I choose initial parameters following the approach suggested by Heckman and Singer (1984) and a detailed discussion is in Appendix F.

The empirical results are presented below. I assume that the overall population consists of four types. There are two occupation abilities type: type 1 and type 2 have the same occupation abilities, and so do type 3 and type 4. Within each occupation ability type, I consider two types of education psychic costs: type 1 and type 2 are different in education psychic costs, although they have the same occupation abilities. Similarly, type 3 and type 4 have different education psychic costs but share the same occupation abilities.

5.1 Occupation-specific Returns to Education, Education and Occupation Choices

As illustrated in Section 4, the nonparametric identification of the finite mixture model heavily relies on the informativeness of the test scores about the unobserved types. If test scores are reflective about unobserved occupation abilities and education psychic costs, we would expect different types to have different test scores. Table 3 reports the estimated parameters in Equation (4). It shows that the test scores vary across the four types and confirms that test scores are helpful to nonparametric identification of the finite mixture model. In addition, table 3 suggests that it is important to take into account the impact of education on test scores. A two-year college attendance significantly increases math and verbal test scores and a four-year college attendance significantly improves all the six test scores. The finding that education improves test scores implies that the test scores are not perfect proxies and using them as proxies would not help addressing the endogeneity issue to give unbiased estimates of occupation-specific returns to attending a two-year college and a four-year college.

Table 4 reports the estimated parameters in Equation (1). It shows that returns to education are occupation specific. The return to attending a two-year college is significantly higher in a blue-collar occupation than a white-collar occupation. A two-year college attendance increases blue-collar hourly payment by 24% and white-collar hourly payment by 17%. Regarding the returns to attending a four-year college, a four-year college attendance significantly increases more white-collar hourly wages than blue-collar wages. A four-year college attendant's hourly wage is 23% higher in a blue-collar occupation and 30% higher in a white-collar occupation than a high school graduate. Comparing a two-year college and a four-year college, these two kinds of postsecondary education institutions increase blue-collar wages similarly while a four-year college attendance is significantly more helpful to enhancing white-collar wages than a two-year college attendance does.

Converting the returns to attending a two-year college and a four-year college into annual returns, the corresponding annual return[11] to two-year college education is 20% in a blue-collar occupation, and 14% in a white-collar occupation. The corresponding annual return to four-year college education is 6% in blue-collar occupation, and 8% in white-collar occupation.

Among the people with post-secondary education in the sample, 27% are two-year college attendants and 73% are four-year college attendants. Hence, on average one year post-secondary education increases blue-collar wages by 10% ($20\% \times 27\% + 6\% \times 73\%$) and white-collar wages by 10%($14\% \times 27\% + 8\% \times 73\%$). Keane and Wolpin (1997) find that

[11]According to Table 1, two-year college attendants have 1.20 year more schooling than high school graduates and four-year college attendants have 3.82 year more schooling than high school graduates on average.

the annual return to postsecondary education is 2.4% in a blue-collar occupation and 7% in a white-collar occupation. The annual return to postsecondary education in a white-collar occupation in this paper is similar to that reported in Keane and Wolpin (1997). However, the annual return to postsecondary education in a blue-collar occupation is higher in this paper than in Keane and Wolpin (1997) where they do not distinguish a two-year college and a four-year college.

The type-specific constants reported in table 4 suggest that individuals are endowed with different occupation abilities. Among the four types of individuals, type 1 and type 2 share the same occupation abilities, but are different in the education psychic costs. Type 3 and type 4 have the same occupation abilities, but different education psychic costs. Although the occupation abilities and the education psychic costs may be correlated, the education psychic costs do not directly affect wages as assumed. So type 1 and type 2 earn the same, and so do type 3 and type 4. As reported in table 4, type 1 and type 2 earn more in a white-collar occupation than a blue-collar occupation. In other words, type 1 and type 2 have a comparative advantage in a white-collar occupation. On the other hand, type 3 and type 4 are similarly productive in a blue-collar and a white-collar occupation, because they earn similar wages in a white-collar and a blue-collar occupation.

Table 5 shows the estimated average partial effects in the occupation choice. Column 1 in table 5 reports the estimated average partial effects in making the occupation choice in the first job. It indicates that individuals with a comparative advantage in a white-collar occupation (type 1 and type 2) are 16% more likely to choose to work in white-collar jobs than those with a comparative advantage in a blue-collar occupation (type 3 and type 4). Moreover, education increases the probability of being employed in a white-collar occupation. Comparing to high school graduates, two-year college attendants are 16% more likely to work in a white-collar occupation and four-year college attendants are 50% more likely to work in a white-collar occupation.

Column 2 in table 5 reports the estimated average partial effects in making the occupation choice in the sequential jobs. It shows that individuals with a comparative advantage in a white-collar occupation (type 1 and type 2) are 6% more likely to work in white-collar jobs than those with a comparative advantage in a blue-collar occupation (type 3 and type 4) in the sequential jobs. Education has smaller influence on occupation in subsequent jobs than initial jobs. Attending a two-year colleges and a four-year college increase the probability of being employed by a white-collar occupation by 4% and 12% respectively. One important factor which affects the occupation choice is the occupation in the previous period. An individual who worked in a white-collar occupation in the previous period is 29% more likely to work in a white-collar occupation in the current period than an individual who worked in a blue-collar occupation in the previous period does.[12]

Regarding the postsecondary education choice, if individuals consider their future occupations when making their education decisions, we would expect individuals with a comparative advantage in a white-collar occupation (type 1 and type 2) to be more likely to attend a four-year college than individuals earn similarly in a blue-collar occupation and a white-collar occupation (type 3 and type 4). Table 6 reports the average partial effects in the postsecondary education choice. It shows that type 1 and type 2 are 53% more likely to attend a four-year college than type 3 and type 4. This finding confirms that the occupation abilities affect the education choice. Further, the results in table 6 suggest that individuals take into account the education psychic costs when making their education decisions. Although type 1 and type 2 share the same occupation abilities, type 1 is 42% more likely to attend a four-year college than type 2 is, which indicates that type 1 has lower psychic costs to attend a four-year college than type 2 does. Similarly, type 4 are more likely to attend a four-year college than type 3 does, which suggests that type 4 has lower psychic costs to attend a

[12] I have consider the case that individuals may have different occupation tastes. For example, those who enjoy working outdoors may prefer a construction worker position to an economist position. To copy with the potential heterogeneity in occupation taste, I estimate a finite mixture model with 8 types. Specifically, I consider two occupation abilities types. Within each occupation abilities type, there are two education psychic costs type. In addition, I look at two occupation taste types for individuals with the same occupation abilities and education psychic costs. I do not find that individuals with the same occupation abilities and education psychic costs behave differently in the occupation choices.

four-year college than type 3 does. Regarding the decision to attend a two-year college, the similarity across the four types in the decision to attend a two-year college suggests no self-selection in attending a two-year college.

To sum up, individuals self select into different education groups based on their occupation abilities and education psychic costs. Their occupation choices are also influenced by their occupation abilities. Failure to address the endogenous education and occupation choices can result in biased estimates of occupation-specific returns. Due to the complication of the self-selection problem here, it is hard to tell the direction of the possible bias. I compare the estimates of occupation-specific returns to education when controlling or not controlling for occupation abilities and/or education psychic costs in table 7. Column (4) presents the estimates of occupation-specific returns to attending a two-year college and a four-year college controlling both the occupation abilities and education psychic costs (the same estimates as those reported in table 4). Column (1) gives the OLS estimates of the occupation-specific returns to education. The OLS estimates of the occupation-specific returns to attending a two-year college are slightly lower than those in column (4) and the OLS estimates of the occupation-specific returns to attending a four-year college are comparable to those in column (3). Column (2) in table 7 shows the OLS estimates of the occupation-specific returns to education when six test scores are included as proxies for occupation abilities and education psychic costs. Using test scores as proxies requires that education does not affect test scores. However, the results in table 3 suggest that education helps to improve the performance in all the six tests. The estimated returns to attending a two-year college and a four-year college in column (2) are around 6 percentage points lower than those reported in column (4). Column (3) in table 7 gives the estimates of the occupation-specific returns to education only controlling for the occupation abilities[13]. The estimates of the occupation-specific returns to attending a two-year college in column (3) are larger than both the OLS estimates in column

[13]Here, I estimate a finite mixture model in which there are two occupation abilities types and everyone has the same education psychic costs. Since the variations in wages across occupations over time are sufficient to nonparametrically identify occupation abilities, I do not use test scores as an additional source of nonparametric identification

(1) and those in column (4). The estimates of the occupation-specific returns to attending a four-year college in column (3) are comparable to those in column (1) and column (4). The results in table 7 suggest that the possible biases are of different direction and cancel out each other although education and occupation choices are endogenous.

5.2 Conditional Independence of Wages and Test Scores

One of the key assumptions of the nonparametric identification of the finite mixture model is that conditional on type and some other observables the six test scores do not directly affect wages. I test the validation of this conditional independence assumption by including the six test scores one by one into the wage equation (Equation (1)). The idea is that assuming the finite mixture model is nonparametrically identified using the other test scores, the coefficient on test score r, which is included in the wage equation, should be zero when test score r and wages are conditionally independent. Table 8 presents the estimated coefficients on the six test scores. The estimated coefficients on coding speed, Rotter locus of control, and Rosenberg self-esteem scale are not significantly different from zero. According to Proposition 4.4, the finite mixture model can be nonparametrically identified because we have three tests satisfied the conditional independence assumption. However, we reject the hypothesis that the conditional independence assumption holds for math skill, verbal skill, and mechanical comprehension. The reason of the finding that math, verbal, mechanical comprehension scores affect wages is that I only consider a small number of types (two occupation abilities types and two education psychic costs types). It is possible that there are heterogeneity in occupation abilities and education psychic costs within each of the four types and math skill, verbal skill, and mechanical comprehension scores are informative about these within type heterogeneity. I check the sensitivity of the estimates in two ways. First, I include the math, verbal, and mechanical comprehension scores together into the wage equation and check whether the estimates of the occupation-specific returns to education are different from those reported in table 4. Table 9 shows that the estimates of the occupation-specific returns to education are around 2 to 5 percentage points smaller than those in table 4. Second, I increase the number of types from 4 (two types of occupation abilities and two types of education psychic costs) to 6 (three types of occupation abilities and two types of education psychic costs) and the corresponding estimates in the wage equation are presented in table 10. Table 10 shows that the estimates of occupation-specific returns to education are around 1 to 4 percentage points smaller than those reported in table 4.

5.3 The Occupation-Specific Returns to A Bachelor's Degree

The wage gap between college dropouts and college graduates are documented in the literature (Jaeger and Page, 1996). It is interesting to examine the occupation-specific returns to college graduate besides the occupation-specific returns to college attendants. Due to the small sample size of the two-year college graduates, I focus on investigating the occupation-specific returns for those obtained a bachelor's degree.

Among the four-year college attendants in my sample, around 70% obtained a bachelor's degree. A simple comparison of the first year wage of the four-year college dropouts and the four-year college graduates shows that the four-year college dropouts and the four-year college graduates earn similarly in a blue-collar occupation, yet the four-year college graduates earn 30% more than the four-year college dropouts in a white-collar occupation[14].

Table 11 presents the estimated parameters of the wage equation (Equation (1)) using the sample where the four-year college dropouts are eliminated. It shows that a bachelor's degree increases blue-collar wages by 26% and white-collar wages by 33% for a high school graduate. Comparing to the returns to attending a four-year college as reported in table 4, ie. 23% and 30% respectively for a blue-collar occupation and a white-collar occupatio, the estimated occupation-specific returns to a bachelor's degree are not much higher.

[14]For more summary statistics of the college dropouts and college attendants, please refer to table A1.

5.4 The Expected Returns to Education

Individuals make their education choices taking into account their future occupations. Yet, they do not know exactly their occupations because of the uncertainty in the labour market. Therefore, their education choices are based on the expected returns to attending a two-year and a four-year college. Below, I calculated the expected returns to education by simulating a sample of 10000 observations.

