EDUCATION CHOICES UNDER UNCERTAINTY

Thesis Submitted for the degree of
Doctor of Philosophy in Economics

Juan Yang
Supervisor: Prof. David J. Mayston

Department of Economics and Related Studies
University of York
January 2007

Abstract

Even though the return to education appears to be significantly positive, some talented young people do not attend post-compulsory education. Since education has been widely acknowledged as a key factor to production in today's information based economy, it is vital to find out the determinants that affect individuals' educational investment decisions. One of the possible affecting factors is uncertainty, which has been addressed three decades ago by some economists, but until now not much progress has been made on how to examine the factor of uncertainty by empirical observations. The thesis aims to design a testable optimal education choices model under uncertainty by using option pricing theory and examines it using British panel data. In addition, we are also interested in the impact of non-pecuniary utility on education choices and will examine this impact using a conditional logit model. In the second part of the thesis, we focus on the influence of a particular source of uncertainty, namely the supply and demand conditions in labour market (i.e. overeducation) to individuals' education choices. With the intensive expansion of higher education in China in recent years, the question of whether its graduates are being oversupplied to the labour market arises. The thesis employs a Chinese graduate survey to discover the role of overeducation on individuals' education choices and corresponding wages. Finally, I analyze the policy effect on individuals' education choices.

Acknowledgements

I would like to thank my supervisor Professor David J. Mayston for his support and advice throughout this PhD. Without his insightful guidance, it would not have been possible for me to complete this thesis.

I would also like to thank Dr. Jackson, Dr. Krause and Dr. Canepa for their constructive suggestions on the TAG meeting.

Thanks also go to my parents and sister for their love and support. Thanks to my friend, Zhuoshi Liu, for his love, understanding and assistance during the preparation of this thesis.

Declaration

Some of the work addressed in this thesis has previously been presented in the following conference:

Yue, Changjun & Juan Yang (2006), Overeducation or Undereducation-Some Evidence from Chinese Graduates, Paper presented to the International Conference on the Economics of Education, Dijon, France.

Mayston, David & Juan Yang (2006), Graduate Overeducation, International Student Demand and the Expansion of Higher Education, Presenting Paper in 25th Arne Ryde Symposium on the Economics of Higher Education and the Education of Economists, University of Lund, Sweden

All of the work contained within this thesis represents the original contribution of the author, except where otherwise indicated.

Table of contents

List of tables

List of figures

Chapter 1 Introduction

Educational choice is a very important task for both young people and the nation, since it concerns individuals' life-time career development as well as a nation's productivity and international competitiveness. For each individual, the level of education will determine his or her access to the labour market and their corresponding wages. The distribution of education and the stock of education accumulation are strongly related to a country's economic status and to economic indicators such as its unemployment rate, prosperity and equity. Education is thus of great concern to each nation's government, as well as to each individual.

Educational choice is also a very complex undertaking, since in reality it involves a number of possible uncertainties and risks. Winston (1999) states the difficulty as: 'People investing in human capital through a purchase of higher education don't know what they are buying and won't and can't know what they have bought until it is far too late to do anything about it.' The statement describes individuals often do not know the subjects they are really interested in, whether they have the capability to be successful in them or what their expected future return will be if they invest in studying them. In addition to the uncertainties and risks involved in the education investment, education is also a good that cannot be insured against or resold. The insurance market is reluctant to insure the return to education because of these uncertainties as well as problems of moral hazard and adverse selection. Belfield (2003) argues that "Adverse selection has students remaining in education for too long; moral hazard may mean students choose easy courses rather than those with the highest return". All these attributes contribute to the fact that educational choice is a very difficult decision.

Levhari & Weiss (1974) analyze the uncertain nature of the input (e.g. ability and quality of schooling) into educational investment and uncertain output (due to, for example uncertain future demand and supply conditions in the labour market) in a two-period model with exogenously determined labour supply. They estimate the optimal level of investment for schooling by maximizing current utility which composed the current and future consumption. They show that an increase in uncertainty about the return to human capital investment decreases the level of investment under plausible assumptions about risk preferences and the risk-return technology.

Williams (1979) estimates the uncertainty from four separate sources: uncertain productivity of education, stochastic depreciation of existing skills, unknown future wages for given skills, and risky returns on marketable assets in a continuous times model. He finds given some restrictions on personal preferences, the introduction of uncertainty produces more complete and more accurate predictions about observable variables. With risky returns to education, both time allocated to training and labour earnings are on average less peaked over the life cycle, which is similar to the results derived by Levhari & Weiss (1974). Williams also concludes with the increasing of human capital, dollars allocated to consumption and marketable assets are also growing accordingly. In other words, human capital accumulation can affect the investment of physical capital and consumption.

However, Kodde (1986) reports the empirical evidence contradicted their prediction and indicate that increased earnings uncertainty increases human capital investment. Snow & Warren (1990) provide an explanation for Kodde's empirical results by extending the Levhari-Weiss model and incorporating future labour supply as a choice variable. They argue that the marginal rate of return to human capital investment depends not only on the amount invested, as in the Levhari-Weiss model, but also on future labour supply. As a result, the return to human capital decreases with an increase in future income since the increased income reduces future labour supply.

Despite these contributions of the risky attribute on the return to education, not much progress has been made during the past three decades on how to examine the implications on education choices of such uncertainty. As recently as this year, Hartog & Vijverberg (2006) employ mean-variance theory to estimate individuals' optimal education strategy. Their model shows that individuals cannot decrease the risk of education choices by diversifying their skills. A shift of high school curriculum from being specific to more general increases variance for both male and female. If individuals cannot diversify their skills, the education choices model does not satisfy the basic assumption of the mean-variance theory and it is thus invalid to use this theory to analyse educational choices. However, using an alternative approach of financial theory to choices under uncertainty is a way to figure out a more attractive model for estimating educational choices under uncertainty.

Motivated by this idea, I will use option theory to evaluate the uncertainty inherent in human capital investment. The use of option theory to estimate education choices under uncertainty was first suggested by Hogan & Walker (2005), who constructed a continuous-time stochastic option model to estimate the optimal time at which to cease education. Although they derived a closed form from the real option model, the solution includes a confluent hypergeometric function, which cannot be examined by empirical evidence. In order to fill in some of the gaps in the empirical evidence on education choices under uncertainty, I use the simple and testable form of the Black-Scholes option model.

From the cell of table 1.1 we can see that not much literature discuss the uncertainty factor in the context of education choices and examined by the empirical observations. In this thesis, we will discuss drop-out risk and other risk including knowledge depreciation, relative position in the wage distribution etc. in the context of education choices in chapter 2 and examined by empirical evidence in chapter 3. Chapter 4 and chapter 5 will analyze the effect of uncertainted future supply and demand condition to individuals' education choices.

In brief, this thesis addresses the uncertainty problem in educational choices and seeks to derive an optimal education investment path for each youngster by using the option theory and treating education as a consumption good as well as an investment good. As part of utility, education can produce current and future consumption value, such as the enjoyment of reading a poem. Regarding the investment character, individuals are expected to acquire a higher return, which increases with the amount invested in education. The investment character of education is quite similar to buying a series of call options which provide individuals' the right to secure or give up a certain level of qualification based on their expected earnings. In order to receive the benefits they have to pay the costs of the option, which is equal to the tuition fees, living expenses and forgone earnings. Based on this idea, we estimated the value of investing in education by using an option model. The estimation results suggest both uncertainty and non-pecuniary utility can explain the behaviour of school attendance. In addition, uncertainty is more important to academic qualification attendees than technical qualification attendees.
In the second part of the thesis, we investigate the influence of a particular uncertainty namely the risk of overeducation, on each individual's expected wages and education choices by using the pecking order theory. With the world-wide expansion of higher education, more and more graduates are being oversupplied to the labour market and employed in non-graduate jobs, which is classified as overeducation. Even though a great amount of empirical evidence on overeducation has been provided for most countries, there is no complete theory to explain the reasons for overeducation. We apply the signal theory and assignment models into the pecking order theory to explain the possible reasons of overeducation and then provide the empirical estimates of key equations of the model by making use of an extensive survey of Chinese graduates. The regression results show unalterable personal characteristics (such as gender, registration belongings) and ability proxy, especially university rank affect

both getting a graduate level of job and the corresponding wages in the same direction. However, some variables (e.g. partymembership, family background, work sector) will play contradict role on getting a graduate job and high wages.

Since we consider the effect of job characteristics to individuals' education choices (i.e. assignment effect), the marginal return is no longer equal to the average return (i.e. the return to education is not constant, but decreasing with the time being). Therefore, the long trend of the supply of qualified labour to the market and education policy is quite critical to individuals' education choices. In the final part the thesis we research the social return to education in order to analyze the future education policy and the supply trend of qualified graduates to the market.

In Brief, five main contributions to the existing literature will be offered in the thesis. Firstly, I employ the Black-Scholes option model to estimate the optimal schooling under uncertainty and provide empirical estimation results. The previous literature on educational choices either examined the education choice model under certainty or discussed some limited aspects of the theoretical impact of uncertainty on educational choices. To my knowledge, this contribution is the first empirical evidence concerning British young people's education choices under uncertainty. Secondly, the thesis provides a comprehensive description of all the potential variables or conditions that may affect individuals' schooling choices, including uncertainty, non-pecuniary utility, information set, discount factor, etc. and provides empirical evidence on the impact of non-pecuniary utility and the discount factor by using the multinomial logit model. Through comparing the multinomial logit model with the Black-Scholes model, we can find out whether non-pecuniary utility and uncertainty play a role in individuals' education choices. Thirdly, the thesis develops the signal theory and assignment models to provide a comparatively complete reason on overeducation. Fourthly, the thesis provides the first empirical study on the important emerging problem of Chinese graduates overeducation. The empirical regression results expose some interesting findings on what may affect an individual's risk of overeducation, the impact of overeducation on wages and the relative attractiveness of investing in different levels of higher education qualifications. Fifthly, the thesis takes Lucas's (1988) endogenous economic growth model as a starting point and then develops its human capital increasing rate by considering the physical capital investment growth rate. This is accomplished by first investigating the socially optimal amount of human capital investment in China.

The thesis is organized as follows: Chapter 2 analyses educational choices under uncertainty by using the Black-Scholes options model and the multinomial logit model. Chapter 3 provides the empirical evidence on educational choices under uncertainty by using the NCDS data set. Chapter 4 and Chapter 5 discuss the effect of overeducation on individuals' education choices based on China's graduates. Chapter 6 estimates the social return to education and what the socially optimal amount of human capital investment is in China in order to provide some policy suggestions on individuals' education choices. Chapter 7 concludes the thesis.

Table 1.1 Risks and uncertainties discussed by current literature

	A	B	C	D
Drop-out risk				
Card (1999)	Y	–	–	–
Hartog et al (1989)	–	–	Y	Y
Altonji (1993)	–	–	Y	Y
Future supply and demand condition (overeducation)				
Levhari & Weiss (1974)	Y	–	–	–
Battu et al (1999)	–	Y	–	–
Chevalier (2003)	–	Y	–	–
Dolton & Vignoles (2000)	–	Y	–	–
Other risk including knowledge depreciation, relative position in the wage distribution etc.				
Levhari & Weiss (1974)	Y	–	–	–
Kodde(1986)	Y	Y	–	–
Williams (1979)	Y	–	–	–
Pereira & Martin (2001)	Y	–	–	–
Harmon et al (2001)	–	–	–	–
Snow & Warren (1990)	–	Y	–	–
Hartog & vijverberg (2006)	–	Y	Y	Y
Hogan & Walker (2005)	–	–	Y	–

Note: A represents the literature which discuss the uncertainty factor of the return to education from the theoretical model; B represents the literature which examine the return to education under uncertainty by empirical observations; C represents the literature which discuss the uncertainty factor in the context of education choices from theoretical model; D represents the literature which examine the uncertainty factor in the context of education choices by empirical observations.

Chapter 2 Education Choices under Uncertainty

2.1. Introduction

Education, in terms of human capital, can be viewed as an investment of current time and money in return for future expected higher payment (Freeman, 1986). Education can also bring the enjoyment of reading a poem, awareness of health problem and high standard social behaviour. However, Mayston (2002) points out that invest in human capital in the form of additional training or further education or other significant changes in career direction may involve a large element of sunk cost that cannot easily be recovered if wrong decision choices are made. Once a career choice has been made, the flexibility of the individual to costlessly reverse their initial choice is lost, suggesting every step move in education choices should be careful.

Due to the significant effects of education choices on individuals' career development, education or career choices received widespread attention in both theoretical and empirical literature. An early educational choice model focused on economic and policy aspects of individuals' preferences over schooling and work alternatives, exemplified by Kohn et al (1976) and Fuller et al (1982). In their papers, students' enrolment decisions regarding college are viewed as choices among a discrete set of schooling and nonschooling alternatives. A multinomial logit model expresses the probability that a given student will select a given alternatives. The effect of tuition fees, scholarship, living expenses, opportunity costs and academic performance on students' evaluation of alternative choices was calculated properly. Keane & Wolpin (1994, 1997) systematically analyze the individuals' selection process among five alternatives: working as a while collar, working as a blue collar, studying, staying at home and serving in the military. They employed a structure model described all the possible choices and their values in each period and assume individuals will choose the one which can maximize the wages utility function. Oosterbeek & Ophem (2000) derive a schooling choice equation from a Cobb-Douglas utility function that considers individuals' discount rate on schooling. They used a maximum likelihood function to estimate the factors that may affect individuals' education choices.

Though the educational choice model has experienced considerable progress, most of the literatures on college choices do not consider uncertainty: they view the individual as able to choose a future level of education with no uncertainty. However, when consider enrolling in any type of schooling, the individual faces considerable uncertainties. Firstly, the individual cannot assess whether (s)he can complete the education, in other words drop-out risks. Secondly, (s)he has imperfect knowledge and information of the value of his(her) abilities, the quality of the schooling and the trend for future demand and supply conditions, which will affect his(her) relative position in the post-education earnings distribution and the value of the education.

For a very long period time, there have been only a few innovative contributions to the literature that discuss the uncertain return to human capital investment. To my knowledge, Altonji (1993) is the one of the very few who models education choices under uncertainty and examines how variables, such as family background, tastes for school, ability, sex and high school curriculum influence individuals' choices on education and ex-ante rate of return, but he leave the task of estimating and test the structural model to later researchers. Hartog et al (1989) develop a dynamic model and treat schooling choices as a sequence of dichotomous decisions. The paper only tackles the reasons and probability for dropping out during the sequential study after correcting self-selection bias. However, none of them gives empirical evidence on education choices under uncertainty.

In the most recent years, scholars have sought suitable financial theory to examine the uncertainty from two directions. One is portfolio theory (e.g.

Hartog & Vijverberg, 2006). Hartog & Vijverberg employ mean-variance (M-V) theory to estimate individuals' optimal choices by linking the curriculum of a school program with a Dictionary of Occupational Titles and labour market wages. The empirical evidence in their model shows there is an upward sloping M-V frontier between variances and expected wages. Their model shows individuals cannot decrease the risk of education choices by diversifying their skills. A shift of the high school curriculum from a specific to more general one increases variance for both male and female. Their empirical results and the nature of human capital determine the restriction of their estimation methods, since it is almost impossible to diversify the human capital once they entered the labour market. The other direction is option theory (e.g. Hogan & Walker, 2005). A continuous-time stochastic model was used in their model to estimate the optimal time to cease education. Once individuals decide to exercise the option, they cannot return to school at a later date. For a number of reasons this approach seems to be less convincing. Firstly, they borrow the results of Dixit & Pindyk (1994), which include complicated forms and hypergeometric functions and it is hard to interpret the exact affecting power and relationships. Secondly, their model does not allow individuals to re-enter the school, which is too restrictive, especially for higher education. In our NCDS data set, more than 20 per cent of individuals receive further formal education after they entered the labour market.

In this chapter, I will also use real option model to evaluate individuals' education choices. In contrast to Hogan & Walker (2005), I treat individuals investing in education as buying a series of European call options, each of which will be exercised in the following working years. The value of the option is the expected excess return by investing more on education, which can also be extended as the excess utility from investing more on education. Individuals will invest in education if and only if the costs of education (option costs) are smaller than the value of the option. Individuals will exercise the option (i.e. secure a level of qualification) if the wage differential is large than the exercise price.

In brief, three major contributions to the existing education choices model will be advanced. Firstly, our evaluation is based on ex-ante wages, instead of ex-post wages (such as considered in Keane & Wolpin, 1997). In addition, our estimation is not based on the return to each year as in Altonji (1993), but qualification level and adjusted by the probability of dropping out. Following the Webbink & Hartog (2004) empirical work that graduates can predict their wages, we model the ex-ante return by all the potential affecting variables (e.g. individuals' secondary school curriculum, academic attitude, IQ, family background).

Secondly, this is the first empirical testable option model to analyze the education choices under uncertainty. Through researching the nature of education investment, I have designed the Black-Scholes (B-S) models to analyze the net payoff from investing more in education. The model not only overcomes the untestable characters of uncertainty brought forward by Levhari & Weiss (1974), but develops the estimation method of Hogan & Walker (2005), namely our model allows individuals' to make education choices at each separate stage.