Panel A of table 12 shows that the expected returns to attending a two-year college are around 23% for type 1 and type 2 (individuals with a comparative advantage in a white-collar occupation) over time[15], and they are around 22% for type 3 and type 4 (individuals with a comparative advantage in a blue-collar occupation) over time. Returns to attending a two-year college are similar to all types, which explains that individuals do not select to attend a two-year college based on their occupation abilities as suggested by the results in table 6. Regarding a four-year college, the expected returns to attending a four-year college for type 1 and type 2 are 34% in the first year and increase to 40% nine years later. Returns to attending a four-year college for type 3 and type 4 are 25% in the first year and increase to 29% in the tenth year. Returns to attending a four-year college are around 9 percentage points higher for type 1 and type 2 than type 3 and type 4 in the first year and the difference increases to 11 percentage points in the 10th year. Therefore, individuals with a comparative advantage in a white-collar occupation (type 1 and type 2) are more likely to attend a four-year college than those with a comparative advantage in a blue-collar occupation (type 3 and type 4) as shown in table 6. Comparing returns to attending a two-year and a four-year college, returns to attending a four-year college are 10 percentage points higher than returns to a two-year college in the beginning and the discrepancy is enlarged to 16 percentage points after nine years for type 1 and type 2. For type 3 and type 4, returns to attending a four-year

[15] Since the occupation choice depend on the occupation abilities and not influenced by education psychic costs directly, the expected returns to attending a two-year college are the same for individuals with same occupation abilities and different education psychic costs. That is to say that type 1 and type 3 have the same expected returns to attending a two-year college, and type 2 and type 4 gain the same in earnings from attending a two-year college

college are 3 percentage points higher than returns to attending a two-year college in the beginning and the difference increases to 7 percentage points after nine years. The difference between returns to attending a two-year college and a four-year college echoes the findings in Belzil and Hansen (2002) that returns can be education-level-specific.

The expected returns to attending a two-year and a four-year college are different across types of people with different occupation abilities. The relationship between returns to education and innate abilities are well documented in the literature (Belzil and Hansen, 2007; Carneiro et al., 2003). The reason of the correlation between the expected returns to education and the occupation abilities is that the expected returns to education are related to the probability of working in a white-collar occupation, which depend on occupation abilities. The expected returns to education and the probability of working in a white-collar occupation can be related in two ways. First, education helps accumulating white-collar and blue-collar skills differently. For example, attending a four-year college increases white-collar skills more than blue-collar skills. Therefore, individuals with a comparative advantage in a white-collar occupation, who are more likely to work in a white-collar occupation, have higher returns to attending a four-year college on average. Second, education enhances the probability of working in a white-collar occupation. For individuals with a comparative advantage in a white-collar occupation, the reward to their occupation abilities are higher in a white-collar occupation than a blue-collar occupation. Attending a four-year college education increases the probability of working in a white-collar occupation and leads to a high reward to their occupation abilities. I decompose the returns to education into these two parts: enhancing occupation-specific skills and increasing the probability of working in a white-collar occupation.

Panel B of table 12 shows the part of returns to education from enhancing occupation-specific skills and panel C of table table 12 presents the part of returns to education from increasing the probability of being employed in a white-collar occupation. Let's first look at the decomposition of the expected returns to attending a two-year college.

For individuals with a comparative advantage in a white-collar occupation (type 1 and type 2), a two-year college attendance increases wages by 22% from enhancing the occupation-specific skills. The part of expected returns from increasing the probability of working in a white-collar occupation is 2% at the beginning and 4% at the end. As type 1 and type 2 become more likely to work in a white-collar occupation, their latter part of the expected returns to attending a two-year college increases. For individuals with a comparative advantage in a blue-collar occupation (type 3 and type 4), the expected returns from enhancing occupation-specific skills are around 22% over the ten years, which is comparable to those of type 1 and type 2 in magnitude. The part of expected returns from increasing the probability of being employed in a white-collar occupation is almost close to zero over time. This is because type 3 and type 4 are rewarded similarly to their occupation abilities in both occupations. Next, let's look at the decomposition of the expected returns to attending a four-year college. For type 1 and type 2, the part of expected returns attending a four-year college due to occupation-specific skills accumulation is around 28%. The part of expected returns from increasing the probability of working in a white-collar occupation increases from 6% to 14% with the probability of working in a white-collar occupation increasing from 71% to 83% over time. After ten years in the labour market, 65% of the total expected returns to attending a four-year college education are from its impact on occupation-specific skills accumulation for type 1 and type 2. For type 3 and type 4. The part of expected returns to attending a four-year college due to occupation-specific skills accumulation is around 26%, which is comparable to that for type 1 and type 2 in magnitude. The part of expected returns to attending a four-year college from its influence on occupation affiliation is close to zero because type 3 and type 4 are rewarded similarly to their occupation abilities in a blue-collar and a white-collar occupation.

The increasing expected returns to attending a four-year college for individuals with a comparative advantage in a white-collar occupation (type 1 and type 2) imply a faster wage growth rate of four-year college attendants than high school graduates. This finding is consistent with Willis and Rosen (1979) where they find that a college attendant's wage grows faster than a high school graduates. This chapter suggests that one important reason for the faster wage growth rate of four-year college attendants is that they switch to the occupation they have a comparative advantage of over time.

5.5 Test Scores and Returns to Education

As shown in table 3, test scores are informative about individuals occupation abilities and education psychic costs. Once the type-specific joint distributions of the six test scores are identified, we can get the probabilities of types conditional on the six test scores using Bayes' rule. In other words, we are able to tell which type an individual is most likely to be given her six test scores and demographic information. Further, we can infer her expect returns to attending a two-year college and a four-year college.

I simulate the six test scores for 10,000 high school graduates, whose parents are high school graduates, who have three siblings, were raised in a two-parent family, lived in the northern urban area of U.S. at age 14, and took the six test scores at age 18. For simplicity, I divide the six test scores into two groups: cognitive tests (math skills, verbal skills, coding speed, and mechanical comprehension) and noncognitive tests (the Rotter Locus of Control and the Rosenberg Self-Esteem Scale). Then I calculate the average scores, $\hat{Q}_c$ and $\hat{Q}_{nc}$, for the two groups. For further simplicity, each of the two average test scores are partitioned into four parts. The proportion of each type conditional on test scores are presented in table 13. For example, look at an individual with high school graduates parents, 3 siblings, raised in a two parents family, lived in the North urban area of U.S. at age 14, and took the cognitive and noncognitive tests at age 18. If all her test scores are at the 10th percentile, her net average scores are in the cell of row 1 and column 1 and she is 0.1% likely to be type 1, 61.5% likely to be type 2, 29.8% likely to be type 3, and 8.6% likely to be type 4. If all her test scores are at the 50th percentile, her net average scores are in the cell of row 2 and column 3 and she is 15% likely to be type 1, 0.1% likely to be type 2, 32.5% likely to be type 3, and 52.3% likely to be type 4. If all her test scores are at the 90th percentile, her net average scores are in the cell of row 4 and column 4 and she is 90.9% likely to be type 1, 0% likely to be type 2, 0% likely to be type 3, and 9.1% likely to be type 4. Once the conditional proportion of types is known, her returns to education can be calculated accordingly. Table 14 shows the expected returns to attending a two-year college and a four-year college for such an individual with test scores at the 10th, 50th, and 90th percentile. Since returns to two-year college are similar to all types, the expected returns to two-year college are almost the same for different test scores. Regarding a four-year college, the expected returns to attending a four-year college increase as test scores increase. The reason is that high test scores imply a high probability of a comparative advantage in a white-collar occupation, and a comparative advantage with a white-collar occupation are associated with a high returns to attending a four-year college.

6 Conclusion

In this paper, I examine the returns to attending a two-year college and a four-year college and how the returns to education differ from those of a white-collar occupation to those of a blue-collar occupation. Despite a vast literature on returns to education, the existing research on how returns to attending a two-year college and a four-year college depend on the occupation choice is limited. The reason for this limitation is that it is difficult to estimate the occupation-specific returns in the presence of endogenous education and occupation choices. On the one hand, individuals are endowed with different abilities to work in a blue-collar occupation or a white-collar occupation. They tend to work in the occupation in which they have a comparative advantage. Moreover, they are more likely to choose the type of postsecondary education that intensively accumulates the skills needed in the occupations they would like to work in when they finish schooling. Therefore, occupation abilities drive wages, education, and occupation. On the other hand, individuals vary in their education psychic costs, which may be correlated with occupation abilities. While the occupation abilities and the education psychic costs are known to the individuals making education and occupation decisions, these abilities and costs are unobserved by the econometrician, thus leading to the missing variable problem. The instrumental variables (IV) approach, conventionally used to deal with the endogeneity issue in the returns to education literature, is difficult to implement here simply because good instruments for both education and occupation are difficult to find.

I address the endogeneity issue in education and occupation by explicitly modeling the sequential education and occupation choices, specifying the unobserved occupation abilities and education psychic costs with a flexible multinomial distribution in a finite mixture model. I show how to nonparametrically identify the occupation abilities using the variations in wages across occupations over time. However, the information from the panel data alone is not enough to identify the education psychic costs. In order to achieve nonparametric identification of the education psychic costs, I use test scores such those of the ASVAB, the Rotter Locus of Control, and the Rosenberg Self-Esteem Scale. I show that conditional on occupation abilities and education psychic costs the education psychic costs can be nonparametrically identified under the assumption that the test scores do not directly affect wages, education, or occupation choices.

Using data taken from the National Longitudinal Survey of Youth (NLSY) 1979, I estimate a parametrically specified finite mixture model for joint wages, education, occupation, and test scores and find that returns to education are occupation-specific. Specifically, I find that attendance of a two-year college enhances blue-collar wages by 24% and white-collar wages by 17% while attendance of a four-year college increases blue-collar wages by 23% and white-collar wages by 30%.

Appendices

A Summary Statistics, College Dropouts vs. College Graduates

In table A1, I divide two-year college attendants into those with and without an associate degree, and divide four-year attendants to those with and without a bachelor's degree. Among the two-year college attendants, 75% do not obtain an associate degree. The average schooling years of the two-year college dropouts are 12.70 years and those of the two-year college graduates are 14.2 years. Among the four-year college attendants, around 30% drop out of four-year college while the majority obtain a bachelor's degree. The average schooling years of the four-year college dropouts are 13.6 years and those of the four-year college graduates are around 16.6 years. Regarding the the first job after schooling, those who obtain an associate degree are slightly more likely to work in a white-collar occupation than those drop out of a two-year college. The probability of initially working in a white-collar position of the two-year college dropouts is 24.4% and that of individuals with an associate degree is 27.5%. Interestingly, although the four-year college dropouts have more schooling years than those with an associate degree, the former are more likely to work in a white-collar occupation entering the labour market than the latter. On average, around 37.4% of the four-year college dropouts initially work in a white-collar occupation. Those with a bachelor's degree are much more likely to start with a white-collar occupation than the others. Around 75.2% of those with a bachelor's degree work in a white-collar occupation as their first jobs. Regarding wages, the two-year college dropouts and the four-year dropouts earn almost the same. The average hourly payment for the two-year college dropouts and the four-year dropouts are \$10.19 and \$10.67 respectively. Those with an associate degree earn around \$12.19 per hour for their first jobs. The hourly payment of those with an associate degree is higher than that of the two-year college dropouts and four-year college dropouts. Those with a bachelor's

Table A1: Discriptive Statistics (More Education Categories)

	2-yr Dropouts			Associate Degree			4-yr Dropouts			Bachelor's Degree		
Variables	Obs	Mean	S.D	Obs	Mean	S.D	Obs	Mean	S.D	Obs	Mean	S.D
Highest grade completed	119	12.697	0.859	40	14.200	0.608	131	13.573	1.336	307	16.619	1.180
Age in 1979	119	17.277	2.004	40	17.525	2.075	131	17.55	2.146	307	17.759	2.246
Initial job(white collar)	119	0.244	0.431	40	0.275	0.452	131	0.374	0.486	307	0.752	0.432
Initial wage	119	10.19	4.579	40	12.185	5.605	131	10.669	5.18	307	14.106	13.981
Initial wage(blue collar)	119	10.524	4.926	40	12.274	6.006	131	10.099	4.513	307	9.935	4.603
Initial wage(white collar)	119	9.155	3.128	40	11.95	4.629	131	11.624	6.067	307	15.478	15.669
Mother education	119	12.126	1.964	40	12.4	1.751	131	12.328	1.854	307	13.166	2.096
Father education	119	12.689	2.626	40	12.9	2.479	131	13.038	3.134	307	14.055	3.072
Number of siblings	119	2.798	1.754	40	2.45	1.694	131	2.748	1.729	307	2.43	1.596
Broken family at age 14	119	0.176	0.383	40	0.075	0.267	131	0.115	0.32	307	0.094	0.293
South at age 14	119	0.244	0.431	40	0.175	0.385	131	0.244	0.431	307	0.248	0.432
Urban at age 14	119	0.84	0.368	40	0.65	0.483	131	0.725	0.448	307	0.814	0.389

degree earn the most among the post-secondary attendants. The average hourly payment of those with a bachelor's degree is around \$14.11. When we take a closer look at wages by separating individuals into two occupation groups, those initially work in a blue-collar occupation and those initially work in a white-collar occupation, the story is different. The two-year college dropouts, the four-year college dropouts, and those with a bachelor's degree earn around \$10 per hour if their first job is blue-collar while those with an associate degree earn \$12.27 per hour if their first job is blue-collar. For those whose first job is white-collar, the two-year college dropouts earn \$9.15 per hour. Those with an associate degree and the four-year college dropouts earn around \$12 per hour. Those with a bachelor's degree earn more than the two-year college dropouts, those with an associate degree, and the four-year college dropouts. The hourly payment of those with a bachelor's degree is \$15.48 per hour. Although this paper mainly examines the impact of attendance of a two-year college and a four-year college on wages, I also provide estimates of the occupation-specific wage gains to obtain a bachelor's degree for a high school graduate by eliminating the four-year college dropouts from the sample. I do not study the occupation-specific returns to an associate degree because the sample size of those with an associate degree is too small or reasonable results.