Thirdly, this chapter brings a comprehensive description of all the potential variables or conditions that may affect individuals' schooling choices including uncertainty, discount factor, non-pecuniary utility, information set, etc. and compare the education choices under certainty and under uncertainty. Through comparison I can highlight the importance of uncertainty, discount factor or non-pecuniary to individuals' education choices. Within my understanding, it is the first attempt to contrast the types of education choices.

The rest of the chapter is organized as follows. Section 2 presents the basic education choices model and related factors on education decisions. In section 3, we analyze the uncertainty involved in educational choices and develop a B-S education choices model. Section 4 discusses the limitations and the connection to econometric models.

2.2. Basic Educational Choices Model

2.2.1 Four possible educational choice

We construct a basic educational choice model under certainty to describe individuals' possible decision choices at each stage in the context of four alternative choices. These choices are working, studying in academic college, studying in technical college and staying at home.

A basic structural framework can be used to analyse the present value of the rewards which individual i expects to receive from their life-time earnings as a result of their making the choice k, where these rewards are assessed in terms of their associated expected utility:

$$E[\sum_{l=t}^{T}(1+\theta_i)^{-l}(U(Y_{ikl}))\,|\,I(t)] \qquad (2.1)$$

where θ_i is the individual's discount factor. $U(Y_{ikl})$ means individual i's utility in period l depending on the available net income Y_{ikl} at time l as a result of making the education choice k. The four alternative choices will be assumed to be working currently in the labour market, studying in academic qualification, studying in technical qualification and staying at home, with the values of k = 1, 2, 3 and 4 respectively. I(t) is the information set available to the individual at time t, which consists of all factors, known to the individual, that affects current utility or the probability distribution of any of the future utility.

The individual's decision process can be described as follows: given their information set I(t) at time t, the individual considers each of the four alternative choices and uses them to calculate the projected future current rewards and thus the four alternative specific value functions, and chooses the alternative that yields the highest present value of their utility. In order to solve this question, we need to work backwards to evaluate the reward of each choice in the last period T, T-1,... and then select the best choice at present t.

choice 1: working (k=1)

If individual i chooses to work directly after compulsory education, at time t+1(the period after compulsory education) he will get:

$$\ln Y_{ik(t+1)} = \alpha_k + \beta_k X_i + \varepsilon_i^1 \tag{2.2}$$

where $\ln Y_{t+1}$ is the log of net wages for the choice of working. X_i is individuals' personal character including schooling, work experience, sex, race, marital status, union membership, region etc. ε_i^1 is the disturbance term for choice 1, which may also affect individuals' wages, but is assumed to be independent of X_i. In the later part of this chapter, we will mainly discuss the nature of ε and how to estimate it.

Choice 2 and 3: continue study (k=2, 3)

If the individual plans to continue study in the next period, he may get a higher return in the later working period after t_1 period's study, given by:

$$\ln Y_{ik(t+t_1)} = \alpha_k + \beta_k X_i + \varepsilon_i^{23} \tag{2.3}$$

The disturbance term ε_i^{23} in equation (2.3) will have a high variance than ε_i^1 in equation (2.1), which will be discussed in detail in the next section. However, before graduation, whilst studying they will bear the costs of tuition fees (W). If SC denotes the scholarship that talented students might obtained, the net income (Y) in the period of continue study can be expressed as:

$$\begin{aligned} &Y_{ik(t+1)} = (SC_i - W)\varepsilon_i^k \\ &(t \le T-1)(k = 2,3) \end{aligned} \tag{2.4}$$

k=2 means academic college was chosen in period t and k=3 means a technical college was selected. Scholarship, tuition fee and living expenses may be related with academic quality, individuals' ability and family background. Good college or university will charge more for tuition, but can be counteracted by excellent academic performance through scholarship. The same amount of tuition fee and living expenses to students coming from various family backgrounds may be significantly different.

choice 4: home production(k=4)

The per period reward function for staying at home is given by:

$$\ln Y_{ik(t+1)} = \xi_{t+1} + \varepsilon_i^k \tag{2.5}$$

where ξ is the income-equivalent value of staying at home. One can explain it as resulting from more leisure time at home and/or the household cost-saving in not having to hire a baby-sitter, cleaner, cook or gardener. Individuals will tend to select this option only if the opportunity cost of leisure is low, when women give birth to a child or when the economic value of staying at home production, such as from avoiding high child nursery fees, is almost equal to the wage from working outside. ξ_{t+1} was generally represented in our empirical studies by the basic wage of those who underwent only compulsory education.

2.2.2 Ex-ante return to education

A key factor in the analysis of educational choices under uncertainty is whether ex-post or ex-ante returns to education should be considered. Previous studies have mainly analysed the ex-post returns to education choices. Though heterogeneity within individuals and consequently within their returns has been emphasized in economic research for several decades now[1], not much knowledge, and especially empirical evidence, is known on ex-ante return to education. Knowledge of the ex-ante risk of educational investment requires a comparison of lifetime earnings under the alternative of not undertaking the educational investment with the expected lifetime earnings from all potential exits once the individual has made the educational investment. Since this information is almost

[1] one of the earliest contributions being Willis and Rosen(1979)

impossible to acquire objectively, individuals have to make the decision based on their subjective estimation of their own characteristics and current labour market information.

Nevertheless if we can estimate the distribution of the disturbance term in equation (2.3), and the associated coefficients on their personal characteristics, including their educational qualifications, it will provide guidance on how individuals might make optimal education decisions under uncertainty, albeit by making the best use one can of the information contained in the ex-post returns on possible education choices. We will consider here first in particular the uncertainty that is associated with the drop-out risk from the chosen educational programme, as dependent upon individual personal characteristics, institutional differences and other residual uncertainty.

Following the basic human capital principle, we revise the Mincerian log linear equation by controlling the education level instead of schooling, since the unequalled return to each year of schooling.[2] The ex-ante wages is given by equation (2.2) if individuals select choice 1 directly after compulsory education. However if individual i choose choice 2 or 3 at any time between t (the time just finishes secondary education) and T (end of lifetime), he has the risk of dropping-out during education. As showed in the previous footnote, the gap between graduate wages and the drop-out wages is not trivial, so that we need to refine $Y_{ik(t+1)}$ in equation (2.3) as $Y_{im(t+1)}$ in equation (2.6) if individuals choose choice 2 after compulsory education in order to consider the drop-out risk. The expected wages in period t+1 considering the drop-out risk can be expressed as:

$$E[Y_{im(t+1)}|I(t)] = p^{g}_{mt}[Y^{g}_{im(t+1)}|I(t)] + (1 - p^{g}_{m(t+1)})[Y^{u}_{im(t+1)}|I(t)] \qquad (2.6)$$

[2] Weisbrod(1962) pointed out the return to school is not linear, such as the reward to the first year of college is not the earnings differential between individuals with 12 and 13 years of schooling and the reward to graduate and dropout cannot be explained by one year of schooling differential as well.

where p_m^g is the probability to graduate from qualification m. $Y_{im(t-1)}$, $Y_{im(t+1)}^g$ and $Y_{im(t+1)}^u$ denote wages, graduate wages and drop-out wages at time t+1 for qualification m conditional available information at time t, respectively and assume $Y_{im(t+1)}^g > Y_{im(t+1)}^u$.

Altonji (1993) estimates the graduation probability through maximizing individuals' earnings, that is to say, agents will drop out from current education as long as further education cannot bring extra income. Hartog et al (1989) explain the two interactive forces behind the drop-out decisions: a push effect and a pull effect. The push effect emphasizes that the individuals cannot meet the requirement of certain qualification due to incompetent or lack of motivation. The pull effect explains the individuals are drawn our of school as they discover that the labour market offers them more favourable returns with a shorter education than initially expected. B-S model will take into account the pull effect and thus we only consider the push effect in our graduation probability model. Duncan et al (2003) also states that the push effect alone cannot explain the empirical observations, there are other reasons may prevent the explanative power of theoretic model, e.g., financial constraints, uncertainty, psychological reasons[3], which from another angel show the reasonability of our model.

The above analysis yields the graduation probability for qualification m[4] as:

$$p_{im}^g = \rho_{1m} A_i + \rho_{2m} P_i + \rho_{3m} D_i + K \tag{2.7}$$

Individuals' abilities A_i shows each student's current and potential skills and talents in a wide range of aspects, such as ability with computers, ability with words, and being athletic, artistic, musical, patient, responsible. P_i represents individual i's family background, which plays an

[3] pupils may have insufficient human capital and ability at the enrolment time to college or university. But in order to satisfy amour-propre or peer pressure, they may choose to attend college or university.
[4] Assume qualification m is higher than compulsory education

important role in student i's financial condition. D_i is self-effort, which expresses subjective intention or interests on study. The last term K is the constant term, which captures the unobservable factors. Johnes & McNabb (2004) estimate a similar equation as (2.7), but involve comprehensive variables, such as peer effect, institution effect, and subject effect. They confirm the importance of family background on individual's drop-out rate, but derive different effect of ability for equation (2.7). They found ability can play a positive as well as negative role on the probability to drop-out. No matter whether individuals' ability is above or below the average ability, it will affect the drop-out probability negatively. We will examine this equation in chapter 3.

Even if we estimate the graduation probability, there is another unpredicted risk involved in deriving the distribution of ex-ante returns, namely the economic risk between the period of decision making and the time to enter the labour market, which include the uncertain nature of demand and supply for a particular field of graduates, uncertain technological innovation, uncertain economic climate and uncertain education and economic policy. Uncertainties of this kind are to some extent also faced by stock or other financial assets, which inspires us to use financial models to estimate the ex-ante return under uncertainty.

2.2.3 Utility of Income and Consumption value (eqns (2.8)-(2.10) were developed by Prof. D.J. Mayston)

In order to pursue our analysis of individual education choice under uncertainty, we can incorporate a degree of risk aversion for individual i by making use of the utility of income function:

$$U(Y_{ik(t+1)}) = Y_{ik(t+1)}^{1-\rho} / (1-\rho) \tag{2.8}$$

where $\rho(<,>1)$ is the individual's coefficient of relative risk aversion. If $Y_{ik(t+1)}$ is lognormally distributed, we may show from Aitchison & Brown (1963), that

$$E(U(Y_{ik(t+1)})) = \exp((1-\rho)\psi_{ik(t+1)})\exp(0.5(1-\rho)^2\sigma^2_{ik(t+1)})/(1-\rho) \qquad (2.9)$$

$$\text{where } \psi_{ik(t+1)} \equiv E(\ln Y_{ik(t+1)}) \text{ and } \sigma^2_{ik(t+1)} \equiv \mathrm{var}(\ln Y_{ik(t+1)}) \qquad (2.10)$$

Equation (2.9) still, however, has income as the only determinant of the individual's utility, which may not provide a good fit to real life. To attempt to better describe the reality, we will extend the utility function by including non-pecuniary benefits and write the overall return in the next period as:

$$V_{ik(t+1)} = E(U(Y_{ik(t+1)})[f_1(m)]^{\chi_i} \qquad (2.11)$$

$f_1(m)$ is a non-pecuniary benefits function which increases with the qualification m and can also be interpreted as the consumption value of schooling and relates both to the present consumption (e.g. the pleasure of attending school) and to future consumption (for instance enjoying reading poems). Present consumption value includes the enjoyment of study as well as effort costs. The future consumption value includes improving the quality of life, personal reputation, job security and so on. Mayston (2002) lists 18 variables that may affect a person's quality of life or job satisfaction including the mental stress or pleasure experienced during the travel, interest involved in the work tasks, time spend with children, working conditions, etc., implying non-pecuniary benefits may affect individuals to make their decisions on education.

χ_i is each individual i's subjective weighting on the consumption value of education or their "tastes for schooling", which depends on information set available at time t. People from a good family background with high academic ability generally attach a high value to χ_i, increasing their probability to attend college and university. Individuals with low academic ability or from a poor family background can experience a very high effort costs and are reluctant to attend extra schooling. Assume school preference χ_i is a linear relationship of self interest on study, family background and ability. It yields

$$\chi_i = \mu_i(E_i + P_i + A_i) = \mu_1 E_i + \mu_2 P_i + \mu_3 A_i \qquad (2.12)$$

There is a dichotomous relationship between $f_1(m)$ and U(Y). On the one hand, with increasing wages, people can spend more money on health care or mental and physical relaxation, thereby increasing in a multiplicative way the utility $f_1(m)$ that they obtain in (2.11) from the education they obtain from higher qualifications. However, this tendency may be mitigated to some extent by the greater pressure, energy and time consuming nature of the work that may be associated with earning the higher wages.

After considering the non-pecuniary utility function and four possible alternative choices, the present life-time utility function can be revised as:

$$V_{ik} = \sum_{l=t}^{T} (1+\theta_i)^{-l} E(U(Y_{ik(t+1)})[f_1(m)]^{\chi} \qquad (2.13)$$
$$(k = 1,2,3,4)$$

where their expected utility of income depends on the choices k in each stage. Individuals are expected to choose the optimal k at each period to maximize V in equation (2.13). In the next section, we will analyze how to solve this problem.

2.3 Deriving the Optimal Education Choices

Under the basic model, we assume the labour market is an efficient market. As defined by Fama in 1965 for financial markets, this involves large numbers of rational, utility-maximizers actively competing, with each trying to predict future market values of individual securities, and where important current information is almost freely available to all participants. In an efficient market, competition among the many intelligent participants leads to a situation where, at any point in time, actual prices of individual securities already reflect the effects of information based both on events that have already occurred and on events which, as of now, the market expects to take place in the future.

Now we apply this concept to the labour market. Each individual in the labour market is a rational expected utility-maximizer that they consider their current consumption as well as their life-long benefits conditional on their information set. Market wages reflect all the available information, such as labour supply and demand for this job, firms' financial situation, required skills and so on. Complete information set (e.g. available jobs, wages) is access to everyone and there is no ethnic and sex discrimination. Individuals with the same qualifications and skills will get the same rewards.

In order to simplify the complicated alternative choices, I draw an easy-to-understand visual graph (see Figure 2.1) to show all the choices, probabilities and consequences in an education choices decision tree. The decision tree provides a highly effective structure within which you can lay out options and investigate the possible outcomes of choosing those options. They also help you to form a balanced picture of the risks and rewards associated with each possible course of action.

From Figure 2.1, we can work out the decision process clearly. Individuals normally finish compulsory education at age 16[5] (time t) and confront four potential choices 1, 2, 3, and 4. To make an optimal choice in period t+1, a rational investor should estimate the rewards in period T conditional on the information set I(t) and the choices in each previous period for each possible education choice set and select the one that can maximize individuals' benefits. The decision choice for period t+2 is similar to period t+1, which require agents to calculate the expected return in time T conditional on the information set I(t+1), k and select the best one.

2.3.1 Factors affecting educational choices

Within the decision tree in Figure 2.1, there are two factors which we will examine first which play a role in educational decisions. These are the

[5] Assuming that the minimum school leaving age is 16 and all the individuals finish the compulsory education

information set, I(t), that is available to the individual decision maker at time t and their discount factor on future levels of expected income. I will examine these two factors here in a more detail.

2.3.1.1 The information set

In this section, I will show how the limited information set, I(t), that is available to the individual decision-maker at a given time t affects their educational decisions. We will assume in particular that the information set will influence the individual's subjective estimate, $g_{kmI(t)}$, of their probability of graduation with qualification m if they opt to study for this qualification in a college of type k =2,3. The subjective probabilities of the two states of the world, of graduating with this qualification or failing to graduate after studying for it, can then be represented by [$g_{ikmI(t)}$, (1-$g_{ikmI(t)}$)]. If we assume the financial returns to individual i at time l from these two states of the world are Y_{iklm} and Y_{iklm0} respectively if they study for qualification m in a college of type k = 2, 3.and the non-pecuniary benefits are $f_1(m)$ and $f_{10}(m)$ respectively, the subjective expected utility that the individual expects to receive from opting to study for qualification m is given by:

$$V_{it}(m,k,I(t)) =$$
$$\sum_{l=t}^{T}(1+\theta_i)^{-l}\{g_{ikmI(t)}EU(Y_{iklm})[f_1(m)]^{\chi_i} + (1-g_{ikmI(t)})EU(Y_{iklm0})[f_{10}(m)]^{\chi_i}\} \tag{2.14}$$

Given the information set, I(t), the individual will then choose the qualification m they want to study, and the college type k=2,3 at which they wish to study, in order to achieve:

$$V_{it}^{*}(I(t)) = \max_{m,k} V_{it}(m,k,I(t)) \tag{2.15}$$

with associated educational decisions by individual i of $m_i(I(t))$ and $k_i(I(t))$ that are clearly contingent upon the information set I(t) that is available to the individual at time t.

If the information set I(t) which they have available to them at time t is imperfect, their subjective probabilities $g_{ikmI(t)}$ of graduation may differ from their true underlying probabilities g_{ikm}. Their expected utility, using these true underlying probabilities, that is expected to result from the educational choices they have made on the basis of the imperfect information set I(t), may then fall short of (2.15), such as through their overestimating their true probability of graduating. We may then have:

$$V_{it}(m,k) = \sum_{l=t}^{T} \phi_i^{-l} \{g_{ikm} EU(Y_{iklm}) f_1(m)^{\chi_i} + (1-g_{ikm}) EU(Y_{iklm0})[f_{10}(m)]^{\chi_i}\}$$
$$<$$
$$V_{it}^{*}(I(t)) = \sum_{l=t}^{T} \phi_i^{-l} \{g_{ikmI(t)} EU(Y_{iklm}) f_1(m)^{\chi_i} + (1-g_{ikmI(t)}) EU(Y_{iklm0})[f_{10}(m)]^{\chi_i} \qquad (2.16)$$

for m = m_i(I(t)) and k= k_i(I(t)), and where $\phi_i \equiv 1+\theta_i$. The individual may then regret making these educational choices and would may have made different educational choices, $m_i(\Omega)$ and $k_i(\Omega)$, if they had had this more perfect information set Ω, namely those which are the solutions to the optimisation:

$$V_{it}^{*} = \max_{m,k} V_{it}(m,k) \qquad (2.17)$$

Improving the information set, I(t), that is available to the individual in making their educational choices will then have a positive benefit of the increase it achieves in $V_{it}(m,k)$ for m = m_i(I(t)) and k= k_i(I(t)), up until the point where this information set is perfect and $V_{it}(m,k) = V_{it}^{*}$. However, if the improvements in the information set are costly to a secure, these positive benefits must be weighed against the costs of improving the information set, with an optimum occuring where the marginal benefits are equal to the marginal costs (see Mayston, 2002).

2.3.1.2 Discount factor

We have already incorporated risk aversion into our analysis of education choice under uncertainty through the coefficient of relative risk aversion

ρ in equations (2.8) and (2.9). Constancy of the coefficient of relative risk aversion across individuals itself implies that absolute risk aversion declines with income (see Pratt, 1964), as one might expect intuitively. The discount factor θ_i in (2.1) reflects the individual's subjective rate of time preference, that also affects their educational choices. Thaler (1999) suggests that individuals from higher-income families will have a higher propensity to save because their discount rate will be smaller than the people from poor family background. In addition, more affluent families may have more ability to borrow money at a lower rate of interest, as well as more accumulated savings of their own, than poorer families. In this thesis, I follow Willis and Rosen (1979) and Oosterbeek and Ophem (2000) by assuming that the discount rate θ_i associated with the subjective rate of time preference depends upon their social background characteristics ϖ_i and a parameter vector b_1; hence

$$\theta_i = -b_1 \varpi_i \tag{2.18}$$

The coefficients b_1 and the taste for schooling parameter χ_i can be estimated from a multinomial logit model of how individuals select a particular regime, as discussed below.

2.3.2 Multinomial logit model (eqns (2.20)-(2.24) were developed by Prof. D.J. Mayston)

The solution of the optimization problem (2.15) can involve considerable complexity of computation. Many scholars discussed various ways to solve this type of problem[6]. Here, for the sake of simplicity, I assume the unobservables ε_i^k are conditionally serially independent. If choice m is chosen in time t, the present value should follows,

[6] See Miller (1984), Pakes (1986), Keane & Wolpin (1994)

$$\begin{aligned} &\Pr[V_m(t)|I(t)] = \Pr[Y_{mt} + \theta E \max(V(t+1)|I(t))] \\ &\geq \Pr[Y_{kt} + \theta E \max(V(t+1)|I(t))] = \Pr[V_k(t)|I(t)] \\ &m \neq k, k = 1,2,3,4 \end{aligned} \tag{2.19}$$

Equation (2.19) is derived from Figure 2.1 and also explains the selection process. Choice m is chosen in period t if and only if its reward in period t plus the highest utility in period t+1 is the highest among the four choices. The selection is an iterative process and we employ the same rule to select j in period t+1, t+2, ..., T. The solution of equation (2.19) involves massive mathematics calculation. In this thesis, I only consider four choices (academic college, academic degree, technical college and technical degree), which changes the dynamic model of Keane and Wolpine into a comparative static model[7] so that individuals only need to consider a limited steps. In the empirical estimation, I then simplify the comparative selection process of Figure 2.3 into a static model of individuals' optimal highest qualification, that is to say individuals should stop at which level, academic college, technical college, academic university or technical university? Please refer to Figure 2.4 for the logic of the estimating process.

This problem can be solved more directly by a conditional logit model. McFadden (1974) estimates the probability that a specific choice m is selected at time t, if and only if its return is conditional on the information set at t is higher than any other possible choices. We can also derive the conditional logit model more directly from our above analysis of education choices under uncertainty by considering the continuous time version of equations (2.1) - (2.14), where we replace the discounting term $(1+\theta_i)^{-l}$ in (2.1) – (2.14) by its continuous analogue $e^{-\theta_i t}$, together with the simplifying case where t = 0 and T goes to infinity, and assume a constant rate of growth φ_i is individual i's expected earnings in each time period, such that:

$$\psi_{ikl} = \psi_{ik0} + \varphi_i l \text{ and } \sigma^2_{ikl} = \sigma^2_{ik0} \tag{2.20}$$

[7] See Figure 2.3 for detail.

Equation (2.13) using (2.9) and (2.10) then becomes:

$$V_{ik} = \int_{t=0}^{\infty} E(U(Y_{ikt}))f_1(k)^{\chi_i} e^{-\theta_i t} dt = \int_{l=0}^{\infty} E(U(Y_{ik0}))f_1(k)^{\chi_i} e^{((1-\rho)\varphi_i - \theta_i)l} dl$$

$$= E(U(Y_{ik0}))f_1(k)^{\chi_i} / (\theta_i - (1-\rho)\varphi_i) \qquad (2.21)$$

The multinomial logit model may be derived from our above analysis if we assume that across different individuals i (2.21) is subject to a proportionate stochastic term ς_{ik} in addition to the systematic influences on V_{ik}, such that they determine their education choices by seeking to maximise:

$$W_{ik} = \varsigma_{ik} V_{ik} \qquad (2.23)$$

If each $\ln \varsigma_{ik}$ is independently and identically distributed according to a type 1 extreme value distribution in standard form, it follows from J.S. Cramer, An Introduction to the Logit Model for Economists, Edward Arnold, 1991, p.51, that the probability of individual i selecting choice k is given by:

$$\Pr[k_i = m | I(t)] = \frac{\exp(\ln V_{im})}{\sum_{k=1}^{4} \exp(\ln V_{ik})} \qquad (2.24)$$

2.4. Real option model

In order to create and sustain a superior economic performance for individuals in a competitive society, attending postsecondary education may well be one of the best choices for agents. Palacios-Huerta (2003) shows that university education has the highest return, modest variance and superior liquid investment. According to British government's White Paper (2003), those who have been through higher education in the UK earn on average 50% more than those who have not. A recent report (2002) addresses that people with a higher education qualification are less likely to be unemployed. In addition, as analyzed in previous section, schooling can increase individuals' future consumption value and quality of life. Nevertheless, a number of academically talented young people do not attend a post-secondary institution.[8]

Some economists analyze that financial constraints or parental education may prevent students attending university. Kane (1994) and Ellwood & Kane (2000) report that the reluctance to attend college mainly results from borrowing constraints. Acemoglu & Pischke (2001) suggest that family income has a significant effect on college enrolments. Others[9] argue that parental education has an important impact on schooling choice.

As we have argued above, investment in human capital may be affected by differences in time preference rates due to differences in parental income, particularly if capital markets are imperfect and individual families face borrowing constraints which prevent them from equating their marginal time preference rates to a common market rate of interest. However, another key factor which may deter individuals from investing in human capital is the existence of uncertainty in the income they may receive from different educational qualifications, with risk aversion to such income variations varying with parental income in the way we have discussed above. As Williams (1979) notes, differences in educational attainment between individuals due to differences in financial wealth are commonly attributed to the uncertainty in the likely returns on educational investment.

Chen (2003) argues that the main reason discouraging high school graduate from investing in higher education is that wage volatility significantly increases by pursuing a four-year university degree compare to a two-year college qualification. Altonji (1993) finds that the fear of not finishing the degree requirements is discouraging students from attending college. These theoretical and empirical articles demonstrate that

[8] Chen(2003) provides statistic data from NLSY that high school graduates between the ages of 32 and 40 in 1997 with a scholastic ability test score in the top quartile, around 16 percent did not attend college.

[9] See Cameron and Heckman(2001), Cameron and Taber(2000), Keane and Wolpin(2001)

uncertainty can explain a significant part of the reasons for education choices.

Primary and secondary education is compulsory in all the developed countries and most of the developing countries. During this period, people will learn the basic common knowledge and skills, which are called generalized education and all these knowledge and skills are transferable and can almost be used in every field. However, people with only primary and secondary education are relatively disadvantaged in processing new information and adjusting to technological change that demands higher skills, so that they may confront higher unemployment risks. However, the uncertainty facing more highly educated people is dichotomous. On the one hand, people will learn more non-transferable skills and spend more time on it that, as Mayston (2002) has argued, becomes a sunk cost, exposing them to larger risks of technological shock and the obsolescence of these more specific skills; on the other hand, educated individuals have a comparatively high analytical and solving problem ability leading to a low unemployment probability. In addition, more educated workers are able to perform a wider range of tasks and are generally easier to be trained in new skills than less educated workers.

As the decision tree in Figure 2.2 emphasises, educational investment is a sequential choice process and the extent of the above two dichotomous aspects is unequal at different levels of education. In this choice process, we may identify three main sources of risk that may affect their academic choices.

Firstly, individuals may have imperfect information about their own abilities to absorb, understand and apply the education they receive that affect their drop-out risk. Due to the existence of the "sheep effect" under which wages are linked to qualifications obtained rather than education received, individuals may not get a corresponding return of the investment they have made before they drop out, but still have to bear the costs of tuition fees and forgone earnings.

Secondly, individuals may not have enough information about their abilities relative to those of their peers' abilities. and hence about their relative and absolute position in the distribution of post-education earnings. Since the range of ability within the same education level increases after higher education expansion10, the distribution of wages for the same qualification becomes wider and wider. Candidates are more concerned about the personalized expected return before making education decisions, suggesting that we should focus on the ex-ante distribution of future wages after graduation rather than the ex-post return as the key factor to educational choices. Despite the importance of ex-ante returns to educational choices, its nature and the empirical estimation are much less developed than that for ex-post returns.

The third source of risk individuals may confront comes from market risks, technological innovations and the risks of knowledge depreciation that may affect their future earnings from any given educational qualification. Due to the high speed of technology innovation, what students learned in school may not be used in the future careers resulting structure unemployment. A good example of this uncertainty is semiconductors. With the invention of computer and other communication technology, semiconductor can be used in a more and more broad way. However, many students in the semiconductor specialty may still learn the outdated knowledge of applying it to a radio resulting them hard to find satisfied jobs.

2.4.1 Real option model

Our multinomial logit model on optimal educational choice that is outlined in Figure 2.4 considers drop-out risk and risk aversion that depend upon parental income, but not market risk. However, as we noted above, market risk, and especially wages volatility, may be critical to individuals' educational choice. We therefore utilise another educational choice model to consider such kinds of risks. Investing in education is

[10] Higher education was expanded in the world level. Higher education attendees are significantly increased no matter in the developing country or developed country.

similar to buy a call option, which gives them the right to make use of the qualification later in the labour market if it is beneficial for them to do so. Rational individuals will invest in education iff the net payoff from investing in a certain level of education is larger than not investing in this level of education. The optimal choice can be expressed as:

$$\Omega = \max[EY^{e}_{i(t+1)} - C, Y^{0}_{i(t+1)}] \tag{2.25}$$

where C represents the option costs including forgone earnings, tuition fees and living expenses. $EY^{e}_{i(t+1)}$ and $Y^{0}_{i(t+1)}$ are the expected earnings of investing and not investing in a certain level of qualification for individual i in period t+1, respectively. Since expected wages Y follows a lognormal distribution, which has been demonstrated by Mincer (1974), allows us to assume the increment of wages follows a stochastic process[11]:

$$dY^{e}_{im} = \gamma_i Y^{e}_{im} dt + \sigma Y^{e}_{im} dz \tag{2.26}$$

Equation (2.26) is also called Geometric Brownian Motion process with drift. γ_i in equation (2.26) is the expected growth rate of Y_i for individual i, which determined by economy-wide factor. σ is the disturbance or uncertainty variable affected by macroeconomic variables. dz is a Wiener process.

Let $R(Y^{e}_{im}) = EY^{e}_{im} - C$. According to equation (2.26) and the Bellman equation, maximizing Ω in equation (2.25) is equivalent to

$$\theta R(Y^{e}_{im})dt = E(dR_i) \tag{2.27}$$

[11] Someone may argue that the growth rate of logarithm wages do not follow a stochastic process, which depends on ability and other personal characters. Yes, this assumption of stochastic process is very strict for wages equation. However, in reality there are very few variables follow the GMB process, which requires Markov property, independent increments and changes that are normally distributed. Even the most liquidity product stock cannot satisfy the last condition (see Dixit and Pindyck, 1993). Evaluate an education investment opportunity is quite similar to assess an investment opportunity for a real business, that both of them are lack of liquidity and cannot be easily sold. When an investment opportunity for a company was estimated, the value of the company was always assumed to follow a GMB process. Whereas, the value of the company may also be affected by CEO performance, the product of this company, the prospect of this industry etc, not a purely stochastic process. This simple stochastic assumption allows us to put more effort on the nature of the uncertainty and avoid the difficult mathematic problems.

Equation (2.27) says the value of option's change rate equals to the expected return from investing a very short period of schooling. Expand dR_i using Ito's Lemma, and we get

$$dR_i = R'(Y_i)dY_i + \frac{1}{2}R''(Y_i)(dY_i)^2 \tag{2.28}$$

Substituting equation (2.26) for dY_i into this expression yields (note E(dz)=0),

$$E(dR_i) = \gamma Y_i R'(Y_i)dY_i + \frac{1}{2}\sigma^2 (Y_i)^2 R''(Y_i)dY_i \tag{2.29}$$

Equation (2.29) now becomes[12]

$$\gamma Y_i R'(Y_i) + \frac{1}{2}\sigma^2 (Y_i)^2 R''(Y_i) = \theta R_i \tag{2.30}$$

In order to solve equation (2.30) or distinguish the wages equation from other security, we need to specify two boundary conditions and an initial condition.

The boundary condition can be obtained by the nature of human capital investment. If no investment happened, the net payoff will be zero, which is denoted by

$$R(Y_i^0) = 0 \tag{2.31}$$

If one's ability is beyond average, the optimal value of education investment equals

$$R(Y_i^*) = \int^{T} e^{-(l-t)}(Y^*_{i(l+1)} - Y^0_{i(l+1)})dl - C \tag{2.32}$$

and $R(Y_i)$ in equation (2.32) must be larger than zero. Finally the smooth pasting condition[13] to ensure $R(Y_i)$ were continuous and smooth at critical value Y, which follows,

[12] Equation (3.27) can also be derived by contingent claim analysis, see Merton (1973) for detail.

$$R'(Y_i^*) = e^{-(l-t)}(Y_{i(t+1)}^* - Y_{i(t+1)}^0) \tag{2.33}$$

One could solve equation (2.33) under three boundary conditions by the standard methods of Fourier transforms or separation of variables, but this would involve massive mathematics calculation. I plan to use the solution derived by Black-Scholes (1973) and Merton (1973) to evaluate the options. The B-S model is expressed in a simple testable way, but imposing strong assumptions. One of the most important assumptions is that the option must be exercised in a fixed date. But individuals' life-time earnings will benefit from educational investment, not on a certain exercise date as assumed by B-S model. Considering these facts, I set up a series of European call option model. Presume investing on education as buying a series of call options, which can be exercised in a sequential number of years to secure agents' earnings are larger than a certain level ($Y_{i(m-1)}^e$) on each year. For example, agents investing on bachelor education were treated as buying 44 shares of European call options and each of which will be exercised in the following each year after graduation. The numbers of options assumed to be bought were calculated by the age of retirement (65 in UK) minus the age of graduation from higher education (average 21 in UK).

Another assumption in B-S model is the fixed exercise price. Since wages were assumed to follow a GMB stochastic process, we have to adjust the wages differential instead of wages as the underlying assets of B-S option. To the same individual i, his wages growth rate for qualification m and qualification m-1 will be very similar, since the ability and economic environment are the same. Therefore the wage differential method is a good way to solve the fixed exercise price. The exercise price in the B-S equation cannot be zero and we will use a very low price L (e.g. 1 pound) to proxy exercise price. Based on the analysis above, the option price

[13] Smooth pasting condition is a very strong assumption for education investment. In order to use standard Black- Scholes model I have to assume the return to schooling is the same across years and do not consider the sheepskin effect.