B Simplification of the Type-Specific Joint Distribution

Below, I show how to simplify the type-specific joint distribution of wages, occupations, education, and the test scores.

$$
\begin{aligned}
& f^m(\{W_{it}, O_{it}\}_{t=1}^T, S_i, \{Q_{ir}\}_{r=1}^6 | \{X_{it}\}_{t=1}^T, Z_{S,i}, \{Z_{ir}\}_{r=1}^6) \\
=& f^m(\{W_{it}, O_{it}\}_{t=1}^T, S_i | \{Q_{ir}\}_{r=1}^6, \{X_{it}\}_{t=1}^T, Z_{S,i}, \{Z_{ir}\}_{r=1}^6) \\
& \times f^m(\{Q_{ir}\}_{r=1}^6 | \{X_{it}\}_{t=1}^T, Z_{S,i}, \{Z_{ir}\}_{r=1}^6) \\
=& f^m(\{W_{it}, O_{it}\}_{t=1}^T, S_i | \{Q_{ir}\}_{r=1}^6, \{X_{it}\}_{t=1}^T, Z_{S,i}) f^m(\{Q_{ir}\}_{r=1}^6 | \{Z_{ir}\}_{r=1}^6) \\
=& f^m(\{W_{it}, O_{it}\}_{t=1}^T, S_i | \{Q_{ir}\}_{r=1}^6, \{X_{it}\}_{t=1}^T, Z_{S,i}) \prod_{r=1}^6 f^m(\{Q_{ir}\}_{r=1}^6 | \{Z_{ir}\}_{r=1}^6) \\
=& f^m(\{W_{it}, O_{it}\}_{t=1}^T, S_i | \{Q_{ir}\}_{r=1}^6, \{X_{it}\}_{t=1}^T, Z_{S,i}) \prod_{r=1}^6 f^m(\{Q_{ir}\}_{r=1}^6 | \{Z_{ir}\}_{r=1}^6) \\
=& f^m(W_{i1} | O_{i1}, S_i) \prod_{t=2}^T f^m(W_{it} | O_{it}, S_i, X_{it}, W_{it-1}, O_{it-1}, X_{it-1}) \\
& \times f^m(O_{i1} | S_i) \prod_{t=2}^T f^m(O_{it} | O_{it-1}, S_i, X_{it}) f^m(S_i | Z_{S,i}) \prod_{r=1}^6 f^m(Q_{ir} | Z_{ir}).
\end{aligned}
$$

The first equality holds under the assumption that the six test scores do not directly affect wages, occupations, and education conditional on type. The second equality holds under the assumption that the regressors and the error terms in Equation (1), Equation (2), Equation (3), and Equation (4) are independent. The third equality holds under the assumption that the error terms in test scores are mutually independent ($\varepsilon_{Q,ir} \perp\!\!\!\perp \varepsilon_{Q,ir'}$ for $r \neq r'$). The fourth equality holds under the assumptions that the error terms in wage follows a first order Markov process ($\varepsilon_{W,it} = \rho\varepsilon_{W,it-1} + \zeta_{it}$) and the occupation choice is only affected by the previous occupation, not the whole occupation history.

C Likelihood Contributions

(a) The likelihood contribution of wages:

$$
\mathcal{L}_W^m(Y_i; \alpha_W, \beta, \sigma_W, \rho) = \phi(\frac{W_{i1} - \mu_{W,i1}}{\sigma_{W,1}}) \prod_{t=2}^T \phi(\frac{W_{it} - \mu_{W,i2}}{\sigma_{W,2}}).
$$

The wage density functions follow a normal distribution according to the assumptions in Equation 1. Specifically,

$$\mu_{W,i1} = \alpha^m_{W1} + \alpha^m_{W,2}O_{it} + \beta_1 2\mathrm{YR}_i + \beta_2 4\mathrm{YR}_i + \beta_3 2\mathrm{YR}_i O_{it} + \beta_4 4\mathrm{YR}_i O_{it} + X'_{it}\beta_5,$$

and

$$\begin{aligned}\mu_{W,i2} =& \alpha^m_{W1} + \alpha^m_{W,2}O_{it} + \beta_1 2\mathrm{YR}_i + \beta_2 4\mathrm{YR}_i + \beta_3 2\mathrm{YR}_i O_{it} + \beta_4 4\mathrm{YR}_i O_{it} + X'_{it}\beta_5 \\ & - \rho(W_{it-1} - (\alpha^m_{W1} + \alpha^m_{W,2}O_{it-1} + \beta_1 2\mathrm{YR}_i + \beta_2 4\mathrm{YR}_i + \beta_3 2\mathrm{YR}_i O_{it-1} \\ & + \beta_4 4\mathrm{YR}_i O_{it-1} + X'_{it-1}\beta_5))\end{aligned}$$

where $D_{O,ijt}$ is a dummy variable, which equals 1 if individual i works in occupation j at time t

(b) The likelihood contribution of occupations:

$$\begin{aligned}\mathcal{L}^m_O(Y_i;\alpha_O,\lambda) =& \Phi(\alpha^m_O + \lambda_1 2\mathrm{YR}_i + \lambda_2 4\mathrm{YR}_i) \\ & \times \prod_{t=2}^{T} \Phi(\alpha^m_O + \lambda_1 2\mathrm{YR}_i + \lambda_2 4\mathrm{YR}_i + \lambda_3 O_{it-1} + X'_{it}\lambda_4).\end{aligned}$$

(c) The likelihood contribution of education:

$$\mathcal{L}^m_S(Y_i;\alpha_S,\delta) = \frac{exp(\alpha^m_{S,j} + Z'_i\delta_j)}{1 + \sum_{j'=2}^{3} exp(\alpha^m_{S,j'} + Z'_i\delta_{j'})}.$$

(d) The likelihood contribution of test scores:

$$\mathcal{L}^m_Q(Y_i;\alpha_Q,\theta,\sigma_Q) = \prod_{r=1}^{6} \phi(\frac{Q_{ir} - \mu_{Q,ir}}{\sigma_{Q,r}}).$$

The density functions of test scores follow a normal distribution according to the assumptions in Equation 4, and $\mu_{Q,ir} = \alpha_r^m + \theta_{r,1}2\text{YR}_{ir} + \theta_{r,2}4\text{YR}_{ir} + Z'_{i,r}\theta_{r,3}$.

D Assumptions and Proofs of Propositions

D.1 Assumptions and proof of Proposition 4.1

Assumption 1. *For m=1,...,M and $t \geq 2$,*

(a)

$$f_t^m(W_t|O_t, X_t, S, \{W_\tau, O_\tau, X_\tau\}_{\tau=2}^{t-1}) = f^m(W_t|O_t, X_t, S, \{W_\tau, O_\tau, X_\tau\}_{\tau=2}^{t-1}),$$

and

$$f_t^m(O_t|X_t, S, \{W_\tau, O_\tau, X_\tau\}_{\tau=2}^{t-1}) = f^m(O_t|X_t, S, \{W_\tau, O_\tau, X_\tau\}_{\tau=2}^{t-1}).$$

(b)

$$f^m(W_t|O_t, X_t, S, \{W_\tau, O_\tau, X_\tau\}_{\tau=2}^{t-1}) = f^m(W_t|O_t, X_t, S),$$

and

$$f^m(O_t|X_t, S, \{W_\tau, O_\tau, X_\tau\}_{\tau=2}^{t-1}) = f^m(O_t|X_t, S).$$

Assumption 1 reduces the number of unknown type-specific distributions and the conditional type-specific joint distributions of wages, occupations, and education can be simplified as follows:

$$\begin{aligned}
&f^m(\{W_t, O_t\}_{t=1}^T, S|\{X_{it}\}_{t=1}^T, Z_S) \\
=&f^m(W_1|O_1, S)\prod_{t=2}^T f^m(W_t|O_t, S, X_t) \\
&\times f^m(O_1|S)\prod_{t=2}^T f^m(O_t|S, X_t)f^m(S|Z_S).
\end{aligned}$$

For the sake of clarity, assume the support of X_t (t=2,...,T) is discrete and known. Let $(\eta_{t,1}, \eta_{t,2}, \ldots, \eta_{t,M-1})$ be elements of $\mathcal{X}_t$ for t=1,...,T. Fix $S = s$ and define, for $(\eta_t, \eta_1, z_S) \in \mathcal{X}_t \times \mathcal{X}_1 \times \mathcal{Z}_\mathcal{S}$,

$$\begin{aligned}
&\lambda_{O,\eta_1}^{*m} = P^m(O_1 = 1|(X_1, S) = (\eta_1, s)), \\
&\lambda_{O,\eta_t}^{m} = P^m(O_t = 1|(X_t, S) = (\eta_t, s)), \\
&\widetilde{\pi}_{z_S}^m = \pi^m P^m(S = s|Z_S = z_S)
\end{aligned}$$

Construct a matrix of type-specific distribution functions and type probabilities as

$$L_t = \begin{pmatrix} 1 & \lambda_{O,\eta_{t,1}}^1 & \cdots & \lambda_{O,\eta_{t,M-1}}^1 \\ \ldots & \ldots & \ddots & \ldots \\ 1 & \lambda_{O,\eta_{t,1}}^M & \cdots & \lambda_{O,\eta_{t,M-1}}^M \end{pmatrix},\ for\ t = 2, \ldots, T$$

$$D_{\eta_1}^O = diag(\lambda_{O,\eta_1}^{*1}, \ldots, \lambda_{O,\eta_1}^{*M}),\ and\ V_{z_S} = diag(\widetilde{\pi}_{z_S}^1, \ldots, \widetilde{\pi}_{z_S}^M).$$

The elements of L_t,$D_{\eta_1}^O$, and V_{z_S} are parameters of the underlying mixture model to be identified. Consider we have data for three time periods i.e. $T = 3$. Fix $O_t = 1$ for all t and define

$$F_{Z_S,X_1,X_2,X_3}^{O*} = \sum_{m=1}^M \widetilde{\pi}_{Z_S}^m \lambda_{O,X_1}^{*m} \lambda_{O,X_2}^m \lambda_{O,X_3}^m$$

.