$R(Y_{imt})$ in B-S equation can be written as an expression of education choice as,

$$R(Y_{im}) = \sum_{l=0}^{N} (Y^{e}_{im(l+1)} - Y^{e}_{i(m-1)(l+1)})(1+\lambda_m - r)^l N(x_1) - LN(x_2) \qquad (2.34)$$

where $N(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp(-\frac{1}{2}x^2)dx$ and

$$x_1 = \frac{\ln[(Y^{e}_{im(t+1)} - Y_{i(m-1)(t+1)})(1+\lambda - r)^l / L] + \sigma^2\tau/2}{\sigma\sqrt{\tau}}$$ and

$$x_2 = \frac{\ln[(Y^{e}_{im(t+1)} - Y_{i(m-1)(t+1)})(1+\lambda - r)^l / L] - \sigma^2\tau/2}{\sigma\sqrt{\tau}} = x_1 - \sigma\sqrt{\tau}$$

λ_m is the average wage growth rate and r is the interest rate. Thus $1+\lambda_m - r$ represents the discounted wage growth rate. L denotes the very low exercise price and equals 1 in our empirical studies. σ is the standard deviation of uncaptured factors that may affect wages, τ is the duration period of this contract, which is equal to 1 year in our case. l corresponds to the year of the contract, which takes the value of 1, 2, … , N. N equals to 44 on the choice of whether to attend university.

Equation (2.34) explains the option in the context of education choices. Namely the investment in an additional qualification to what the individual already has involves a bundle of N call options, which give individuals the right to work at the wage corresponding to that qualification if exercised at the future date. The call option will be exercised at the future exercise date iff the wage differential exceeds the exercise price. Individuals should compare the benefits of investing in qualification m, $T(Y_{imt})$, with the education costs (C) before making any education decisions. It should be noted that the cost of education should be separated into N shares, when one assesses the value of $T(Y_{imt})$.

This estimating method overcomes the fixed exercise date limitation of a single European option model and provides much more freedom of education choices. Comparing to Hogan & Walker (2005), this model allows individuals make education choices at each separate stage based on the available information at that time. In addition, individuals could break several years between studies and choose all types of education routes by combining both academic and technical. The logic of the option model on optimal educational choice was explained in Figure 2.5.

2.5 Limitations and Connections with Empirical Work

Although this model can satisfy the multi-purpose estimation, it involves several very strong and unrealistic assumptions. Firstly, the model requires that the distribution of wages growth rate satisfies a drifted geometric Brownian motion, while the wages may not follow this process. Secondly, B-S model assumes the financial assets can be continuous hedged, but it is impossible to be realized for human capital. We have to assume individuals' are risk neutral, which is a pretty strong assumption for education choices.

In next chapter, I will use empirical observations to examine the education choices model by considering the risks and not considering the risks in order to find out which provides a better fit to the empirical data. The results will imply whether uncertainty plays an important role in education choices.

To estimate the optimal investment from the real option model, we only need to work out $Y^e_{im(t+1)}$, $Y_{i(m-1)(t+1)}$ and σ. The beginning wages $Y_{i(m-1)(t+1)}$ without receiving a further education and expected wages after receiving a further education $Y^e_{im(t+1)}$ will be evaluated by equation (2.2) and (2.6) separately. Volatility σ can be collected from historic data.

The evaluation of the multinomial logit model is mainly based on individuals' utility function, which is composed by ex-ante wages as well as non-pecuniary return. The later will be represented by family background and individuals' interests or willingness on study.

Figure 2.1 Education choices decision tree

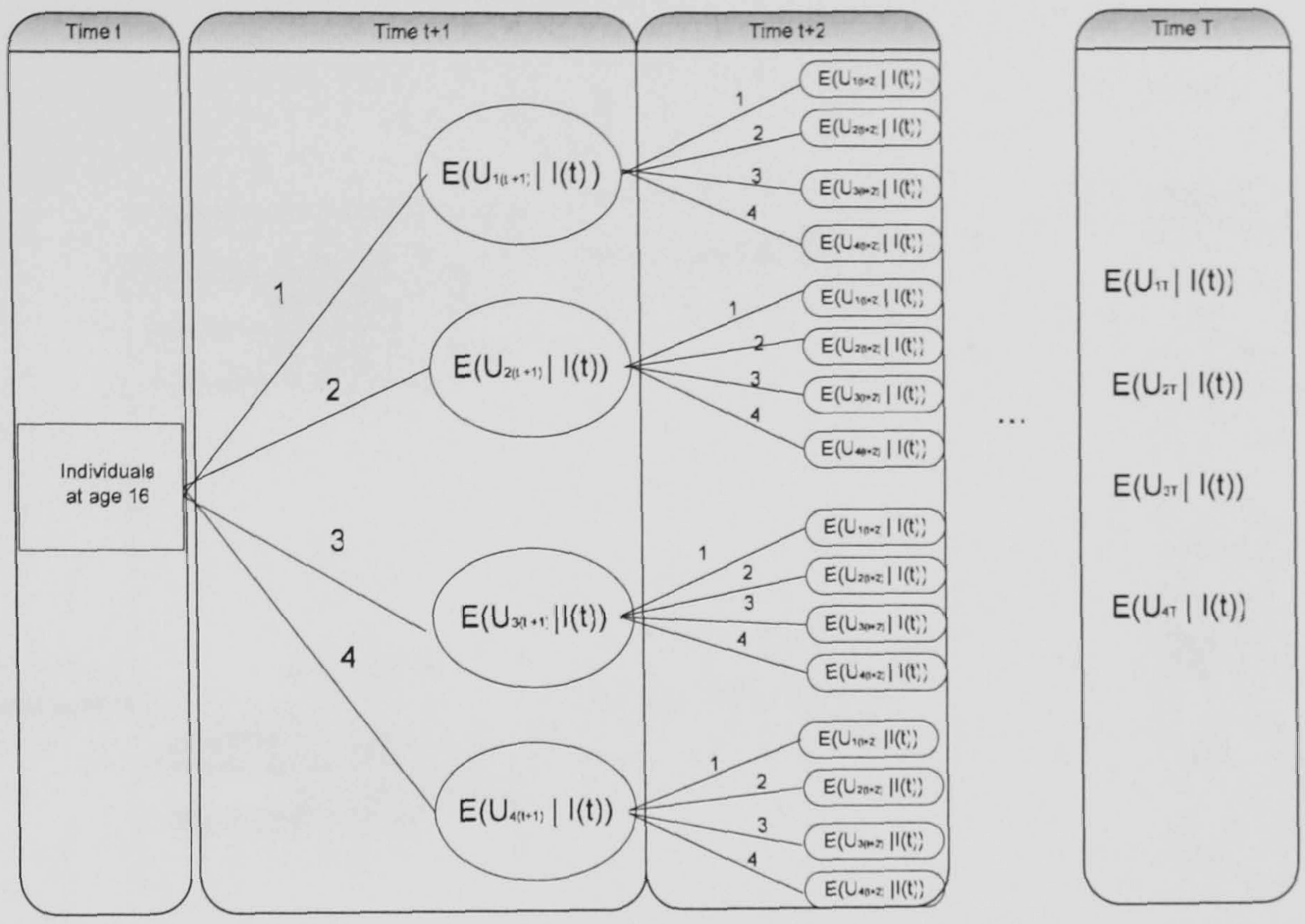


Figure 2.2 Decision tree for education investment

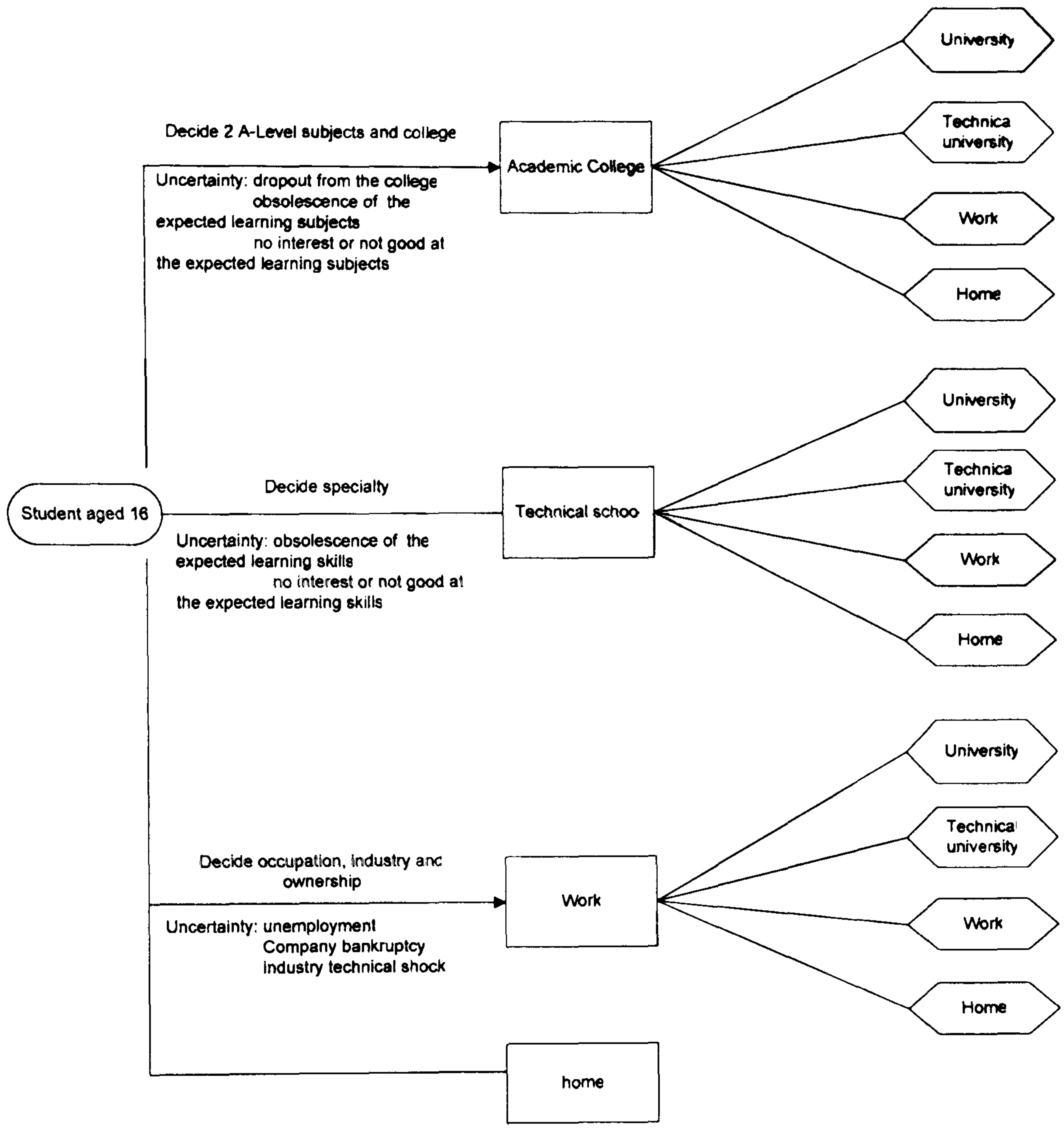
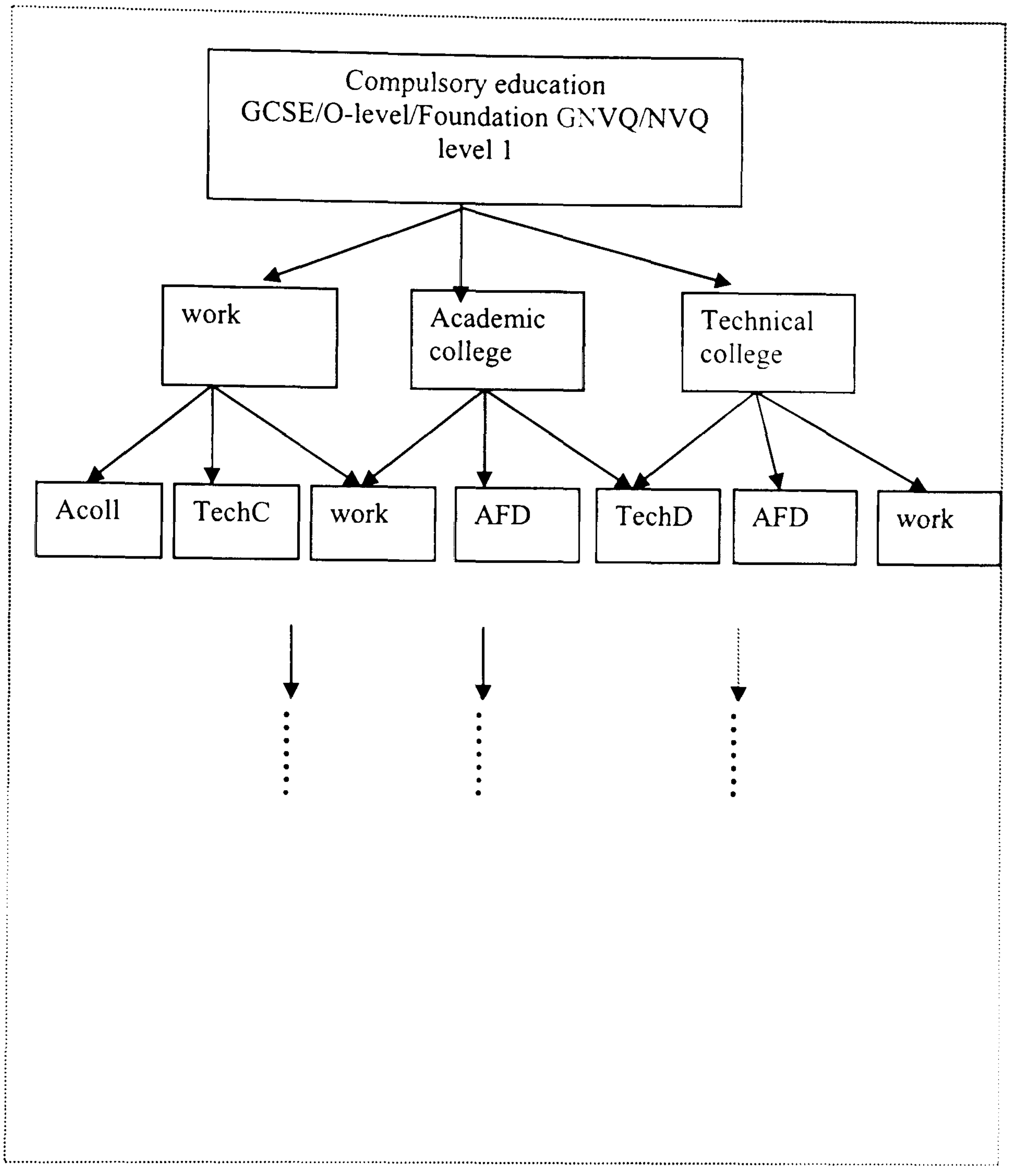


Figure 2.3 Possible selection process after compulsory education



AColl	GCE 'A' level,
TechC	TEC / BEC Certificate or Diploma, ONC/ OND (or SNC/SND), Advanced/ Final /Part II or III, Polytechnic (or Central Institute) Diploma or Certificate, RSA – Stage 3
AFD	University or CNAA First Degree
TechD	HNC/ HND (or SHNC/SHND), TEC / BEC Higher or Higher National Certificate or Diploma, Professional qualifications, Nursing qualifications

Figure 2.4 Modelling optimal educational choices under uncertainty using a multinomial logit model

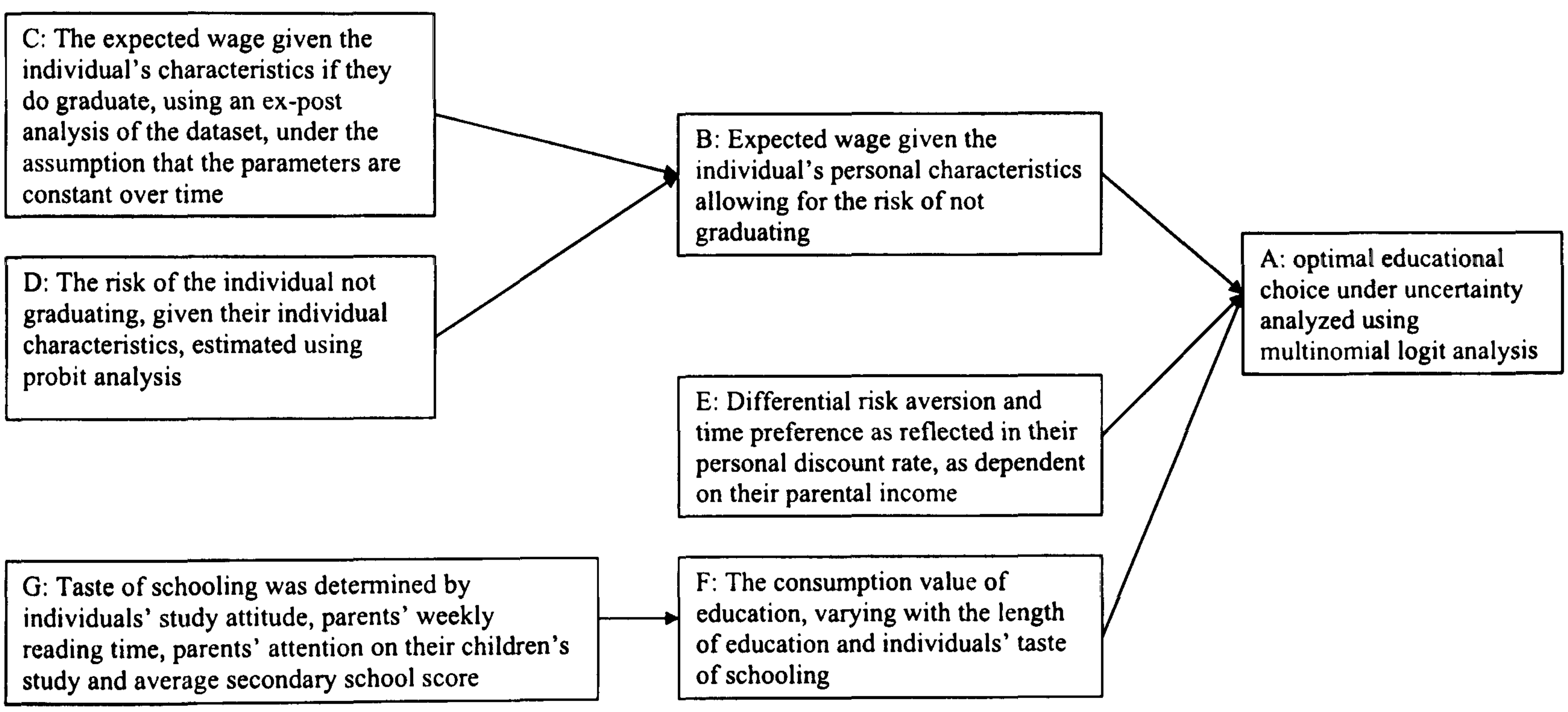
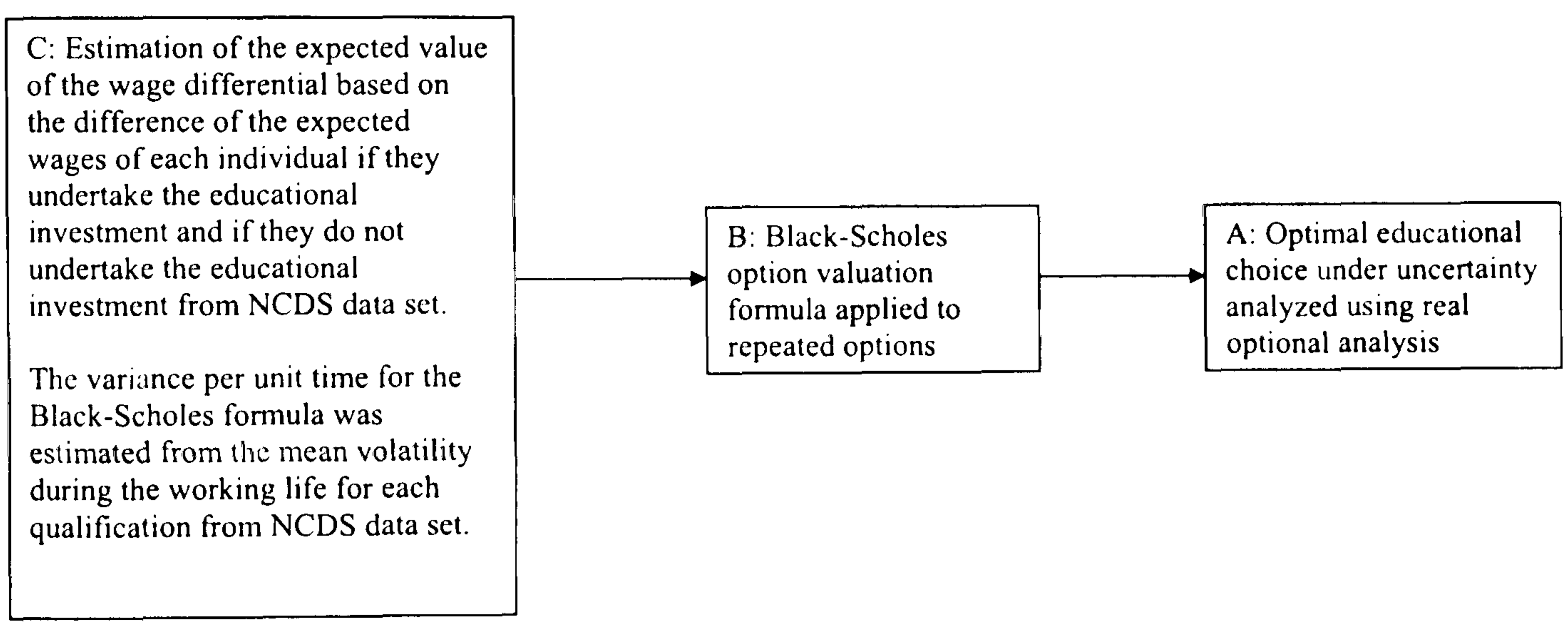


Figure 2.5 Modelling optimal educational choices under uncertainty using a real option analysis



Chapter 3 Empirical Evidence on Education Choices under Uncertainty

3.1. Introduction

We analyzed the education choices model in a great detail in Chapter 2 by using two methods, namely B-S option method and multinomial logit model. However, we do not yet know whether they are suitable for our empirical dataset. This chapter aims to test the two models and ascertain which model explains the empirical data better. After comparing the educational choices from the series options model and the multinomial logit model of education decisions, we can also shed light on the following questions: Does market uncertainty play a role in explaining education choices? Will non-pecuniary benefits affect individuals' schooling choices? and Are current education decisions optimal?

Alongside answering the above questions, we find quite a few innovative empirical findings from our dataset that is composed of six waves of the longitudinal dataset National Child Development Survey (NCDS). The main contributions of this chapter can be summarized as follows: Firstly, it will estimate ex-ante wages, which complements current research on British ex-ante wages estimation. William & Gordon (1981)'s innovative work shows individuals prediction on their future earnings follows a parabola shape. Webbink & Hartog (2004) use real wages observations confirms students can predict their future earnings subjectively. These empirical findings strengthen our confidence on assessing individuals' ex-ante wages objectively through individuals' characteristics. Previous empirical evidence on discrete choice structure was mainly based on ex-post wages under certainty, which cannot avoid the problem of selection bias. They (e.g. Keane & Wolpin, 1997; Oosterbeek & Ophem, 2000) attempt to model post-secondary choices conditional on past decisions of students and schools. Our model predicts the ex-ante return conditioning on individuals' secondary school curriculum, academic attitude, family background as well as graduation probability. The difference test between ex-ante and ex-post wages demonstrates that our estimation method is quite close to the real value.

Secondly, this chapter is the first empirical study to examine the optimal educational choices under uncertainty. By employing the B-S option model, we overcome the untestable attrition of uncertainty in human capital investment as advanced by Levhari & Weiss (1974) and Williams (1979) and can estimate individuals' optimal education choices under uncertainty. NCDS dataset recording of individuals' comprehensive personal characteristics and employment information at ages 6, 11, 23, 33 and 41 satisfies our quite demanding data request.

Another contribution of the chapter is to estimate sequential education-going behaviour in both academic and technical institutions. Dearden et al (2002) state the design of education and training in both academic and technical qualifications are determined by the relative benefits of each. It is meaningful to discuss them separately based on individuals' real situations. To my knowledge, it is the first empirical work to tackle this problem. Appendix A depicts all possible education paths that individual may take to a certain level of qualification. Developing the standard multinomial logit model, we can describe various educational paths that are depicted by Appendix A under uncertainty. The series of option models can also evaluate the separate probability of achieving each qualification level.

Fourthly, the comprehensive information in the NCDS allows us to first explore some empirical evidence that has been predicted by theoretic models a long time ago. The life-time earnings will be less for those who acquired qualification later in their lives. In addition, those who entered tertiary education with several years experience generally come from poor family background. Since our multinomial logit model also considers individuals' different discount rates, we can also predict the situation with several years' break between schooling.

Finally, we manage to estimate the effect of non-pecuniary utility on individuals' education choices. This is also the first study to evaluate utility by some empirical variables, such as ability, family background and individuals' attitudes towards study. Through adding a discount rate and a non-pecuniary utility function, we find that the estimation results are closer to the reality and effort costs of study do discourage individuals from attending education.

The following part of the chapter is organized as follows: section 2 describes the data, descriptive statistics and ex-post wages regression. Section 3 provides the regression results for ex-ante return. The corresponding results from the multinomial logit model and real option model will be analyzed in section 4 and section 5, respectively. Section 6 discusses the limitations of empirical data and further extensions.

3.2. The Data

3.2.1 Data description

We will employ NCDS to assess the micro factors that may affect individuals' expected returns. The NCDS data set is a continuing, multi-disciplinary UK longitudinal study, which examines 17,000+ babies who were born in a particular week in March, 1958. Follow-up surveys of the whole cohort were carried out at ages 7, 11, 16, 23, 33, and, most recently, in 1999 at the age of 41. The study includes the detailed information about parental background, secondary school curriculum and ability in early waves and about earnings and other employment information in later waves, which allows us to assess the behavioural and cognitive development of the children as well as labour market behaviour. The detailed family background variables include parents' highest qualification, socio-economic background and everyday newspaper reading time. It also records respondents' examination information, such as the total number of GCSE grade A-C, the math and English score at GCSE and A-level. The study even surveys the parents' interest in children's education and respondents' willingness to study by means of the question of when and why they plan to leave school. Since the respondents were born in the same year, it helps us to control the general economic environment and cohort effect. All the individuals in the data set face the same education policy, the same demand and supply conditions in the labour market, the same wage levels, which allows us to pay more attention to the effect of uncertainty on wages. In addition, the dataset inscribes wage level and qualification level at their 30s and early 40s, when they entered a stable period and such kinds of data are not easy to obtain.

After connecting the six waves' survey together and deleting the observations that did not report all the necessary information, the sample size sharply decreases to 581. The variables that will be used in this chapter are summarized in Table 3.1.

Insert Table 3.1

Table 3.2 describes the respondents' characteristics of ex-post education distribution by the father's occupation, education and family culture environment when the children are seven years old, secondary school curriculum and individuals' monthly ex-post wages at age 41. As expected, higher qualifications correspond, on average, to more favourable family background and better examination scores in the statistic description. However, for the technical degree and technical college, family background may not seem to be such an obvious factor. Current job earning variables also suit the logic of the assumption that the mean increases with the level of increase in education. For the same level of education, the wages for academic qualifications are higher than for technical qualifications. One possible reason is that the academic qualifications require higher ability and initial human capital, which may determine the employees' later performance in the labour market.

Insert Table 3.2

3.2.2 Ex-post wages analysis

In order to estimate ex-ante wages, an estimate is needed of the effect of individuals' characteristics on ex-post wages. General human capital theory (Becker, 1962 and Mincer, 1974) explains that people forego their possible earnings (including all costs of schooling) and accumulate skills and knowledge in school in return for receiving higher lifetime earnings. The expected wages are

$$EW_i = \alpha + a_1 s_i + a_2 T_i + a_3 T_i^2 + \varepsilon_i \qquad (3.1)$$

where s represents schooling and T is experience. According to this equation, individuals with the same qualifications and experience should earn the same wages, which is not the case in reality. Many scholars argue about the large constant ε_i , which expressed the unexplained facts attributed to individuals' heterogeneity, especially ability.

In this chapter, we will control for the qualification level[14] and some job and personal characteristics, such as gender and ability. Although the return to some job characteristics variables is significant, they are balanced in a competitive labour market. Individuals can trade off leisure time, job security and other welfare or benefits with a higher salary. To find out the return to each individual character, we regressed the following equation:

$$w_{inm} = Ew_{nm} + \eta_1 A_{inm} + \eta_2 Z_{inm} + \eta_3 O_j + \varepsilon_i \qquad (3.2)$$

where w_{inm} is individual i's natural log wages for qualification m and type of education n. m=1, 2, 3 express A-level, first degree and post degree. n=1,2 denotes academic and technical, respectively. Individuals are

[14] using qualifications instead of actual years as a measure of human capital can avoid repeated years, inefficient years and drop-out years.(see Sloane et al, 1999 for detail)
Appendix B lists how do we group the showcard qualifications into technical degree, university degree, technic A-level and academic college.

expecting to earn Ew_{nm}, the market rental price for qualification m and type of education n without experience, which can be represented by the average natural log wages at the same level. The idea to employ natural log of real wages instead of real wages comes from the Mincerian equation and the basic human capital model (e.g. Becker, 1962). A_i represents individual i's demographic characteristics include gender, family background and ability. In this chapter, family background was represented by father's social class (fsc); the Best Grade of Maths (GMath) and English in O-levels (GEng) were used to estimate ability. Fsc, GMath and GEng were classified into several levels in order to reflect the affecting power of each level to their future wages. Surely test performance at ages 7 or 11, which is given by NCDS is a better estimate of ability, however they do not play a significant role on their later wages and I have to use O-level scores to estimate ability.

Z is a vector denoting other job characteristics variables, for instance the size of the firm, company location, type of organization. O_j represents the standard occupation code. All the occupations were grouped into 9 sectors according to the showcard of the NCDS. They are managers and senior officials (occu1), professional occupations (occu2), associate professional and technical occupations (occu3), administrative and secretarial occupations (occu4), skilled trades occupations (occu5), personal service occupations (occu6), sales and customer service occupations process (occu7), plant and machine operatives (occu8) and elementary occupations (occu9).

One may notice that we did not add a quite important variable, experience into this estimation. Ben-Porath (1967) states that life-time wages trajectory is diverged by qualification: the higher the academic qualification, the higher the return to experience. Since we control the qualification level in our estimation method and all the respondents are in the same age group, the return to experience for the same cohort within the same level of qualification may not be diverged. In addition, the return to experience in our series of the option model is not that important. I use the average growth rate for each qualification to represent the return to experience.

Of course, there are other variables that may affect individuals' wages, such as teaching quality, university reputation, union membership and labour contract. We group it all into the constant. Table 3.3A is derived from regressing lnwages in the equation (3.2) when individuals are at the age of 33. The reason we use the wages at age 33 instead of at 23 or 41 is that the wages at age 33 are quite typical during the life-time. In addition, the regression results show that there is little effect of personal characteristics on individuals' wages at age 23, but quite a significant effect on wages at age 33 and 41. Since we will use these wages to estimate the life-time wages, it is more suitable to use wages at age 33 instead of 23.

The regression results in Table 3.3A show that gender plays an important role in all levels of education. Being a male may contribute up to 30.7 per cent to the wages of technical college level graduates. Besides gender, the most important factor in individuals' wages determinants is the best English and mathematics grade at O-level, which corresponds to the findings of Dolton & Vignoles (2002). However GMath and GEng play various roles for different qualification level. Generally good GMath and GEng will affect future earnings positively for all kinds of qualification, except technical degree; but poor GMath and GEng will play a negative role for academic university only. This implies individuals with poor academic background will not be in an advantaged position even if they enter the academic university.

Insert Table 3.3

In a similar way to the findings of Webbink & Hartog (2004), there is no relationship between family background and future wages. The type of organization and firm size may have various effects on individuals' wages for different qualification. Working in the central government may influence university graduates' wages negatively, but have a positive impact on postgraduates' wages. Compared to working in a small company, large company employees generally earn more than those who are self-employed or employed by small companies in the UK. The statistics shows that post degree and first degree holders generally occupy the occu1 to occu4 occupational categories and individuals with a lower qualification generally have a lower social class of occupation. Managers or professional occupations will only have a positive return to higher degree holders.

3.2.3 An ex-post analysis of educational choices

Chen (2003) and other authors argue that financial restrictions can be a main factor that blocks individuals from attending university and they also employ all kinds of models to explain their ideas. However until now no empirical study has examined this problem. Detailed questions in the six waves of the NCDS dataset allow us to examine whether family background plays an important role in students' education choices and who will attend universities after several years of experiences.

Table 3.4[15] gives us the answer as to why some talented individuals do not attend higher education directly. Regardless of the first degree or higher degree, those who attend universities after a break of several years have mathematics and English grades that are average higher than those who attend universities directly. But those who attend universities after a break up of several years have an average unfavourable family background than those who attend universities directly. This confirms Chen (2003)'s findings that financial restrictions constrain some talented youngsters attending higher education. It is also quite interesting that the average secondary score of the higher degree is lower than that of the first degree, which suggests that second best students want to remedy their disadvantage through receiving more education.

Table 3.4 Individuals' characteristics comparison

[15] This table and the following analysis did not consider the students from Scotland, since their education and examination system is different.

	GMath	GEng	Fsc	Fce	lsa	fsch	Samp Size
Acquired first degree before 23	2.22	2.16	3.6	2.6	3.8	20.2	193
Acquired first degree after 23	2.21	2.12	4.4	3.1	3.6	18.6	40
Acquired higher degree before 23	2.27	2.23	3.7	3.4	3.2	19.9	13
Acquired higher degree after 23	2.17	2.20	3.7	2.9	3.4	17.8	64

In addition, we are interested to find out whether attending universities later will affect their wages. In order to answer this question, I summarize respondents' wages at ages 23, 33 and 41 for those who attend university directly and those after a break up of several years (see table 3.5 below). The results demonstrate the basic human capital theory of Becker (1964) and Ben-Porath (1967) that those who get qualification in their earlier life enjoy a higher level of average wages than those who get a qualification later.