Now fix $O_2 = O_3 = 1$ and define

$$F^{O}_{Z_S,X_2,X_3} = \sum_{m=1}^{M} \widetilde{\pi}^m_{Z_S} \lambda^m_{O,X_2} \lambda^m_{O,X_3}.$$

Similarly, define the following functions

$$\begin{aligned}
F^{O*}_{Z_S,X_1,X_2} &= \sum_{m=1}^{M} \widetilde{\pi}^m_{Z_S} \lambda^{*m}_{O,X_1} \lambda^m_{O,X_2}, \\
F^{O*}_{Z_S,X_1,X_3} &= \sum_{m=1}^{M} \widetilde{\pi}^m_{Z_S} \lambda^{*m}_{O,X_1} \lambda^m_{O,X_3}, \\
F^{O*}_{Z_S,X_1} &= \sum_{m=1}^{M} \widetilde{\pi}^m_{Z_S} \lambda^{*m}_{O,X_1}, \\
F^{O}_{Z_S,X_2} &= \sum_{m=1}^{M} \widetilde{\pi}^m_{Z_S} \lambda^m_{O,X_2}, \\
F^{O}_{Z_S,X_3} &= \sum_{m=1}^{M} \widetilde{\pi}^m_{Z_S} \lambda^m_{O,X_3}.
\end{aligned}$$

Arrange these into two $M \times M$ matrices:

$$P^O_{z_S} = \begin{pmatrix} 1 & F^O_{z_S,\eta_{3,1}} & \cdots & F^O_{z_S,\eta_{3,M-1}} \\ F^O_{z_S,\eta_{2,1}} & F^O_{z_S,\eta_{2,1},\eta_{3,1}} & \cdots & F^O_{z_S,\eta_{2,1},\eta_{3,M-1}} \\ \vdots & \vdots & \ddots & \vdots \\ F^O_{z_S,\eta_{2,M-1}} & F^O_{z_S,\eta_{2,M-1},\eta_{3,1}} & \cdots & F^O_{z_S,\eta_{2,M-1},\eta_{3,M-1}} \end{pmatrix},$$

and

$$P^{O*}_{z_S,\eta_1} = \begin{pmatrix} F^{O*}_{z_S,\eta_1} & F^{O*}_{z_S,\eta_1,\eta_{3,1}} & \cdots & F^{O*}_{z_S,\eta_1,\eta_{3,M-1}} \\ F^{O*}_{z_S,\eta_1,\eta_{2,1}} & F^{O*}_{z_S,\eta_1,\eta_{2,1},\eta_{3,1}} & \cdots & F^{O*}_{z_S,\eta_1,\eta_{2,1},\eta_{3,M-1}} \\ \vdots & \vdots & \ddots & \vdots \\ F^{O*}_{z_S,\eta_1,\eta_{2,M-1}} & F^{O*}_{z_S,\eta_1,\eta_{2,M-1},\eta_{3,1}} & \cdots & F^{O*}_{z_S,\eta_1,\eta_{2,M-1},\eta_{3,M-1}} \end{pmatrix}.$$

To achieve identification, further assume:

Assumption 2. *There exist some* $\{\eta_{t,1}, \ldots, \eta_{t,M-1}\}_{t=2}^{T}$ *such that* $P^O_{z_S}$ *is of full rank and that all the eigenvalues of* $(P^O_{z_S})^{-1}P^{O*}_{z_S,\eta_1}$ *take distinct values.*

Proof of Proposition 4.1. $P^O_{z_S}$ and $P^{O*}_{z_S,\eta_1}$ can be expressed as the follows:

$$P^O_{z_S} = L_2'V_{z_S}L_3, \text{ and } P^{O*}_{z_S,\eta_1} = L_2'V_{z_S}D^O_{\eta_1}L_3.$$

Because $P^O_{z_S}$ is full rank, it follows that L_2 and L_3 are full rank. We can construct a matrix $A_{z_S} = (P^O_{z_S})^{-1}P^{O*}_{z_S,\eta_1} = L_3^{-1}D^O_{\eta_1}L_3$. Because $A_{z_S}L_3^{-1} = L_3^{-1}D^O_{\eta_1}$ and the eigenvalues of A_{z_S} are distinct, the eigenvalues of A_{z_S} determines the elements of $D^O_{\eta_1}$.

Moreover, the right eigenvectors of A_{z_S} are the columns of L_3^{-1} up to multiplicative constants. Denote $L_3^{-1}K$ to be the right eigenvectors of A_{z_S} where K is some diagonal matrix. Now we can determine $V_{z_S}K$ from the first row of $P^O_{z_S}L_3^{-1}K$ because $P^O_{z_S}L_3^{-1}K = L_2'V_{z_S}K$ and the first row of L_2' is a vector of ones. Then L_2' is determined uniquely by $L_2' = (P^O_{z_S}L_3^{-1}K)(V_{z_S}K)^{-1}$. Similarly, by construct a matrix $B_{z_S} = (P^{O\prime}_{z_S})^{-1}(P^{O*\prime}_{z_S,\eta_1})$, we can uniquely determine L_3'.

We can determine V_{z_S} from the first row of $P^O_{z_S}L_3^{-1}K$ because $P^O_{z_S}L_3^{-1}K = L_2'V_{z_S}K$ and the first row of L_2' is a vector of ones. Till now we have identified $\{\widetilde{\pi}^m_{z_S}\}_{m=1}^M$, $\{\lambda^m_{O,\eta_1}\}_{m=1}^M$ and $\{\{\lambda^m_{O,\eta_{t,j}}\}_{j=1}^{M-1}\}_{m=1}^M$ for $t = 2, 3$.

Next I show how to identify $D^O_{x_1}$ for any $x_1 \in \mathcal{X}_1$. Let's construct $P^{O*}_{z_S,x1}$ in the same way as $P^{O*}_{z_S,\eta_1}$. It follows that $D^O_{x_1} = (L_2'V_{x_1})^{-1}P^*_{O,x_1}L_3^{-1}$. So $\{\lambda^{*m}_{O,x_1}\}_{m=1}^M$ for any $x_1 \in \mathcal{X}_1$ is identified.

To identify $\{\lambda^m_{O,x_2}\}_{m=1}^M$ for any $x_2 \in \mathcal{X}_2$, construct the following matrices:

$$L^{x_2} = \begin{pmatrix} 1 & \lambda^1_{O,x_2} \\ \vdots & \vdots \\ 1 & \lambda^M_{O,x_2} \end{pmatrix},$$

and

$$P^{x_2} = \begin{pmatrix} 1 & F^O_{z,\eta_{3,1}} & \dots & F^O_{z,\eta_{3,M-1}} \\ F^O_{z,x_2} & F^O_{z,x_2,\eta_{3,1}} & \dots & F^O_{z,x_2,\eta_{3,M-1}} \end{pmatrix}.$$

P^{x_2} can be expressed as $P^{x_2} = (L^{x_2})'V_{z_S}L_3$. So $(L^{x_2})' = P^{x_2}(V_{z_S}L_3)^{-1}$. So $\{\lambda^m_{O,x_2}\}_{m=1}^M$ is identified. With similar approach $\{\lambda^m_{O,x_3}\}_{m=1}^M$ for any $x_3 \in X_3$ can also be identified.

To identify $V_{z'_S}$ for any $z'_S \in \mathcal{Z}_\mathcal{S}$, construct $P^O_{z'_S}$ by replacing z_S with z'_S in $P^O_{z_S}$. $P^O_{z'_S}$ can be expressed as $P^O_{z'_S} = L'_2 V_{z'_S} L_3$. Then $V_{z'_S} = (L'_2)^{-1} P^O_{z'_S} L_3^{-1}$ and $\{\widetilde{\pi}^m_{z'_S}\}_{m=1}^M$ for any $z'_S \in \mathcal{Z}_\mathcal{S}$ is identified. By integrating out S, we can get $\{\pi^m\}_{m=1}^M$ and $f^m(S|Z_S) = \widetilde{\pi}^m_{z'_S}/\pi^m$.

I have shown the identification of π^m, $f^m(S|Z_S)$,$f^m(O_1|X_1,S)$ and $f^m(O_t|X_t,S)$ for any $(\{X_t\}_{t=1}^3, S, Z_S) \in \prod_{t=1}^3 \mathcal{X}_t \times \mathcal{S} \times \mathcal{Z}_\mathcal{S}$. The rest is to show the identification of the type-specific wage marginal distributions. Define

$$\lambda^{*m}_{W,(w_1,x_1)} = f^m((W_1,O_1) = (w_1,1)|(X_1,S) = (\eta_1,s)),$$
$$D^W_{w,\eta_1} = diag(\lambda^{*1}_{W,(w,\eta_1)}, \dots, \lambda^{*M}_{W,(w,\eta_1)}).$$

Fix $W_1 = w_1$, $O_t = 1$ for $t = 1, 2, 3$, and define the following functions:

$$F^{W*}_{Z_S,X_1,X_2,X_3} = \sum_{m=1}^{M} \widetilde{\pi}^m_{Z_S} \lambda^{*m}_{W,(w_1,X_1)} \lambda^m_{O,X_2} \lambda^m_{O,X_3},$$

$$F^{W*}_{Z_S,X_1,X_2} = \sum_{m=1}^{M} \widetilde{\pi}^m_{Z_S} \lambda^{*m}_{W,(w_1,X_1)} \lambda^m_{O,X_2},$$

$$F^{W*}_{Z,X_1,X_3} = \sum_{m=1}^{M} \widetilde{\pi}^m_{Z_S} \lambda^{*m}_{W,(w_1,X_1)} \lambda^m_{O,X_3},$$

$$F^{W*}_{Z,X_1} = \sum_{m=1}^{M} \widetilde{\pi}^m_{Z_S} \lambda^{*m}_{W,(w_1,X_1)}.$$

Arrange these to an $M \times M$ matrix:

$$P^{W*}_{z_S,\eta_1} = \begin{pmatrix} F^{W*}_{z_S,\eta_1} & F^{W*}_{z_S,\eta_1,\eta_{3,1}} & \cdots & F^{W*}_{z_S,\eta_1,\eta_{3,M-1}} \\ F^{W*}_{z_S,\eta_1,\eta_{2,1}} & F^{W*}_{z_S,\eta_1,\eta_{2,1},\eta_{3,1}} & \cdots & F^{W*}_{z_S,\eta_1,\eta_{2,1},\eta_{2,M-1}} \\ \vdots & \vdots & \ddots & \vdots \\ F^{W*}_{z_S,\eta_1,\eta_{2,M-1}} & F^{W*}_{z_S,\eta_1,\eta_{2,M-1},\eta_{3,1}} & \cdots & F^{W*}_{z_S,\eta_1,\eta_{2,M-1},\eta_{3,M-1}} \end{pmatrix}.$$

$P^{W*}_{z_S,\eta_1} = L_2' V_{z_S} D^W_{w_1,\eta_1} L_3$. Then $D^W_{w_1,\eta_1} = (L_2' V_{z_S})^{-1} P^{W*}_{z_S,\eta_1} L_3^{-1}$ and $f^m(W_1, O_1|X_1, S)$ is identified. Further $f^m((W_1|O_1, X_1, S) = f^m(W_1, O_1|X_1), S)/f^m((O_1|X_1), S)$.