Table 3.5 Hourly Wages Comparison

	wages81	wages91	wages99
Acquired higher education before 23	2.01	8.08	13.97
Acquired higher education after 23	1.79	7.45	13.25

3.3. Ex-ante Return under Uncertainty

Since individuals make education investment decisions before entering the labour market, ex-ante wages will have more guidance than ex-post wages. In addition, Freeman (1971), Dominitz & Manski (1996) and Webbink & Hartog (2004) show that individuals can predict the structure and size of the future wages accurately based on their own secondary education mark, family background, academic interests, etc. The evidence gives us the theoretical background to model the ex-ante wages according to available information. According to Webbink & Hartog (2004), gender, grade in secondary education and degree of faculty play a role in prediction and also in ex-post wages while family background does not. Students from wealthy families do have a higher attendance rate for higher education according to our statistics. Since we group the family background into different ranks, it may give us more accurate information on prediction and will be considered in the wages analysis.

The estimation process is as follows: we first examine the effect of personal idiosyncrasy on future return by ex-post wages (equation 3.2) for each level and type of qualification, which has been done in table 3.3A. The process of classifying individuals into a particular regime may generate a selection bias to the effects of individuals' characteristics on wages. A multinomial logit model was used in table 3.3B to examine the selection bias. In other words, whether there is a significant impact of individuals' characteristics to individuals selecting a particular regime. The results show only NOAC plays a significant role on the selection process and the selection bias of other variables are not significant. Secondly, we estimate the ex-ante wages by summarizing all the products of coefficients that we regressed in step one and the value of each characteristic.

NCDS data did not record whether individuals drop out during their study or not, so we do not have perfect observations to represent the probability to graduate. And so the probability to graduate will not be considered in the empirical analysis. Since the dummy variables on academic background can reflect individuals' with different abilities' expected wages, this will not significantly affect our estimation results.

The estimated results of the ex-ante return were reported in Table 3.4. From Table 3.4 we notice that the estimated return can predict expected wages very well. The coefficients of the estimated return for most of the variables have the same direction and size as the coefficients of realized return. Parents' social background and secondary English grade play a more important role in estimated return than in realized return. In addition, the estimated return is not distributed as polarized as realized return as we expected, since some parts of wages are the compensation for high consumer price index, high job demands or insecurity.

We then compare the lnwages differential between objectively predict lnwages and realized lnwages by equation (3.4).

$$w - w_e = (\alpha - \alpha_e) + (\beta - \beta_e)A_{ij} + (\gamma - \gamma_e)X_{ij} + (\varepsilon - \varepsilon_e) \qquad (3.3)$$

where α is the constant, β and γ are the coefficients for ability and individual character X respectively. The subscript e means predicted wages. We include the father's education, individuals' motivation, occupation and gender into individuals' characteristics. Most of the estimated variables go in the same direction as the real results. The predicted wages for students who will work as managers and senior officials are higher than the average, as is the case in reality. The third column in Table 3.6 shows that the estimated return can predict future return very well. Most of the variables are explained by the estimated return, except for the average secondary mark and extrinsic motivation. We assume that individuals with higher secondary marks are less likely to drop out and earn a corresponding level of wages. However, in reality, average secondary marks have no relations with future wages.

Insert Table 3.6

Our prediction method overcomes the subjectivity of self-prediction that Webbink & Hartog (2004) discussed in their pioneering work, such as students from higher income families tending to predict higher salary than they actually earn and students from science subjects expecting more income than they can realize. However, our objective predicting method may not have as much information as the subjective method does.

3.4. Optimal Choices from the Multinomial Logit Model

In the theoretical analysis of chapter 2, individuals will select the optimal educational choice by computing backwards to the starting decision from looking at the expected payoffs in the outer branches and then moves inwards. In other words, individuals educational choice in next period (t+1) is based on the expected utility of all kinds of choice in period t+2, the choice in t+2 is selected according to the maximal utility in t+3 ... Since the selecting process was based on qualification instead of years as analyzed by Keane and Wolpine (1997) and only four possible educational choices were considered in this chapter, I can simplify individuals dynamic selecting process as a static model of individuals' optimal highest qualification. In other words, which qualification is individuals' optimal highest qualification, Acoll, TechC, AFD or TechD? In this sense, individuals optimal educational choices can be estimated by a standard multinomial logit model. The optimal choice mn is chosen if and only if the choice mn can provide the highest utility to individual i and is expressed as:

$$\Pr[k_i = mn | I(t)] = = \frac{\exp(\ln V_{imn})}{\sum_{k=1}^{4} \exp(\ln V_{ik})} \tag{3.4}$$

In chapter 2, the utility for each choice includes future income as well as the consumption value of greater education. The consumption value in different choices has different possible values and may also vary across individuals. If we take the logarithms of equations (2.10) and (2.21), we get

$$\ln V_{ik} = \ln E(U(Y_{ik0})) + \chi_i \ln f_1(k) - \ln(\theta_i - (1-\rho)\varphi_k) \qquad (3.5)$$

where $\chi_i = \mu_i(E_i + P_i + A_i) = \mu_1 E_i + \mu_2 P_i + \mu_3 A_i$

$\theta_i = -b_1 \varpi_i$

The logarithm of the expected income utility $\ln E(U(Y_{ik0}))$ was represented by the expected logarithm of ex-ante wages in eqn (2.10) that we discussed in the previous section. $f_1(k)$ is the consumption value, which increases with qualification level. "Taste of schooling" χ_i is assumed to be a linear function of individuals' self interest on study (E_i), family background (p_i) and ability (A_i). E_i can be represented by logarithm of self preferred school leaving age, p_i is denoted by the logarithm of family culture environment and ability A_i is the logarithm of number of O level grade A-C. In this thesis, $f_1(k)$ only refers to current consumption value, since the future consumption value (e.g. better living condition) is hardly observed and can be replaced by wages. I use minus tuition fees to represent current consumption value. Wages growth rate φ_m can be estimated by the average annual growth rate in the mean of the log of annual earnings for each qualification.

In order to estimate equation (3.4) and (3.5), we have to work out the value of θ_i and χ_i. Card (1999) gives the exact forms for discount factor and consumption value considering individual's heterogeneity in a covariance and variance form. However, these forms are too difficult to be measured by econometricians. Oosterbeek & Ophem (2000) estimated the discount rate by assuming current schooling choices are the best educational choices and use maximum likelihood function to estimate the discount rate. In this thesis, we have to assume individuals' current qualification level is the best education choices. Through a logit model, we can work out μ_1, μ_2, μ_3 and b_1, which when multiplied by each individual's characteristics will be the discount factor and consumption value for each individual.

In the model, the effect of family background and personal ability will be applied to the taste for schooling parameter (χ_i) and the discount rate (θ_i) separately. Oosterbeek & Ophenm (2000) said even if two vectors have all their elements in common, their regression coefficients can still be identified. To make identification not solely dependent on the functional form, I also randomly assign different elements to χ_i and θ_i. The family culture environment and the father's social class were allocated to χ_i and θ_i, respectively.

Table 3.7 lists the mean value of the discount factor and taste of schooling for diversified qualifications. Consistent with our expectations, the estimated value of discount rate rises when the qualification decreases and for the same qualification level, the value for the technical type is a little higher. The school preference tells the same story, namely that the mean increases with qualification level and those who attend university have the highest value.

Table 3.7 Estimated mean consumption value and discount rate for each qualification

	AColl	TechC	AFD	TechD
Discount rate	0.026 (0.016)	0.028 (0.021)	0.019 (0.032)	0.021 (0.051)
School preference	-0.148 (0.098)	-0.330 (0.119)	0.011 (0.003)	-0.179 (0.029)

3.4.1 Education choices without considering consumption value

Firstly, we will estimate the multinomial logit education choices model (equation (3.4)) without considering the consumption value. In other words, the utility is represented solely by earnings as showed in equation (3.6).

$$\ln V_{ik} = \ln E(U(Y_{ik0})) - \ln(\theta_i - (1-\rho)\varphi_i) \qquad (3.6)$$

After substituting equation (3.6) into equation (3.4), we obtain the estimation of individuals' optimal educational choice. Through counting the exact number in each qualification estimated by equation (3.4) and (3.6), we derived the Figure 3.1.

Figure 3.1 Distributions of Postsecondary Activity Choices without considering non-pecuniary utility

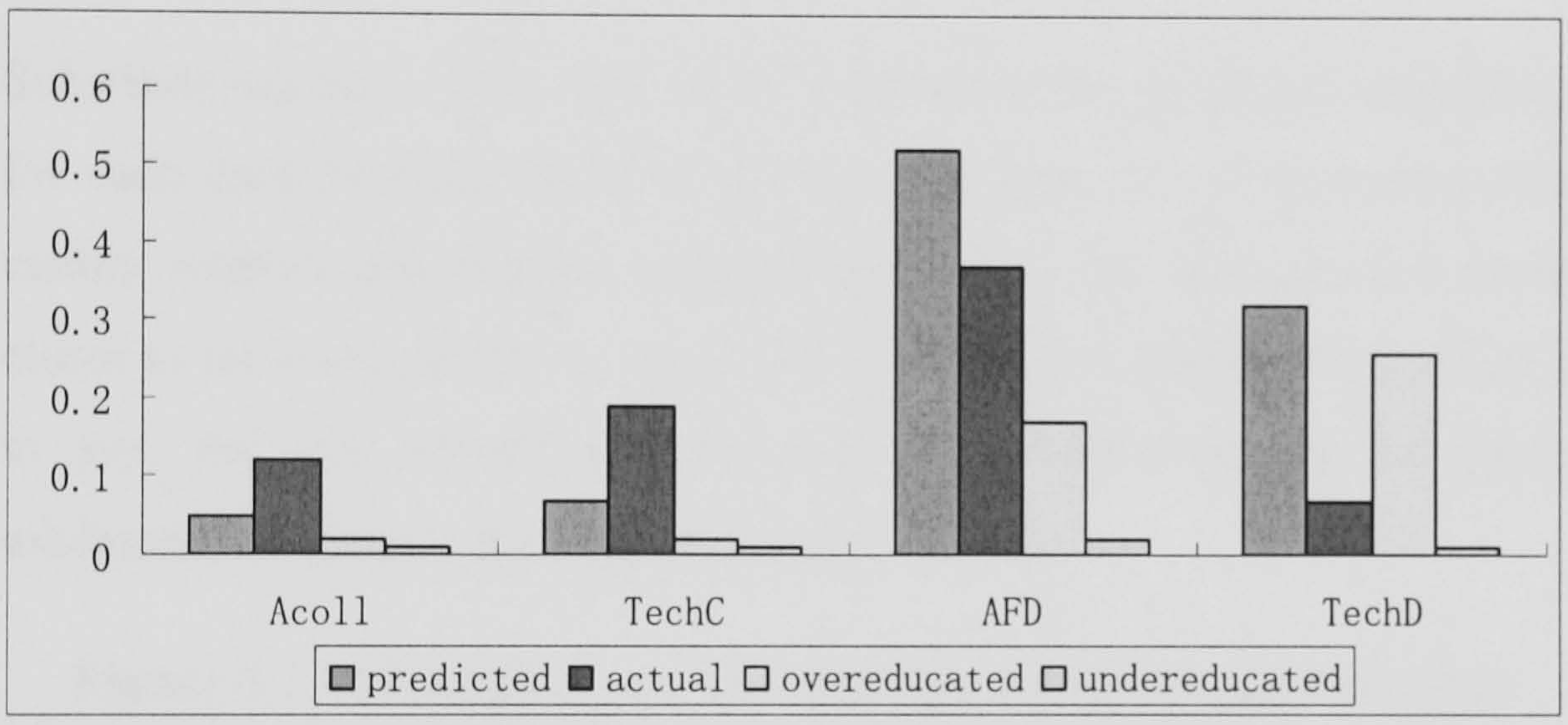


Figure 3.1 describes the predicted results for all levels of education from the multinomial logit model. The predicted attendance is a little bit higher than the actual figure for both academic and technical qualification. The third and fourth column for each qualification are the about overeducated and undereducated percentages compared to the predicted optimal results, respectively. This figure can examine the difference between predictions and actual results as well as whether we group the people that belong to other qualifications into this particular qualification (e.g. prediction accuracy). If our optimal prediction is accurate, there are 17 percent more individuals who should attend university. The prediction about technical attendance is comparatively inaccurate and much higher than real results. According to our theoretical model, individuals will attend either academic or technical college since they can acquire higher wages by doing that. However, in reality there are all kinds of reasons to discourage

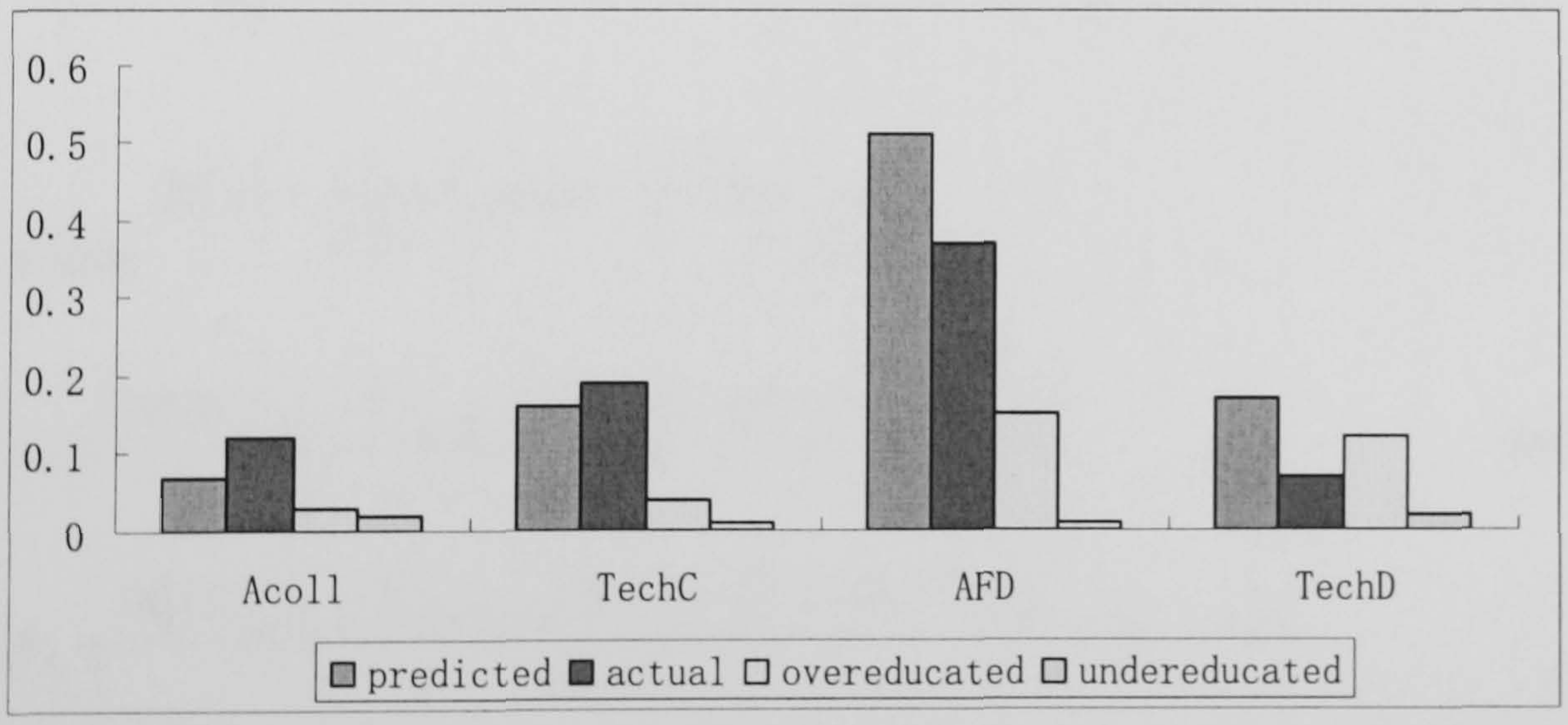
individuals from attending higher levels of education. We are then interested in finding out whether considering consumption value can increase the accuracy of prediction.

3.4.2 Education choices considering consumption value

Predicted low college attendance and comparatively high university attendance in Figure 3.1 may partly show that wage is not the only factor affecting education choices. In this section, we take into account the consumption value to see whether the function after adding this variable can provide a better fit.

Substitute equation (3.5) into (3.4), will derive the predicted attendance for each qualification, which is showed in Figure 3.2. Compared to the results without considering consumption value, the prediction is much closer to the realized results, especially for technical qualifications. That is to say, the low attendance of technical qualification can be partly explained by respondents' low consumption value.