To identify $f^m(W_t|O_t, X_t, S)$ for $t = 2$, define

$$\lambda^m_{W,(w_2,\eta_2)} = f^m((W_2, O_2) = (w_2, 1)|(X_2, S) = (\eta_2, s)).$$

Fix $W_2 = w_2$, $O_t = 1$ for $t = 2, 3$, and define the following functions:

$$F^W_{Z_S,X_2,X_3} = \sum_{m=1}^{M} \widetilde{\pi}^m_{Z_S} \lambda^m_{W,(w_2,X_2)} \lambda^m_{O,X_3},$$

$$F^W_{Z_S,X_2} = \sum_{m=1}^{M} \widetilde{\pi}^m_{Z_S} \lambda^m_{W,(w_2,X_2)}.$$

Then construct the following matrices:

$$L^{w_2} = \begin{pmatrix} 1 & \lambda^1_{W,(w_2,\eta_2)} \\ \vdots & \vdots \\ 1 & \lambda^M_{W,(_2,\eta_2)} \end{pmatrix},$$

and

$$P^{w_2} = \begin{pmatrix} 1 & F^O_{z_S,\eta_{3,1}} & \dots & F^O_{z_S,\eta_{3,M-1}} \\ F^W_{z_S,x_2} & F^W_{z_S,\eta_2,\eta_{3,1}} & \dots & F^W_{z_S,\eta_2,\eta_{3,M-1}} \end{pmatrix}.$$

P^{w_2} can be expressed as $P^{w_2} = (L^{w_2})' V_{z_S} L_3$. Then $(L^{w_2})' = P^{w_2}(V_{z_S} L_3)^{-1}$ and $\{f^m(W_2, O_2|X_2, S)\}_{m=1}^M$ is identified and $f^m(W_2|O_2, X_2, S) = f^m(W_2, O_2|X_2, S)/f^m(O_2|X_2, S)$ for $m = 1, \dots, M$. With similar approach, $f^m(W_3|O_3, X_3, S)$ can be identified for $m = 1, \dots, M$. This completes the proof of Proposition 4.1. □

D.2 Assumptions and proof of Proposition 4.2

Assumption 3. *For m=1,...,M and* $t \geq 2$,

(a)

$$f^m_t(W_t|O_t, X_t, S, \{W_\tau, O_\tau, X_\tau\}_{\tau=2}^{t-1}) = f^m(W_t|O_t, X_t, S, \{W_\tau, O_\tau, X_\tau\}_{\tau=2}^{t-1}),$$

and

$$f^m_t(O_t|X_t, S, \{W_\tau, O_\tau, X_\tau\}_{\tau=2}^{t-1}) = f^m(O_t|X_t, S, \{W_\tau, O_\tau, X_\tau\}_{\tau=2}^{t-1}).$$

(b)

$$f^m(W_t|O_t, X_t, S, \{W_\tau, O_\tau, X_\tau\}_{\tau=2}^{t-1}) = f^m(W_t|O_t, S, X_t, W_{t-1}, O_{t-1}, X_{t-1}),$$

and

$$f^m(O_t|X_t, S, \{W_\tau, O_\tau, X_\tau\}_{\tau=2}^{t-1}) = f^m(O_t|O_{t-1}, S, X_t).$$

Under Assumption 3, the conditional joint distribution of wages, occupations, and education can be simplified as follows:

$$\begin{aligned}
&f^m(\{W_t, O_t\}_{t=1}^T, S|\{X_t\}_{t=1}^T, Z_S) \\
=&f^m(W_1|O_1, S)\prod_{t=2}^T f^m(W_t|O_t, S, X_t, W_{t-1}, O_{t-1}, X_{t-1}) \\
&\times f^m(O_1|S)\prod_{t=2}^T f^m(O_t|O_{t-1}, S, X_t) f^m(S|Z_S).
\end{aligned}$$

The transition process of (W_t, O_t, X_t) becomes a stationary first-order Markov process. Define $Y_t = (W_t, O_t, X_t)$. The variation of Y_t affects both the type-specific conditional joint distribution at period t and that at period $t+1$. This makes it difficult to construct factorization equations as before. To solve this problem, we look at every other period. Fix Y_t to be $\bar{y}_t$ for odd t and define

$$\begin{aligned}
&\widetilde{\pi}^m_{\bar{y},z_S} = \pi^m f^m(\bar{y}_1, s|z_S), \\
&\lambda^m_{\bar{y}}(Y_t) = f^m(\bar{y}_{t+1}|Y_t, s) f^m(Y_t|\bar{y}_{t-1}, s), \\
&\lambda^{*m}_{\bar{y}}(Y_T) = f^m(Y_T|\bar{y}_{T-1}, s).
\end{aligned}$$

Let ξ_t be element of $\mathcal{Y}_t$ and define

$$L_{t,\bar{y}} = \begin{pmatrix} 1 & \lambda^1_{\bar{y}}(\xi_{t,1}) & \dots & \lambda^1_{\bar{y}}(\xi_{t,M-1}) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \lambda^M_{\bar{y}}(\xi_{t,1}) & \dots & \lambda^M_{\bar{y}}(\xi_{t,M-1}) \end{pmatrix},$$

$$V_{\bar{y}} = diag(\widetilde{\pi}^1_{\bar{y},z_S}, \ldots, \widetilde{\pi}^M_{\bar{y},z_S}), \text{ and } D^O_{Y_T|\bar{y}} = diag(\lambda^{*1}_{\bar{y}}(Y_T), \ldots, \lambda^{*M}_{\bar{y}}(Y_T)).$$

Then construct

$$P^O_{\bar{y}} = L'_{2,\bar{y}} V_{\bar{y}} L_{4,\bar{y}},$$

$$P^{O*}_{\bar{y}} = L'_{2,\bar{y}} D^O_{Y_T|\bar{y}} V_{\bar{y}} L_{4,\bar{y}}.$$

Further, assume

Assumption 4. *There exist some $\{\xi_{t,1}, \ldots, \xi_{t,M-1}\}_{t=1}^T$ such that $P^O_{\bar{y}}$ is of full rank and that all the eigenvalues of $(P^O_{\bar{y}})^{-1} P^{O*}_{\bar{y}}$ take distinct values.*

Proof of Proposition 4.2. Without loss of generality, set $T = 6$. Fix $(Y_1, Y_3, Y_5) = (y_1, y_2, y_5)$ and define

$$F^{*O}_{Y_2,Y_4,Y_6} = \sum_{m=1}^M \widetilde{\pi}_{\bar{y},Z_S} \lambda^m_{\bar{y}}(Y_2) \lambda^m_{\bar{y}}(Y_4) \lambda^{*m}_{\bar{y}}(Y_T),$$

$$F^{*O}_{Y_2,Y_6} = \sum_{m=1}^M \widetilde{\pi}_{\bar{y},z_S} \lambda^m_{\bar{y}}(Y_2) \lambda^{*m}_{\bar{y}}(Y_T),$$

$$F^{*O}_{Y_6} = \sum_{m=1}^M \widetilde{\pi}_{\bar{y},z_S} \lambda^{*m}_{\bar{y}}(Y_T),$$

$$F^O_{Y_2,Y_4} = \sum_{m=1}^M \widetilde{\pi}_{\bar{y},z_S} \lambda^m_{\bar{y}}(Y_2) \lambda^m_{\bar{y}}(Y_4),$$

$$F^O_{Y_2} = \sum_{m=1}^M \widetilde{\pi}_{\bar{;},z_S} \lambda^m_{\bar{y}}(Y_2),$$

$$F^O = \sum_{m=1}^M \widetilde{\pi}_{\bar{y},z_S}.$$

And construct matrices as follows:

$$P_{\bar{y}}^{O} = \begin{pmatrix} F^{O} & F^{O}_{\xi_{4,1}} & \cdots & F^{O}_{\xi_{4,M-1}} \\ F^{O}_{\xi_{2,1}} & F^{O}_{\xi_{2,1},\xi_{4,1}} & \cdots & F^{O}_{\xi_{2,1},\xi_{4,M-1}} \\ \vdots & \vdots & \ddots & \vdots \\ F^{O}_{\xi_{2,M-1}} & F^{O}_{\xi_{2,M-1},\xi_{4,1}} & \cdots & F^{O}_{\xi_{2,M-1},\xi_{4,M-1}} \end{pmatrix},$$

and

$$P_{\bar{y}}^{O*} = \begin{pmatrix} F^{O*}_{\xi_6} & F^{O*}_{\xi_{4,1},\xi_6} & \cdots & F^{O*}_{\xi_{4,M-1},\xi_6} \\ F^{O*}_{\xi_{2,1},\xi_6} & F^{O*}_{\xi_{2,1},\xi_{4,1},\xi_6} & \cdots & F^{O*}_{\xi_{2,1},\xi_{4,M-1},\xi_6} \\ \vdots & \vdots & \ddots & \vdots \\ F^{O*}_{\xi_{2,M-1},\xi_6} & F^{O*}_{\xi_{2,M-1},\xi_{4,1},\xi_6} & \cdots & F^{O*}_{\xi_{2,M-1},\xi_{4,M-1},\xi_6} \end{pmatrix}.$$

Then repeat the argument of the proof of Proposition 4.1 and we achieve the identification of $\widetilde{\pi}^{m}_{\bar{y},z_S}$, $\lambda^{m}_{\bar{y}}(\xi_t)$, and $\lambda^{*m}_{\bar{y}}(Y_T)$. Then integrate out the other elements and apply Bayes' rule, we can get π^m, $f^m(W_1|O_1, X_1, S)$, $f^m(O_1|X_1, S)$, $f^m(S|Z_S)$, $f^m(W_t|O_t, S, X_t, W_{t-1}, O_{t-1}, X_{t-1})$, and $f^m(O_t|O_{t-1}, S, X_t)$.

$\square$

D.3 Assumptions and proof of Proposition 4.3

Denote the support of Q_1, Q_2, and Q_3 by $\mathcal{Q}_1$, $\mathcal{Q}_2$, and $\mathcal{Q}_3$ respectively. Partition $\mathcal{Q}_1$ into M mutually exclusive and exhaustive subsets and denote the partitions as $\triangle_{Q_1} = \{\delta^1_{Q_1}, \ldots, \delta^M_{Q_1}\}$. Similarly denote the partitions of $\mathcal{Q}_2$ as $\triangle_{Q_2} = \{\delta^1_{Q_2}, \ldots, \delta^M_{Q_2}\}$. Let $\triangle = \triangle_{Q_1} \times \triangle_{Q_2}$. Also partition $\mathcal{Q}_3$ into 2 mutually exclusive and exhaustive subsets as $\triangle_{Q_3} = \{\delta^1_{Q_3}, \delta^2_{Q_3}\}$.

Let's define

$$p^m_{Q_1} = (P^m(Q_1 \in \delta^1_{Q_1}|s, z_s), \ldots, P^m(Q_1 \in \delta^M_{Q_1})|s, z_s)',$$

$$p^m_{Q_2} = (P^m(Q_2 \in \delta^1_{Q_2}|s, z_s), \ldots, P^m(Q_2 \in \delta^M_{Q_2})|s, z_s)',$$

$$p^m_{Q_3}(h) = P^m(Q_3 \in \delta^h_{Q_3}|s, z_s),$$

$$\widetilde{\pi}^m = \pi^m f^m(s, z_s).$$

Collect the type-specific distributions into following matrices

$$L_{Q_1} = (p^1_{Q_1}, \ldots, p^M_{Q_1}),$$

$$L_{Q_2} = (p^1_{Q_2}, \ldots, p^M_{Q_2}),$$

$$V = diag(\widetilde{\pi}^1, \ldots, \widetilde{\pi}^M),$$

and

$$D_h = diag(p^1_{Q_3}(h), \ldots, p^M_{Q_3}(h)).$$

Let $P_s(Q_1 \in \delta^m_{Q_1}, Q_2 \in \delta^{m'}_{Q_2})$ be the probability that $Q_1 \in \delta^m_{Q_1}$ and $Q_2 \in \delta^{m'}_{Q_2}$ for $S = s$ and $P_s(Q_1 \in \delta^m_{Q_1}, Q_2 \in \delta^{m'}_{Q_2}, Q_3 \in \delta^h_{Q_3})$ be the probability that $Q_1 \in \delta^m_{Q_1}$, $Q_2 \in \delta^{m'}_{Q_2}$, and $Q_3 \in \delta^h_{Q_3}$ for $S = s$. Define two $M \times M$ matrices as follows:

$$P_\triangle = \begin{pmatrix} P_s(Q_1 \in \delta^1_{Q_1}, Q_2 \in \delta^1_{Q_2}) & \ldots & P_s(Q_1 \in \delta^1_{Q_1}, Q_2 \in \delta^M_{Q_2}) \\ \vdots & \ldots & \vdots \\ P_s(Q_1 \in \delta^M_{Q_1}, Q_2 \in \delta^1_{Q_2}) & \ldots & P_s(Q_1 \in \delta^M_{Q_1}, Q_2 \in \delta^M_{Q_2}) \end{pmatrix},$$

$$P_{\triangle,h} = \begin{pmatrix} P_s(Q_1 \in \delta^1_{Q_1}, Q_2 \in \delta^1_{Q_2}, Q_3 \in \delta^h_{Q_3}) & \dots & P_s(Q_1 \in \delta^1_{Q_1}, Q_2 \in \delta^M_{Q_2}, Q_3 \in \delta^h_{Q_3}) \\ \vdots & \dots & \vdots \\ P_s(Q_1 \in \delta^M_{Q_1}, Q_2 \in \delta^1_{Q_2}, Q_3 \in \delta^h_{Q_3}) & \dots & P_s(Q_1 \in \delta^M_{Q_1}, Q_2 \in \delta^M_{Q_2}, Q_3 \in \delta^h_{Q_3}) \end{pmatrix}.$$

Assume:

Assumption 5. *There exists a partition $\triangle \times \triangle_{Q_3}$ on the variables (Q_1, Q_2, Q_3) for which the matrix $P_\triangle$ is nonsingular and the eigenvalues of $P_{\triangle,h}P_\triangle^{-1}$ are distinct for partition level $h = 1$ of the variable Q_3.*

Proof of Proposition 4.3. $P_\triangle$ and $P^*_{\triangle,h}$ can be expressed as the follows:

$$P_\triangle = L_{Q_1}V(L'_{Q_2}), \text{ and } P_{\triangle,h} = L_{Q_1}D_hV(L'_{Q_2}).$$

Since $P_\triangle$ is nonsingular, both L_{Q_1} and L_{Q_2} are nonsingular. Construct $A_h = P_{\triangle,h}P_\triangle^{-1} = L_{Q_1}D_hL_{Q_1}^{-1}$, and we have $A_hL_{Q_1} = L_{Q_1}D_h$. The distinct eigenvalues of A_h determines the elements of D_h, and its eigenvectors determine the columns of L_{Q_1} uniquely up to a multiplicative constant. Then L_{Q_1} is uniquely determined since the elements of each column of L_{Q_1} must sum to one. Construct $B_h = (P'_{\triangle,h})(P'_\triangle)^{-1} = L_{Q_2}D_hL_{Q_2}^{-1}$, and L_{Q_2} is determined using the similar argument. Once L_{Q_1} and L_{Q_2} are determined, V is uniquely determined by $V = (L_{Q_1})^{-1}P_\triangle(L'_{Q_2})^{-1}$. Then $\{\pi^m\}_{m=1}^M$ is identified by integrating out S and Z_S, and $f^m(S|Z_S) = \widetilde{\pi}^m/(\pi^m f(Z_S))$.

For any $q_1 \in \mathcal{Q}_1$, denote $p_{q_1} = (P^1_{Q_1}(q_1), \dots, P^M_{Q_1}(q_1))$ and define $P_{q_1,\triangle Q_2} = p_{q_1}V(L_{Q_2})'$. Then $p_{q_1} = P_{q_1,\triangle Q_2}(V(L_{Q_2})')^{-1}$, and $\{P^M_{Q_1}(q_1)\}_{m=1}^M$ is identified. Define $P_{\triangle Q_1,q_2}$ and $P_{\triangle Q_1,q_3}$ analogously and apply the same argument, $\{P^M_{Q_2}(q_2), P^M_{Q_3}(q_3)\}_{m=1}^M$ are identified.

$\square$

D.4 Assumptions and proof of Proposition 4.4

Denote $p^m_{O_t} = P^m(O_t = 1|(S, Z_S) = (s, z_S))$, and $D_{O_t} = diag(p^1_{O_t}, \ldots, p^M_{O_t})$. Construct an $M \times M$ matrix

$$P_{\triangle,O_t} = \begin{pmatrix} P(Q_1 \in \delta^1_{Q_1}, Q_2 \in \delta^1_{Q_2}, O_t = 1) & \ldots & P(Q_1 \in \delta^1_{Q_1}, Q_2 \in \delta^M_{Q_2}, O_t = 1) \\ \vdots & \ldots & \vdots \\ P(Q_1 \in \delta^M_{Q_1}, Q_2 \in \delta^1_{Q_2}, O_t = 1) & \ldots & P(Q_1 \in \delta^M_{Q_1}, Q_2 \in \delta^M_{Q_2}, O_t = 1) \end{pmatrix},$$

Assume

Assumption 6. *The eigenvalues of $P_{\triangle,O_t} P^{-1}_{\triangle}$ are distinct.*

Proof of Proposition 4.4. The proof of the nonparametric identification of education psychic costs using test scores is already shown in the proof of Proposition 4.3. Below, I prove the nonparametric identification of occupation abilities using test scores.

Express $P_{\triangle,O_t}$ as $P_{\triangle,O_t} = L_{Q_1} D_{O_t} V L'_{Q_2}$. Replacing $P_{\triangle,h}$ in the proof of Proposition 4.3 by $P_{\triangle,O_t}$, and repeating the proof, π^m, $f^m(S|Z_S)$, and $f^m(O_t|X_t, S)$ are identified.

Next, denote $p^m_{W_t} = F^m((W_t, O_t) = (w_t, 1)|(S, Z_S) = (s, z_S))$ and $D_{W_t} = diag(p^1_{W_t}, \ldots, p^M_{W_t})$. Let $P(Q_1 \in \delta^m_{Q_1}, Q_2 \in \delta^{m'}_{Q_2}, (\omega_t, 1))$ be the probability that $Q_1 \in \delta^m_{Q_1}$, $Q_2 \in \delta^{m'}_{Q_2}$, $W_t = \omega_t$, and $O_t = 1$ for $S = s$. The corresponding $M \times M$ matrix is

$$P_{\triangle,w_t} = \begin{pmatrix} P_S(Q_1 \in \delta^1_{Q_1}, Q_2 \in \delta^1_{Q_2}, (\omega_t, 1)) & \ldots & P_S(Q_1 \in \delta^1_{Q_1}, Q_2 \in \delta^M_{Q_2}, (\omega_t, 1)) \\ \vdots & \ldots & \vdots \\ P_S(Q_1 \in \delta^M_{Q_1}, Q_2 \in \delta^1_{Q_2}, (\omega_t, 1)) & \ldots & P_S(Q_1 \in \delta^M_{Q_1}, Q_2 \in \delta^M_{Q_2}, (\omega_t, 1)) \end{pmatrix}.$$

$P_{\triangle,w_t} = L_{Q_1} D_{w_t} V L'_{Q_2}$. Then $D_{w_t} = L^{-1}_{Q_1} P_{\triangle,w_t} (V L'_{Q_2})^{-1}$, and $f^m(W_t, O_t|X_t, S)$ is identified. By Bayes' rule, $f^m(W_t|O_t, X_t, S) = f^m(W_t, O_t|X_t, S)/f^m(O_t|X_t, S)$.

□

E EM Algorithm

Consider $(k+1)$th iteration. In E step, calculate the expectated log-likelihood ϕ based on the estimates from the kth iteration:

$$\phi^{(k)} = \sum_{i=1}^{n} \sum_{m=1}^{M} \mu_i^{m(k)} (log\pi^m + log\mathcal{L}_W^m + log\mathcal{L}_O^m + log\mathcal{L}_S^m + log\mathcal{L}_Q^m),$$

where

$$\mu_i^{m(k)} = \frac{\pi^{m(k)} \mathcal{L}_W^{m(k)} \mathcal{L}_O^{m(k)} \mathcal{L}_S^{m(k)} \mathcal{L}_Q^{m(k)}}{\sum_{m=1}^{M} \pi^{m(k)} \mathcal{L}_W^{m(k)} \mathcal{L}_O^{m(k)} \mathcal{L}_S^{m(k)} \mathcal{L}_Q^{m(k)}}.$$

In M step, compute the parameters by maximizing the expected log-likelihood ϕ:

$\pi^{m(k+1)}$ satisfies $\frac{\partial \phi^k}{\partial \pi^{m(k+1)}} = 0$. Correspondingly,

$$\pi^{m(k+1)} = \frac{\sum_{i=1}^{n} \mu_i \pi^{m(k)}}{n}.$$

$\beta_W^{m(k+1)}$ satisfies $\frac{\partial \phi^k}{\partial \beta_W^{m(k+1)}} = 0$. And it can be simplified to

$$\frac{\partial \sum_{i=1}^{n} \sum_{t=1}^{T} \mathcal{L}_W^{m(k+1)}}{\partial \beta_W^{m(k+1)}} = 0,$$

which is an OLS regression.

$\gamma_O^{m(k+1)}$ satisfies $\frac{\partial \phi^k}{\partial \gamma_O^{m(k+1)}} = 0$, and it can be simplified to

$$\frac{\partial \sum_{i=1}^{n} \sum_{t=1}^{T} \mathcal{L}_O^{m(k+1)}}{\partial \gamma_O^{m(k+1)}} = 0.$$

$(\theta_R^{m(k+1)}$ satisfies $\frac{\partial \phi^k}{\partial \theta_R^{m(k+1)}} = 0$. And it can be simplified to

$$\frac{\partial \sum_{i=1}^{n} \sum_{r=1}^{R} \mathcal{L}_Q^{m(k+1)}}{\partial \theta_R^{m(k+1)}} = 0,$$

which is a probit.

$\delta_S^{m(k+1)}$ satisfies $\frac{\partial \phi^k}{\partial \delta_S^{m(k+1)}} = 0$, and it can be simplified to

$$\frac{\partial \sum_{i=1}^{n} \mathcal{L}_S^{m(k+1)}}{\partial \delta_S^{m(k+1)}} = 0,$$

which is a multinomial logit.

F Choice of Initial Values

The strategy is to start with estimating the parameters in Equation (7) when the population is homogenous (M=1) and then add one more type at a time and re-estimate the parameters. Let $\mathcal{L}_i^m$ denote the likelihood for individual i and define

$$\mu^m = \sum_{i=1}^{n} (1 - \frac{\mathcal{L}_i^m}{\sum_{k=1}^{m-1} L_i^k \pi^k})$$

The estimation follows the algorithm as below:

(a) Set $M = 1$ and $\pi^1 = 1$. Choose initial values for α_W, α_O, α_S, α_R, β, λ, δ, θ, σ_W, σ_Q, and ρ in Equation (7).

(b) Given the current value of M, maximize the likelihood over α_W, α_O, α_S, α_R, β, λ, δ, θ, σ_W, σ_Q, ρ, and π^m.

(c) Evaluate μ^{M+1} for a grid of values of the type-specific parameters.

(d) Set the type-specific parameters to the values that yield the smallest value for μ^{M+1}.

(e) Maximize the likelihood. Increase the value of M by 1. Return to Step (b).

References

Belzil, C. and Hansen, J. (2002). Unobserved ability and the return to schooling. *Econometrica*, 70:2075–2091.

Belzil, C. and Hansen, J. (2007). A structural analysis of correlated random coefficent wage regression model. *Journal of Econometrics*, 140:827–848.

Blaug, M. (1985). Where are we now in the economics of education. *Economics of Education Review*, 4(1):17–28.

Card, D. (2001). Estimating the return to schooling: Progress on some persistent econometric problems. *Econometrica*, 69(5):1127–1160.

Carneiro, P., Hansen, K. T., and Heckman, J. J. (2003). Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice. *International Economic Review*, 44:361–422.

Cunha, F. and Heckman, J. J. (2008). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of Human Resources*, 43(4):738–782.

Cunha, F., Heckman, J. J., and Schennach, S. M. (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, 78(3):883–931.

Dempster, A., Laird, N., and Rubin, D. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society.Series B (Methodological)*, 39(1):746–773.

Dickens, W. T. and Lang, K. (1985). A test of dual labor market theory. *The American Economic Review*, 75(4):792–805.

Duncan, G. and Homan, S. (1981). The incidence and wage effects of overeducation. *Economics of Education Review*, 1(1):75–86.

Grubb, W. N. (1997). The returns to education in the sub-baccalaureate labor market, 1984-1990. *Economics of Education Review*, 16(3):231–245.

Hansen, K., Heckman, J. J., and Mullen, K. J. (2004). The effects of schooling and ability on achievement test scores. *Journal of Econometrics*, 121(1-2):39–98.

Heckman, J. J., Lochner, L. J., and Todd, P. E. (2006a). Earnings functions, rates of return and treatment effects: The mincer equation and beyond. I:307–458.

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

Heckman, J. J., Stixrud, J., and Urzua, S. (2006b). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labour Economics*, 24(3):746–773.

Hoffmann, F. (2011). An empirical model of life-cycle earnings and mobility dynamics. working paper.

Jaeger, D. A. and Page, M. E. (1996). Degrees matter: New evidence on sheepskin effects in the returns to education. *The Review of Economics and Statistics*, 78(4):733–740.

Kane, T. J. and Rouse, C. E. (1995). Labor market returns to two- and four-year college. *American Economic Review*, 85(3):600–614.

Kasahara, H. and Shimotsu, K. (2009). Nonparametric identification of finite mixture models of dynamic discrete choices. *Econometrica*, 77:135–175.

Kasahara, H. and Shimotsu, K. (2012). Nonparametric identification of multivariate mixtures. Discussion papers 2010-09, Graduate School of Economics, Hitotsubashi University.

Keane, M. P. and Wolpin, K. I. (1997). The career decisions of young men. *Journal of Political Economy*, 105:473–522.