Figure 3.2 Distributions of Postsecondary Activity Choices after considering non-pecuniary benefits



3.5. Educational Decisions under Uncertainty in a Real Option Model

Though adding individual utility functions into the multinomial logit model largely improves the accuracy of prediction, the predicted attendance is still higher than actual attendance. It is interesting to figure out whether uncertainty can explain the difference. Following the discussion of chapter 2 and estimated ex-ante wages in part 2, we can finish this task by the real option model.

Expected ex-ante return will enter the real option model from two aspects, one is education costs and the other is exercise price. In this model, we will compute each individual's net payoff of an investment opportunity in each year based on available information at that time. We assume that all the individuals will obey the law and finish the compulsory secondary schooling. Therefore the schooling decisions are between whether to attend A-levels or not and whether to attend universities or not. The percentage of individuals who attend post tertiary education is quite small, hence we will not consider it. For each level of education, individuals could choose academic or technical education. As analyzed in chapter 2 we treat investment in education as buying a series of European call options with a fixed one year contract and the real option model on how to choose an optimal education level m is adjusted to yield

$$H(Y_{im}) = \sum_{l=0}^{N} (Y^e_{im(l+1)} - Y^e_{i(m-1)(l+1)})(1+\lambda_m - r)^l N(x_1) - LN(x_2) \tag{3.8}$$

where $N(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp(-\frac{1}{2}x^2)dx$ and

$$x_1 = \frac{\ln[(Y^e_{im(l+1)} - Y_{i(m-1)(l+1)})(1+\lambda - r)^l / L] + \sigma^2\tau/2}{\sigma\sqrt{\tau}}$$ and

$$x_2 = \frac{\ln[(Y^e_{im(l+1)} - Y_{i(m-1)(l+1)})(1+\lambda - r)^l / L] - \sigma^2\tau/2}{\sigma\sqrt{\tau}} = x_1 - \sigma\sqrt{\tau}$$

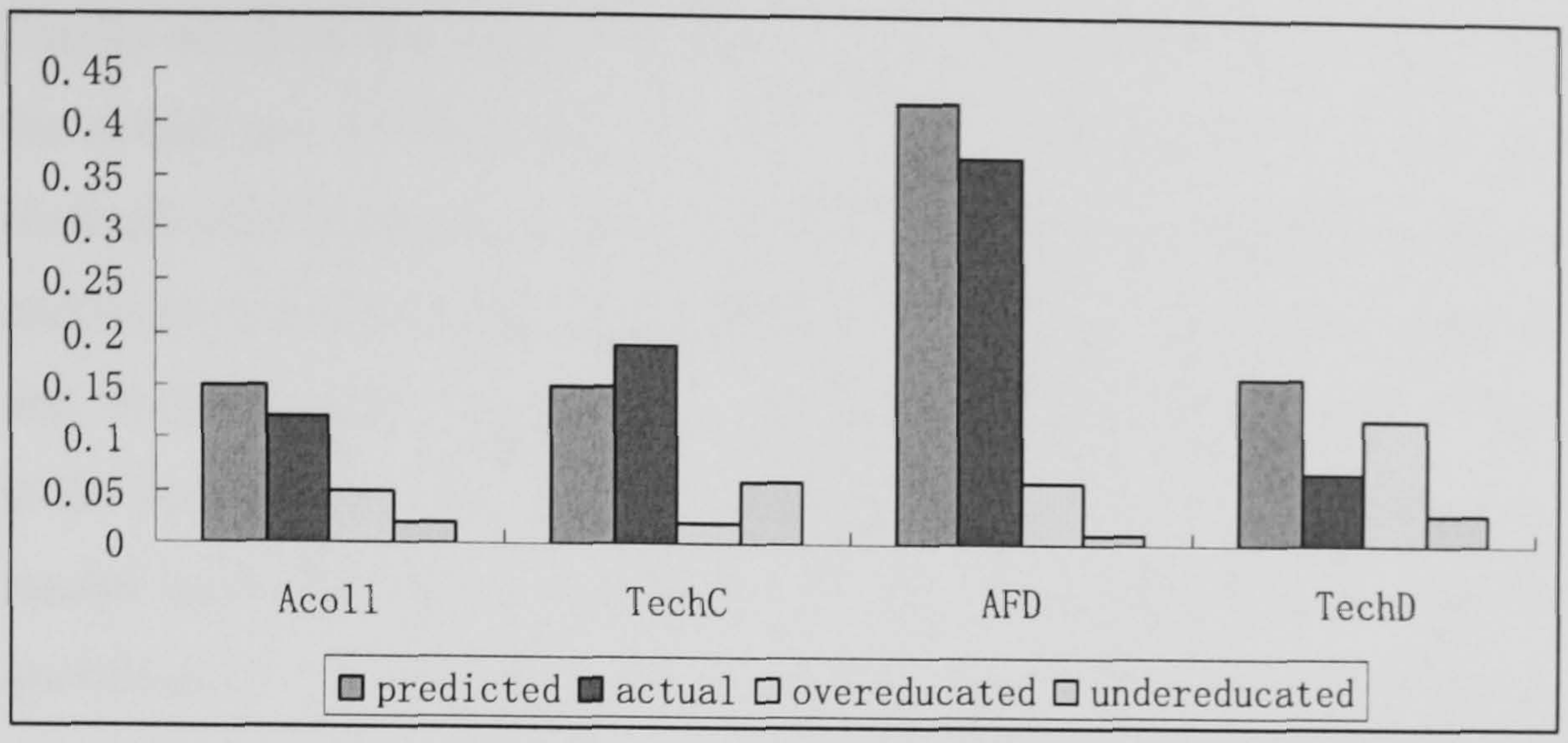
where $Y^e_{im(l+1)}$ is the expected beginning wages for qualification m and individual i, λ_m is the annual increasing rate for qualification m and r is the interest rate. σ is the mean volatility during the working life for each qualification m, which is distributed from 0 to 1. Here σ is calculated by the weighted average wages and is the same for people sharing the same qualification. Weighted average wages for qualification m was estimated by the beginning wages at age 23 and deflated wages at age 33 and 41. And the deflated wage at age 33 and 41 were processed by the average annual wages growth rate for each qualification m.

Since the NCDS records individuals' wages at ages 23, 33 and 41, we can use the life-time average wage volatility to denote σ instead of annual wage volatility. Period l equals 0, 1, 2, ..., N and the value of N depends on m. N=46, 44 for A-level and first degree separately.

For simplicity, the education costs were estimated by foregone wages only. Through comparing the value of $\sum_{r=0}^{N} T(Y_{im})$ with option costs C, we derived the Figure 3.3. [16] The estimated results reported in Figure 3.3 are surprisingly similar to the actual results in both direction and size for academic attendance. Since we do not consider the tuition fees and living expenses in our estimation, the predicted results of first degree attendance is a little bit higher than the actual situation. What is more, in our prediction, around 10 percent of individuals did not attend university directly after A-levels, which cannot be estimated by Hogan & Walker (2005).

Figure 3.3 Estimated distributions of postsecondary activity choices by wages real option model

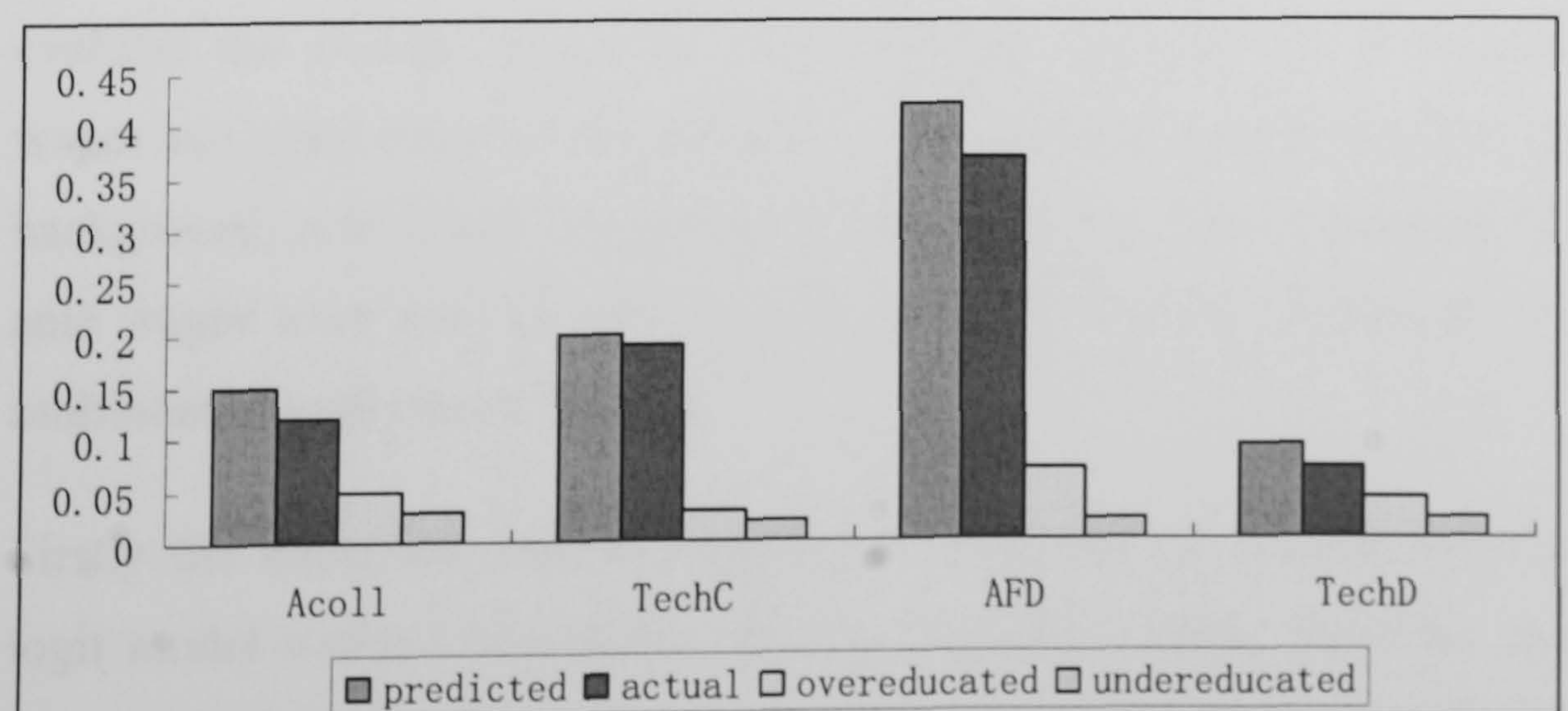
[16] Figure 3.3 was obtained by comparing individuals' expected wages for different educational choices and assumed individuals will select the one which can maximize their expected income.



The B-S wages estimating method can predict uncertainty and the associated wage distribution very well, but cannot figure out the type of education (i.e. academic or technical education) and individuals' preference. From the statistics in section two, we know that individuals from different family backgrounds may have different discount rates and consumption value for investing in education resulting in various educational choice routes. We have then tried to substitute the wages in equation (3.8) by utility function and re-evaluate the value of option to see whether this will provide a better fit for technical qualification.

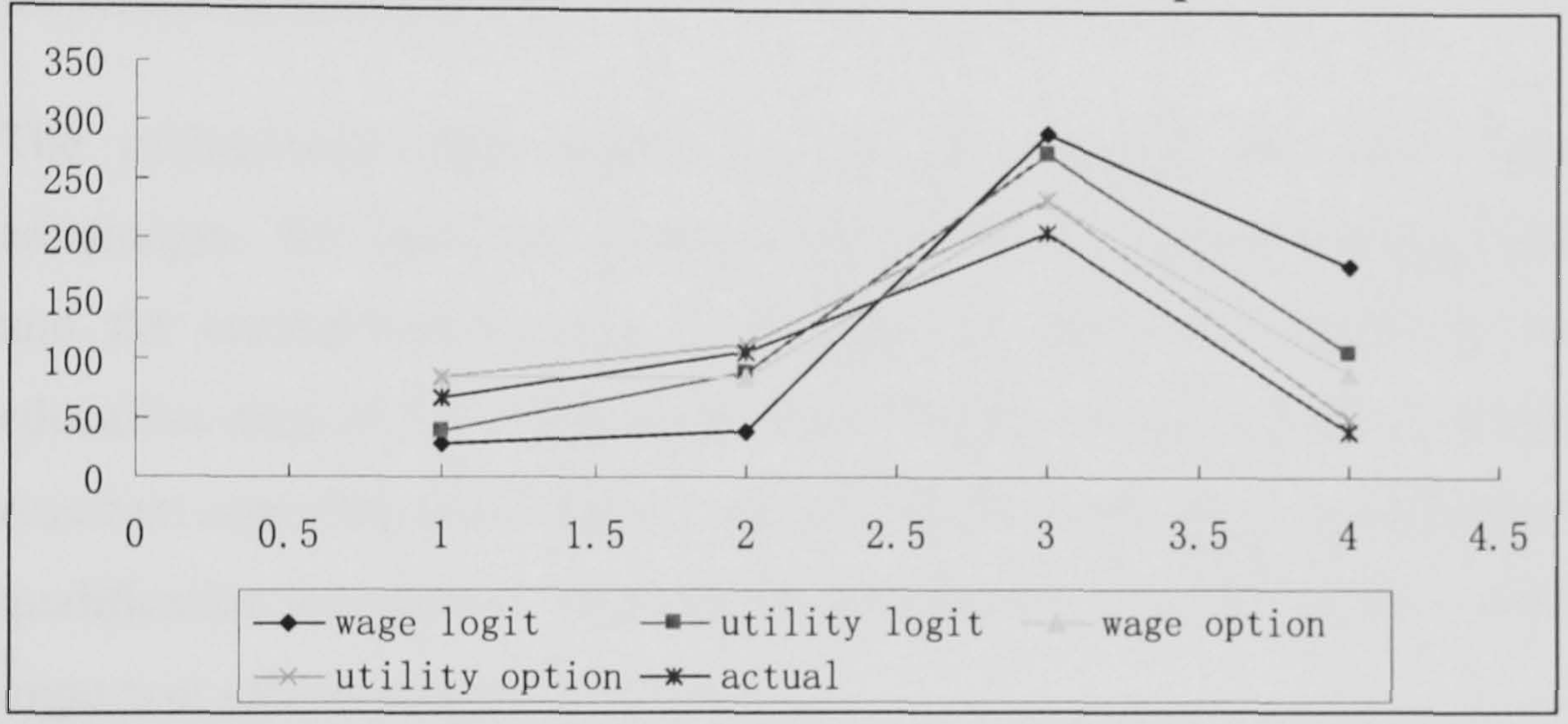
The estimation results in Figure 3.4 for technical qualification is significant improved after considering current consumption value or effort costs. However, adding consumption value has not much effect on the results for academic qualification.

Figure 3.4 Estimated distributions of postsecondary activity choices by utility real option model



Finally we draw the Figure 3.5 to compare the four estimation results with the actual one. From Figure 3.5 we can conclude that those estimation methods considering consumption value give a better fit than those did not and those considering the uncertainty provide closer results than those did not. In other words, the estimation method by the utility real option model is the one which is the closest to the actual attendance. In addition, these results imply that the main concern for individuals who choose academic qualifications is wage volatility, while those who select technical schooling put more weight on the effort that will be spent on study and have a high discount rate.

Figure 3.5 Three estimation methods comparison



3.6. Conclusions

This chapter tries to examine the education choices by using the real option model and multinomial logit model that we discussed in chapter 2 separately through the NCDS dataset. In order to realize that goal, we first evaluate the ex-ante wages for each education choices. The predicted wages are quite close to the realized wages through considering family background, individuals' ability and academic attitude. The estimated ex-ante wages were used in both the B-S education choices model and the multinomial logit choices model.

Firstly the education choices model was evaluated by the multinomial logit model without considering the non-pecuniary utility function. The prediction results have some distance with the actual results, especially for technical qualification. After adding consumption value and effort costs into the regression model, the percentage of attending education is quite close to the real results. This suggests that non-pecuniary utility does affect individuals' career move. We then examine the effect of uncertainty on education choices by the B-S without considering utility value model. The prediction results are quite similar to the real results for the academic qualification, but not the technical one. This suggests that volatility is not the main reason to stop individuals from being educated for technical qualifications but is the main factor to discourage individuals from attending academic qualification. Following this, wages were substituted by utility value in the B-S model and the prediction was much improved for technical qualification.

The multinomial logit model and the B-S model both have their advantages: the first can consider individuals' diversified discount rate and the second can manage to estimate the effect of uncertainty on education choices. By comparison, the B-S utility model provides a better result on capturing individuals' behaviour for both academic and technical qualification attendance. This implies uncertainty and utility play a very important role on education choices.

One important finding in this empirical study is even if we consider all the necessary factors, the attendance of technical qualification is still lower than the optimum level. The predicted comparatively high attendance can explore two problems: one is not everyone is a rational investor and farsighted. The other is the asymmetric information in labour market. Students from poor family background have few ideas on the advantages of obtaining qualifications and would not like to invest in education.