Light, A. and Strayer, W. (2004). Who receives the college wage premium? assessing the labor market returns to degrees and college transfer patterns. *The Journal of Human Resources*, 39(3):411–482.

Marcotte, D. E., Bailey, T., Borkoski, C., and Kienzl, G. S. (2005). The returns of a community college education: Evidence from the national education longitudinal survey. *Educational Evaluation and Policy Analysis*, 27(2):157–175.

Roy, A. D. (1951). Some thoughts on the distribution of earnings. *Oxford Economic Papers*, 3(2):135–146.

Sicherman, N. (1991). "overeducation" in the labor market. *Journal of Labor Economics*, 9(2):101–122.

Willis, R. J. (1986). Wage determinants: A survey and reinterpretation of human capital earnings functions. I:525–602.

Willis, R. J. and Rosen, S. (1979). Education and self-selection. *Journal of Political Economy*, 87:S7–S36.

Table 1: Discriptive Statistics

	Overall			High School			2-yr College			4-yr College		
Variables	Obs	Mean	S.D	Obs	Mean	S.D	Obs	Mean	S.D	Obs	Mean	S.D
2-yr college attendant	934	0.175	0.380	318	0.000	0.000	163	1.000	0.000	453	0.000	0.000
4-yr college attendant	934	0.485	0.500	318	0.000	0.000	163	0.000	0.000	453	1.000	0.000
Highest grade completed	934	13.943	2.249	318	11.880	0.567	163	13.074	1.034	453	15.700	1.864
Age in 1979	934	17.217	2.082	318	16.447	1.628	163	17.362	2.033	453	17.706	2.223
Initial job(white collar)	934	0.394	0.489	318	0.110	0.313	163	0.264	0.442	453	0.640	0.480
Initial wage	934	11.603	15.738	318	9.879	22.438	163	10.795	5.004	453	13.103	12.024
Initial wage(blue collar)	934	10.249	17.129	318	10.060	23.754	163	10.910	5.234	453	10.089	4.535
Initial wage(white collar)	934	13.684	13.068	318	8.410	3.431	163	10.472	4.340	453	14.797	14.374
Mother education	934	12.269	2.086	318	11.355	1.835	163	12.202	1.919	453	12.934	2.066
Father education	934	12.726	3.038	318	11.226	2.469	163	12.748	2.604	453	13.770	3.111
Number of siblings	934	2.767	1.748	318	3.110	1.850	163	2.755	1.757	453	2.530	1.631
Broken family at age 14	934	0.127	0.334	318	0.151	0.359	163	0.147	0.355	453	0.104	0.305
South at age 14	934	0.239	0.427	318	0.239	0.427	163	0.221	0.416	453	0.245	0.431
Urban at age 14	934	0.730	0.444	318	0.619	0.486	163	0.791	0.408	453	0.786	0.411

Table 2: Test Scores by Education and Initial Occupation

(a) By Intial Occupation

	Blue Collar			White Collar		
Variable	Obs	Mean	S.D.	Obs	Mean	S.D.
Math skill	566	-0.323	0.961	368	0.497	0.844
Verbal skill	566	-0.279	1.066	368	0.428	0.700
Coding speed	566	-0.231	0.966	368	0.356	0.947
Mechanical	566	-0.065	1.044	368	0.101	0.920
Locus of control	566	-0.099	0.982	368	0.153	1.009
Self-esteem	566	-0.110	1.000	368	0.169	0.978

(b) By Education

	High School			2-yr College			4-yr College		
Variable	Obs	Mean	S.D.	Obs	Mean	S.D.	Obs	Mean	S.D.
Math skill	318	-0.684	0.814	163	-0.236	0.870	453	0.565	0.812
Verbal skill	318	-0.672	1.107	163	-0.068	0.867	453	0.497	0.607
Coding speed	318	-0.430	0.911	163	-0.099	0.955	453	0.338	0.953
Mechanical	318	-0.248	1.096	163	0.049	0.969	453	0.156	0.903
Locus of control	318	-0.256	0.956	163	-0.017	1.026	453	0.186	0.982
Self-esteem	318	-0.325	0.947	163	0.058	0.943	453	0.207	0.999

(c) By Education, Blue-Collar

	High School			2-yr College			4-yr College		
Variable	Obs	Mean	S.D.	Obs	Mean	S.D.	Obs	Mean	S.D.
Math skill	283	-0.708	0.801	120	-0.255	0.925	163	0.294	0.910
Verbal skill	283	-0.734	1.105	120	-0.071	0.855	163	0.359	0.705
Coding speed	283	-0.476	0.916	120	-0.126	0.965	163	0.116	0.934
Mechanical	283	-0.257	1.097	120	0.063	0.959	163	0.294	0.910
Locus of control	283	-0.263	0.971	120	0.006	1.001	163	0.108	0.943
Self-esteem	283	-0.348	0.943	120	0.071	0.969	163	0.171	1.024

(d) By Education, White-Collar

	High School			2-yr College			4-yr College		
Variable	Obs	Mean	S.D.	Obs	Mean	S.D.	Obs	Mean	S.D.
Math skill	35	-0.495	0.899	43	-0.181	0.698	290	0.718	0.708
Verbal skill	35	-0.178	1.004	43	-0.061	0.908	290	0.574	0.531
Coding speed	35	-0.061	0.783	43	-0.022	0.933	290	0.462	0.942
Mechanical	35	-0.178	1.108	43	0.011	1.009	290	0.718	0.708
Locus of control	35	-0.204	0.834	43	-0.079	1.101	290	0.23	1.003
Self-esteem	35	-0.139	0.975	43	0.021	0.877	290	0.228	0.986

Table 3: Estimated Test Scores Parameters (Equation (4))

	Math	Verbal	Coding	Mechanic	Self-control	Self-esteem
Constant						
Type 1	-0.021	-1.122 ***	-1.315 ***	-2.430 ***	-1.081 ***	-1.548 ***
	(0.315)	(0.285)	(0.404)	(0.367)	(-0.446)	(0.474)
Deviation of type 2 from type 1	-1.271 ***	-0.868 ***	-0.742 ***	-0.734 ***	-0.380 ***	-0.140 ***
	(0.074)	(0.078)	(0.113)	(0.099)	(-0.122)	(0.120)
Deviation of type 3 from type 1	-2.052 ***	-2.187 ***	-1.445 ***	-1.661 ***	-0.588 ***	-0.535 **
	(0.096)	(0.076)	(0.104)	(0.110)	(-0.128)	(0.129)
Deviation of type 4 from type 1	-0.622 ***	-0.271 ***	-0.528 ***	-0.226 ***	-0.410 ***	-0.257 ***
	(0.067)	(0.076)	(0.089)	(0.086)	(-0.100)	(0.096)
2-year college	0.019 ***	0.038 ***	0.019	0.030	0.005	0.005
	(0.014)	(0.014)	(0.017)	(0.017)	(-0.021)	(0.022)
4-year college	0.041 ***	0.024 ***	0.018 ***	0.017 **	0.006 **	0.020 **
	(0.009)	(0.010)	(0.012)	(0.012)	(-0.015)	(0.014)
Mother education	0.012 **	-0.027 ***	-0.012	-0.010 *	-0.010	-0.033 *
	(0.013)	(0.013)	(0.019)	(0.018)	(-0.020)	(0.020)
Father education	-0.064 ***	-0.090 ***	0.115 ***	-0.092	-0.025	0.003
	(0.068)	(0.069)	(0.102)	(0.093)	(-0.100)	(0.113)
#siblings	-0.034	0.033 ***	0.042	-0.146	0.048	0.003
	(0.055)	(0.056)	(0.074)	(0.067)	(-0.082)	(0.083)
Broken family at age 14	0.008 *	0.013	-0.005	-0.186 *	-0.035	0.029
	(0.051)	(0.053)	(0.075)	(0.070)	(-0.087)	(0.079)
South at age 14	-0.012	0.044 *	0.067	0.134 **	0.072	0.083
	(0.013)	(0.013)	(0.019)	(0.017)	(-0.023)	(0.024)
Urban at age 14	0.278	0.168	0.129	-0.052 ***	0.224	0.104
	(0.076)	(0.082)	(0.110)	(0.097)	(-0.15)	(0.156)
Age	0.542 **	0.305 ***	0.210 ***	-0.286 ***	0.145 ***	0.210 ***
	(0.066)	(0.070)	(0.091)	(0.086)	(-0.121)	(0.122)

Dependent variable: math, verbal, coding speed, mechanical comprehension, locus of control, and self-esteem scores. [a]

Type 1 and type 2 have the same occupation abilities. So do type 3 and type 4.

Type 1 and type 2 have different education psychic costs. So do type 3 and type 4.

Standard errors are in parenthesis.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

[a] All test scores are normalized to mean zero and standard deviation one.

Table 4: Estimated Wage Parameters (Equation (1))

	Blue Collar	White Collar
Constants		
Type 1 & 2	6.932 ***	7.044 ***
	(0.021)	(0.033)
Type 3 & 4	6.485 ***	6.457 ***
	(0.021)	(0.038)
2-year college	0.243 ***	0.170 ***
	(0.024)	(0.036)
4-year college	0.231 ***	0.296 ***
	(0.022)	(0.031)
Blue-collar experience	0.078 ***	0.033 ***
	(0.008)	(0.009)
Blue-collar experience squared	-0.285 ***	-0.076
	(0.069)	(0.075)
White-collar experience	0.040 ***	0.079 ***
	(0.012)	(0.008)
White-collar experience squared	-0.044	-0.191 **
	(0.169)	(0.084)
Hypothesis testing		p-value
blue-collar return=white-collar return, 2-year college		0.016
blue-collar return=white-collar return, 4-year college		0.018
2-year college return=4-year college return, blue-collar		0.311
2-year college return=4-year college return, white-collar		0.000

Dependent variable: log hourly salary

Type 1 and type 2 have the same occupation abilities. So do type 3 and type 4.

Type 1 and type 2 have different education psychic costs. So do type 3 and type 4.

Standard errors are in parenthesis.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 5: Estimated Average Partial Effects, Occupation Choice

	Initial job	Subsequent jobs
Constant		
Type 1 & 2	-0.246 ***	-0.137 **
	(0.057)	(0.071)
Deviation of type 3 & 4 from type 1 & 2	-0.162 ***	-0.056 **
	(0.063)	(0.027)
2-year college	0.159 ***	0.042 **
	(0.044)	(0.019)
4-year college	0.503 ***	0.122 ***
	(0.028)	(0.046)
Blue-collar experience		-0.044
		(0.193)
Blue-collar experience squared		0.262 **
		(0.128)
White-collar experience		0.054
		(0.120)
White-collar experience squared		-0.266 **
		(0.137)
White-collar job in the last period		0.287 ***
		(0.079)

The average partial effect is calculated as the average of the partial effect of each individual.

Type 1 and type 2 have the same occupation abilities. So do type 3 and type 4.

Type 1 and type 2 have different education psychic costs. So do type 3 and type 4.

Standard errors are in parenthesis.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 6: Estimated Average Partial Effects, Educational Choice

	2-year College	4-year College
Constant		
Type 1	-0.028	-0.293
	(0.179)	(0.283)
Deviation of type 2 from type 1	0.083	-0.416 ***
	(0.828)	(0.11)
Deviation of type 3 from type 1	0.092	-0.525 ***
	(0.112)	(0.116)
Deviation of type 4 from type 1	0.094	-0.204 ***
	(0.792)	(0.085)
Mother education	-0.004	0.038
	(0.645)	(0.078)
Father education	0.004	0.031
	(0.735)	(0.087)
#siblings	-0.001	-0.021
	(0.103)	(0.086)
Broken family at age 14	0.032	-0.066
	(0.063)	(0.126)
South at age 14	-0.013	0.077 **
	(0.059)	(0.033)
Urban at age 14	0.059	0.044
	(0.070)	(0.043)

The average partial effect is calculated as the average of the partial effect of each individual.

Type 1 and type 2 have the same occupation abilities. So do type 3 and type 4.

Type 1 and type 2 have different education psychic costs. So do type 3 and type 4.

Standard errors are in parenthesis.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 7: Occupation-Specific Returns

	(1)	(2)	(3)	(4)
2-year college × blue-collar	0.224 ***	0.188 ***	0.277 ***	0.243 ***
	(0.024)	(0.023)	(0.024)	(0.024)
2-year college × white-collar	0.136 ***	0.105 ***	0.192 ***	0.170 ***
	(0.034)	(0.031)	(0.035)	(0.036)
4-year college × blue-collar	0.246 ***	0.172 ***	0.237 ***	0.231 ***
	(0.022)	(0.023)	(0.022)	(0.022)
4-year college × white-collar	0.295 ***	0.227 ***	0.293 ***	0.296 ***
	(0.029)	(0.027)	(0.030)	(0.031)

Column (1): OLS estimates of the occupation-specific returns to education

Column (2): OLS estimates of the occupation-specific returns to education when six test scores are included as proxies for occupation abilities

Column (3): estimates of the occupation-specific returns to education when controlling for occupation abilities only

Column (4): estimates of the occupation-specific returns to education when controlling for both occupation abilities and education psychic costs

All the regressors in Equation (1) are included.

Standard errors are in parenthesis.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 8: Testing Conditional Independence of Wages and Test Scores

	(1)	(2)	(3)	(4)	(5)	(6)
2-year college × blue-collar	0.203 ***	0.275 ***	0.250 ***	0.235 ***	0.243 ***	0.240 ***
	(0.024)	(0.025)	(0.025)	(0.024)	(0.024)	(0.024)
2-year college × white-collar	0.123 ***	0.198 ***	0.176 ***	0.163 ***	0.170 ***	0.165 ***
	(0.037)	(0.034)	(0.035)	(0.037)	(0.036)	(0.036)
4-year college × blue-collar	0.185 ***	0.291 ***	0.239 ***	0.205 ***	0.231 ***	0.227 ***
	(0.024)	(0.023)	(0.023)	(0.022)	(0.022)	(0.022)
4-year college × white-collar	0.241 ***	0.357 ***	0.305 ***	0.267 ***	0.296 ***	0.291 ***
	(0.034)	(0.031)	(0.032)	(0.032)	(0.031)	(0.031)
Math	0.027 ***					
	(0.010)					
Verbal		-0.044 ***				
		(0.008)				
Coding			-0.009			
			(0.008)			
Mechanic				0.025 ***		
				(0.007)		
Locus of control					0.0003	
					(0.008)	
Self-esteem						0.008
						(0.008)

All the regressors in Equation (1) are included.

Standard errors are in parenthesis.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 9: Estimated Wage Parameters, with Three Test Scores in the Wage Equation

	Blue Collar	White Collar
Constants		
Type 1 & 2	6.914 ***	7.048 ***
	(0.021)	(0.032)
Type 3 & 4	6.461 ***	6.443 ***
	(0.023)	(0.038)
2-year college	0.222 ***	0.121 ***
	(0.023)	(0.034)
4-year college	0.228 ***	0.273 ***
	(0.024)	(0.032)
Blue-collar experience	0.075 ***	0.034 ***
	(0.008)	(0.009)
Blue-collar experience squared	-0.278 ***	-0.125 **
	(0.069)	(0.074)
White-collar experience	0.050 ***	0.080 ***
	(0.012)	(0.009)
White-collar experience squared	-0.158	-0.205 ***
	(0.161)	(0.084)
Test Scores		
Math	0.085 ***	
	(0.013)	
Verbal	-0.137 ***	
	(0.012)	
Mechanical comprehension	0.068 ***	
	(0.009)	

Dependent variable: log hourly salary

Type 1 and type 2 have the same occupation abilities. So do type 3 and type 4.

Type 1 and type 2 have different education psychic costs. So do type 3 and type 4.

The coefficients on the three test scores are restricted to be the same across occupations.

Standard errors are in parenthesis.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 10: Estimated Wage Parameters in the Wage Equation (6 types)

	Blue Collar	White Collar
Constants		
Type 1 & 2	6.998 ***	7.116 ***
	(0.027)	(0.034)
Type 3 & 4	6.501 ***	6.439 ***
	(0.026)	(0.041)
Type 5 & 6	6.609 ***	6.679 ***
	(0.021)	(0.037)
2-year college	0.235 ***	0.126 ***
	(0.023)	(0.036)
4-year college	0.238 ***	0.283 ***
	(0.021)	(0.032)
Blue-collar experience	0.076 ***	0.03 ***
	(0.008)	(0.009)
Blue-collar experience squared	-0.284 ***	-0.063
	(0.07)	(0.078)
White-collar experience	0.042 ***	0.078 ***
	(0.011)	(0.008)
White-collar experience squared	-0.075	-0.211 ***
	(0.162)	(0.086)

Dependent variable: log hourly salary

Type 1 and type 2 have the same occupation abilities. So do type 3 and type 4 as well as type 5 and type 6.

Type 1 and type 2 have different education psychic costs. So do type 3 and type 4 as well as type 5 and type 6.

Standard errors are in parenthesis.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 11: The Occupation-Specific Returns to A Bachelor's Degree

	Blue Collar	White Collar
Constants		
Type 1 & 2	6.914 ***	6.983 ***
	(0.024)	(0.035)
Type 3 & 4	6.465 ***	6.425 ***
	(0.023)	(0.042)
2-year college	0.257 ***	0.191 ***
	(0.025)	(0.036)
Bachelor's degree	0.257 ***	0.328 ***
	(0.035)	(0.033)
Blue-collar experience	0.084 ***	0.047 ***
	(0.009)	(0.01)
Blue-collar experience squared	-0.326 ***	-0.162 **
	(0.076)	(0.08)
White-collar experience	0.038 ***	0.083 ***
	(0.014)	(0.01)
White-collar experience squared	-0.049	-0.214 **
	(0.189)	(0.094)

Dependent variable: log hourly salary

Type 1 and type 2 have the same occupation abilities. So do type 3 and type 4.

Type 1 and type 2 have different education psychic costs. So do type 3 and type 4.

The four-year college dropouts are eliminated from the sample.

Standard errors are in parenthesis.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 12: Expected Returns to Education

	2-Year College		4-Year College	
	Type 1 & 2	Type 3 & 4	Type 1 & 2	Type 3 & 4
Panel A: Total				
1st year	0.238	0.225	0.337	0.251
5th year	0.225	0.212	0.349	0.255
10th year	0.239	0.22	0.397	0.287
Panel B: Occupation-Specific Skills Accumulation				
1st year	0.216	0.229	0.277	0.264
5th year	0.206	0.221	0.283	0.271
10th year	0.201	0.221	0.288	0.275
Panel C: Better Occupation Match				
1st year	0.022	-0.004	0.06	-0.013
5th year	0.019	-0.009	0.065	-0.016
10th year	0.038	-0.001	0.109	0.011

Calculation is based on the simulation of 10000 observations.

Panel A: total expected returns to attending a two-year college and a four-year college

Panel B: expected returns to education from enhancing the occupation-specific skills

Panel C: expected returns to education from increasing the probability of being employed in a white-collar occupation

Total expected returns to education (Panel A) is the sum of the expected returns from enhancing the occupation-specific skills (Panel B) and the expected returns from increasing the probability of being employed in a white-collar occupation (Panel C).

Type 1 and type 2 have the same occupation abilities. So do type 3 and type 4.

Type 1 and type 2 have different education psychic costs. So do type 3 and type 4.

Table 13: Posterior Probabilities of Types Conditional on Test Scores

(a) Type 1

	$\bar{Q}_{nc} \leq -0.5$	$-0.5 < \bar{Q}_{nc} \leq 0$	$0 < \bar{Q}_{nc} \leq 0.5$	$\bar{Q}_{nc} \geq 0.5$
$\bar{Q}_c \leq -0.5$	0.001	0	0.013	0.007
$-0.5 < \bar{Q}_c \leq 0$	0.065	0.127	0.150	0.191
$0 < \bar{Q}_c \leq 0.5$	0.328	0.459	0.565	0.680
$\bar{Q}_c \geq 0.5$	0.782	0.777	0.873	0.909

(b) Type 2

	$\bar{Q}_{nc} \leq -0.5$	$-0.5 < \bar{Q}_{nc} \leq 0$	$0 < \bar{Q}_{nc} \leq 0.5$	$\bar{Q}_{nc} \geq 0.5$
$\bar{Q}_c \leq -0.5$	0.615	0.54	0.445	0.357
$-0.5 < \bar{Q}_c \leq 0$	0.000	0.001	0.001	0.002
$0 < \bar{Q}_c \leq 0.5$	0.000	0.000	0.000	0.000
$\bar{Q}_c \geq 0.5$	0.000	0.000	0.000	0.000

(c) Type 3

	$\bar{Q}_{nc} \leq -0.5$	$-0.5 < \bar{Q}_{nc} \leq 0$	$0 < \bar{Q}_{nc} \leq 0.5$	$\bar{Q}_{nc} \geq 0.5$
$\bar{Q}_c \leq -0.5$	0.298	0.353	0.432	0.529
$-0.5 < \bar{Q}_c \leq 0$	0.313	0.343	0.325	0.327
$0 < \bar{Q}_c \leq 0.5$	0.056	0.043	0.033	0.031
$\bar{Q}_c \geq 0.5$	0.000	0.006	0.000	0.000

(d) Type 4

	$\bar{Q}_{nc} \leq -0.5$	$-0.5 < \bar{Q}_{nc} \leq 0$	$0 < \bar{Q}_{nc} \leq 0.5$	$\bar{Q}_{nc} \geq 0.5$
$\bar{Q}_c \leq -0.5$	0.086	0.106	0.111	0.107
$-0.5 < \bar{Q}_c \leq 0$	0.623	0.529	0.523	0.481
$0 < \bar{Q}_c \leq 0.5$	0.616	0.498	0.402	0.289
$\bar{Q}_c \geq 0.5$	0.218	0.217	0.127	0.091

Calculation is based on the simulation of 10000 high school graduates whose parents are high school graduates, who have three siblings, were raised in a two-parent family, lived in the northern urban area of U.S. at age 14, and took the six test scores at age 18.

$\bar{Q}_c$ is the average (standardized) math, verbal, coding speed, mechanical comprehension scores.

$\bar{Q}_n c$ is the average (standardized) Rotter Locus of Control and Rosenberg Self-Esteem Scale.

Each cell shows the probability of belonging to a specific type given that $\bar{Q}_c$ and $\bar{Q}_{nc}$ fall in a specific region.

Table 14: Expected Returns to Education, By Test Scores

	2-Year College			4-Year College		
	10th Percentile	50th Percentile	90th Percentile	10th Percentile	50th Percentile	90th Percentile
1st year	0.230	0.232	0.238	0.280	0.293	0.333
5th year	0.217	0.219	0.225	0.286	0.300	0.342
10th year	0.225	0.228	0.236	0.323	0.339	0.388

Calculation is based on the simulation of 10000 observations.

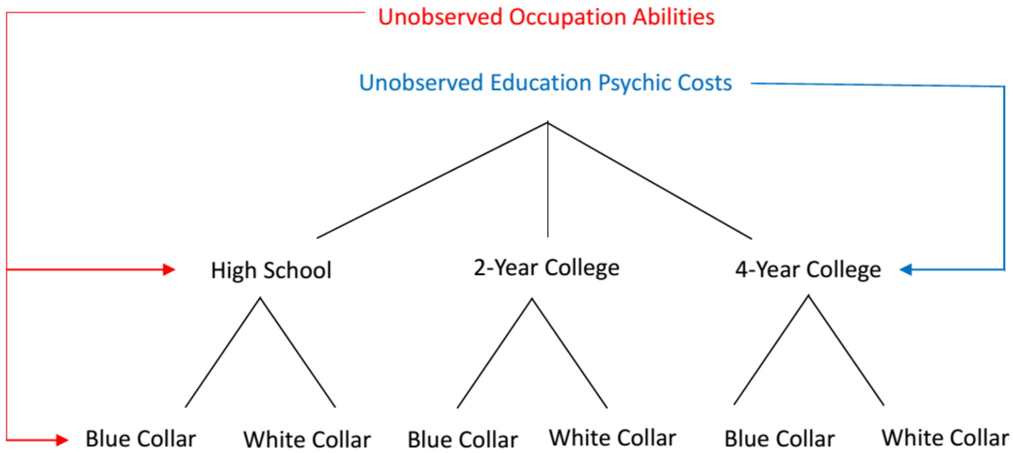

Figure 1: Sequential Education and Occupation Choices