Despite the contribution of the chapter to current research, there is a long agenda in predicting education choices. In our empirical education choices model, we can either predict education choices under uncertainty without considering discount factor or education choices under certainty by considering discount factor. Later research should try to release the assumption of risk neutrality in the real option model and combine both the uncertainty and individuals' preference in their education choices model. Besides, we cannot depict a vivid picture of life-time career moves. In other words, the estimating method did not capture individuals' behaviour in each year and did not consider the type of education (full-time or part-time) and the continuity of education (i.e. whether broken up or suspended for several years between studies). This should be designed in later education choice models. Another area that needs to be devoted is specialty. How to estimate or even suggest individuals' college and university majors and relate it to both individuals' prospects and the country's strategic development direction may be the subject of future study.

Appendix A

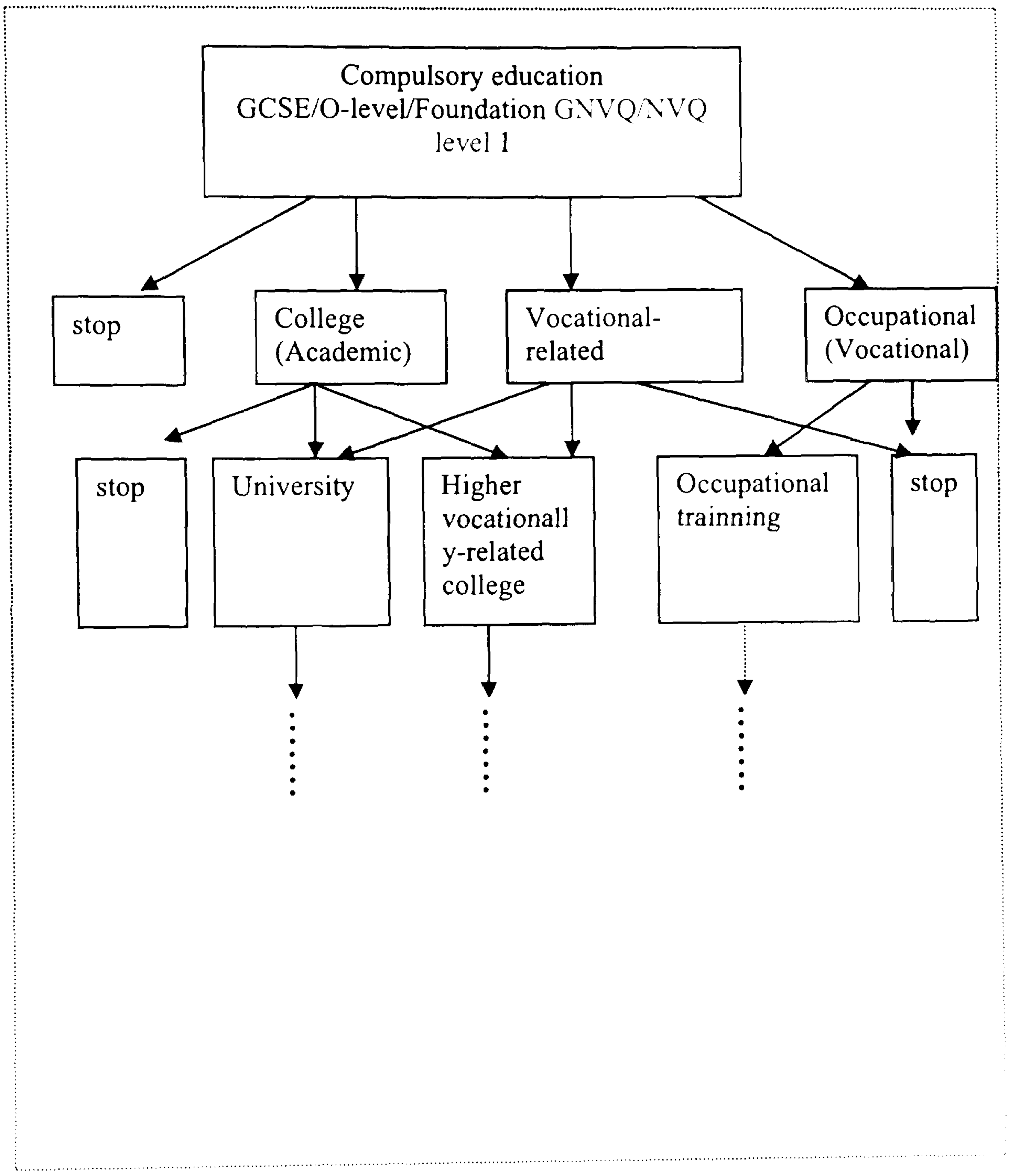

Compulsory education
GCSE/O-level/Foundation GNVQ/NVQ
level 1
stop
College
(Academic)
Vocational-
related
Occupational
(Vocational)
stop
University
Higher
vocationall
y-related
college
Occupational
trainning
stop

Table 3.1 Variable description

variable	Description
Fce	Family culture environment, which is estimated by the weighted average of two separate variables (i.e. mother and father's spare time reading newspaper and father's attention to child's study) and two variables share the same weight Mother and father's spare time reading newspapers were used to estimate this variable. 2. yes, most days. 3 yes, occasionally 4. hardly ever Father's attention to Child's study. 1. care a great deal 2 care 3 care, but not much 4. do not care
Fsc	Father's social class when child is 7 years old. Fsc1=1 if social class= I, Fsc2=1 if social class=II, Fsc3=1 if social class=III, Fsc4=1 if social class= IV or below.
Fsch	Father's school leaving age
Noac	Numbers of O-level grade A-C from 1 to11
GMath	Best grades obtained in Maths at O-level. GMath1=1 if individuals got grade A, GMath2=1 if individuals got grade B or C, GMath3=1 if individuals got grade D or below.
GEng	Best grades obtained in English at O-level. GEng1=1 if individuals got grade A, GEng2=1 if individuals got grade B or C, GEng3=1 if individuals got grade D or below.
Lsa	age likely to leave school study
Hquali81	Highest qualification obtained before 1981
Hquali91	Highest qualification obtained before 1991

Table 3.1 variable description-continued

AHD	University or CNAA higher degree (including master and doctorate)
AFD	University or CNAA first degree
APdpl	University or CNAA post graduate diploma
Adpl	University or CNAA Diploma or certificate (including teaching Cert.)
Acoll	Academic college (e.g. 3 A-level graduates)
TechD	Vocational or technical degree
Techdpl	Polytechnic diploma
TechC	Technic A-level
Pricom	Private company
Cengov	Central government
La/lea	Local authority/ Local education authority
Natindu	National industry

Table 3.2 Statistics description of individual characteristics

Value Label	Fce	Fsc	Fsch	GMath	GEng	Ex-post wages in 1999 per hour
AHD	2.17 0.03	4.03 0.3	21.25 0.3	1.80 0.10	1.76 0.08	16.50 2.67
APdpl	2.20 0.09	4.11 0.06	19.54 0.19	2.60 0.15	2.34 0.07	15.98 1.99
AFD	2.19 0.02	4.01 0.09	18.62 0.29	2.23 0.05	1.99 0.04	17.78 2.59
Adpl	2.25 0.06	4.13 0.07	18.16 0.24	2.74 0.18	2.03 0.10	15.94 1.83
TechD	2.20 0.04	4.46 0.13	17.6 0.38	2.86 0.13	2.65 0.15	16.01 2.01
Techdpl	2.28 0.07	4.46 0.21	17.99 0.65	2.38 0.07	2.44 0.15	15.87 1.79

Note: the second rows of each cell are std. Err.

Table 3.3A Regressing results from log hourly earnings

	ACoII	TechC	AFD	TechD	AHD
Gender	0.260***	0.307***	0.244***	0.277***	0.150**
	0.103	0.046	0.023	0.102	0.098
Noac	0.011	0.021	-0.005	0.094*	-0.001
	0.025	0.017	0.004	0.034	0.060
GEng1	0.037***	0.135**	0.036*	0.115	0.047***
	0.019	0.089	0.029	0.101	0.022
GEng2	0.054	0.019	-0.015	0.246*	0.021
	0.079	0.027	0.016	0.135	0.091
GEng3	0.041	0.395	-0.006	0.059	0.001
	0.087	0.298	0.017	0.071	0.099
GMath1	0.036*	0.101***	0.057***	-0.102	0.036***
	0.025	0.013	0.017	0.009	0.002
GMath2	0.031*	0.224	0.119	0.247***	0.190***
	0.024	0.190	0.199	0.111	0.017
GMath3	0.039***	0.185**	-0.037***	0.116	-0.097
	0.017	0.100	0.019	0.099	0.079
Fsch	-0.033	-0.033	-0.018	-0.081	0.027
	0.043	0.027	0.011	0.143	0.024
Fcel	0.064	-0.243	0.120	0.074	-0.002
	0.049	0.233	0.039	0.141	0.015
Lsa	-0.027	-0.291	-0.017	-0.321**	0.063
	0.036	0.147	0.035	0.141	0.039
Fsc1	0.075	0.094	0.044**	0.015	0.002
	0.057	0.105	0.027	0.016	0.007
Fsc2	0.029	0.048*	0.049	0.039	0.011
	0.021	0.029	0.041	0.049	0.091
Fsc3	-0.046	0.060	0.019	0.031***	0.021
	0.051	0.045	0.037	0.013	0.030
Fsc4	-0.015*	0.049	-0.031*	0.026	-0.029
	0.009	0.051	0.020	0.019	0.022
Pricom	-0.130	0.424***	-0.007	0.763***	0.016
	0.151	0.194	0.064	0.321	0.100

Table 3.3 A Regressing results from log hourly earnings -continued

Cengov	0.146 0.107	-0.159 0.103	-0.278* 0.148	-0.264 0.318	0.285 0.317
La/lea	-0.139 0.221	-0.152 0.146	-0.092 0.091	0.103 0.087	-0.129 0.216
Natindu	-0.101 0.261	1.636 1.887	-0.039*** 0.007	-0.165** 0.075	-0.233 0.175
occu1	-0.185 0.181	0.067 0.119	0.286*** 0.059	0.492** 0.271	0.344*** 0.128
occu 2	-0.051 0.081	-0.133 0.241	0.051 0.046	-0.342 0.324	0.185 0.143
occu 3	-0.146 0.131	-0.342 0.298	-0.072 0.073	-0.523 0.325	0.303** 0.172
occu 4	-0.183 0.345	-0.095 0.102	-0.274** 0.131	-0.432 0.324	-0.195*** 0.007
Adj. R2	0.1684	0.1756	0.2364	0.1934	0.1143
obs	67	106	194	39	13

Note, NCDS wave2 (1963), wave3 (1974) and wave5 (1991) were employed in the regressing process. Wave 2 and wave 3 provide family background and students ability proxies; wave 5 reveals the wages information and highest qualification obtained at age 33.

* indicates significance at the 10% level ** indicates significance at the 5% level ***indicates significance at the 1% level

Table 3.3 B Multinomial logit model selector

	ACoIl	TechC	AFD	TechD
Gender	0.689*	0.053	0.314	0.422
	0.388	0.666	0.378	0.633
Noac	0.228**	0.146	0.157***	0.162*
	0.094	0.107	0.060	0.100
GEng1	0.337	0.262	0.340	0.612*
	0.286	0.350	0.214	0.318
GEng2	0.396	0.261	0.311	0.632
	0.313	0.213	0.343	0.592
GEng3	0.388	0.259	0.327	0.611
	0.356	0.251	0.298	0.509
GMath1	-0.224	0.171	0.127	0.139
	0.240	0.273	0.163	0.261
GMath2	0.221	0.170	0.129	0.137
	0.232	0.158	0.113	0.131
GMath3	0.223	0.173	-0.131	-0.141
	0.191	0.179	0.127	0.119
Fsch	0.030	-0.029	0.043**	0.013
	0.033	0.049	0.021	0.037
Fce	0.267	0.257	-0.049	-0.310
	0.230	0.258	0.161	0.302
Lsc	-0.266	-0.495*	0.423*	-0.402
	0.276	0.283	0.226	0.288
Fsc1	-0.364	0.115	-0.034	0.050
	0.282	0.272	0.156	0.275
Fsc2	0.335	0.125	0.039	0.049
	0.299	0.111	0.059	0.191
Fsc3	0.279	0.153	0.013	0.061
	0.265	0.121	0.009	0.049
Fsc4	0.398	0.171	0.041	0.065
	0.313	0.159	0.121	0.061
cons	-2.748	-2.386	0.043	-1.938
	2.071	2.131	1.368	2.122
Loglihood	-69.791	-124.454	-256.786	-57.341
Pseudo R2	0.1121	0.0998	0.1420	0.1009

Table 3.6 Regression analysis of estimated and realized return

	Realized return		Estimated return		Difference	
	Coeff.	Std.err	Coeff.	Std.err	Coeff.	Std.err
constant	3.247***	0.992	3.257***	0.763	0.106	0.189
gender	0.276***	0.076	0.309***	0.008	0.013	0.014
fsch	0.106	0.004	0.009	0.006	-0.002	0.011
Fsc1	0.027*	0.019	0.028***	0.006	0.017	0.017
Fsc2	0.018	0.017	0.019***	0.009	0.013	0.019
Fsc3	-0.022	0.019	-0.021***	0.011	0.016	0.016
Fsc4	-0.021	0.019	-0.022***	0.013	0.019	0.011
Secondary education						
Noac	0.035	0.021	0.055***	0.006	-0.027***	0.011
Gmath 1	0.231***	0.055	0.491***	0.026	0.057	0.109
Gmath2	0.139***	0.039	0.377***	0.041	0.086	0.191
Gmath3	-0.111	0.123	-0.211***	0.019	0.099	0.101
GEng1	0.111	0.067	0.046***	0.013	-0.003	0.742
GEng2	0.121	0.054	-0.031***	0.009	0.091	0.499
GEng3	-0.101	0.039	-0.029***	0.003	-0.071	0.191
Motivation						
Extrinsic (fce)	0.102	0.031	-0.116	0.103	0.015*	0.040

Table 3.6 Regression analysis of estimated and realized return-continued

Intrinsic (lsa)	0.051**	0.028	0.059***	0.007	-0.009	0.049
Occupation						
Occ1	0.193*	0.105	0.421***	0.027	0.197	0.217
Occ2	0.093**	0.094	0.161	0.172	0.176	0.101
Occ4	-0.048	0.111	-0.085***	0.021	0.009	0.121
Occ5	0.063	0.052	0.047	0.029	-0.031	0.028
Occ6	-0.189	0.241	-0.015	0.027	0.053	0.239
Occ7	-0.288*	0.239	-0.008	0.028	0.313	0.251
Adj-R	0.219		0.589		0.071	
Obs	581		581		581	

Chapter 4 A Particular Uncertainty on Education Choices

-- some evidence from Chinese graduates[17]

Co-written with Prof. David Mayston

4.1. Introduction

In this chapter, we will analyze a particular and also a very important uncertainty on individuals' education decisions—overeducation (i.e. he has more education than is required to do his job). In chapter 3, we considered the uncertainty in individuals' education choices from personal characteristics and macroeconomic risk aspect, separately. But we do not consider the effect of economic growth, economic structure and job characters to individuals' demand for education (i.e. supply and demand condition). However, individuals' education choices are not only affected by their own personal characteristics, but the anticipated capacity of the economy to absorb graduates into productive employment and their peers' qualification level. That is to say how many percents of individuals have the same quality as themselves can find the same level of jobs in the labour market (i.e. whether there is an overeducation problem in the labour market).

Given the magnitude of China's current production of graduates, with a net entry rate to tertiary education at 39% annum in 2003[18], the emerging problem of overeducation merits the attention of economists and policy makers. Notwithstanding China has experienced an impressive rate of economic growth in recent years, its growth rate of 12 per cent in GDP in 2004 is far outstripped by the expansion rate of its higher education system, as we illustrate in table 4.1. Against the background of the current

[17] Thank Prof. David Mayston's contribution on Pecking order model. The data was provided by Changjun Yue from Peking university.
[18] UK is 48% and USA is 63% for comparison (see Chevalier, A., 2003 for detail)

expansion rate of higher education and economic growth rate, the probability of overeducation becomes an essential issue for individuals to make education decisions. In this chapter, we analyse the determinants that may affect an individual's risk of being under- or over-educated.

Though a significant amount of discussion and empirical work were developed on overeducation, there is not a complete theoretic model to explain the reasons that individuals will be overeducated. Therefore, in the first part of this chapter, we fill in this part of blank by developing a pecking order theory model based upon the assignment approach of Sattinger (1993). Our framework used for the analysis is that of a pecking order model in which each individual's characteristics, including their educational achievement, contribute to their relative attractiveness in the job selection process compared to other individuals who they are competing against in the competition for jobs, through an index of their attractiveness in which each of these individual characteristics has a relative weight. The number of jobs available at a given level is determined by demand side factors, such as the level of government expenditure, GDP growth and so on. In the pecking order model, employers choose for the jobs at the highest level of individuals whose individuals characteristics make then the most attractive to hire for the umber of top level jobs employers are offering.

In the second part of this chapter, we analyze the determinants of getting a graduate level of job from a Chinese graduate survey by ordered logit model. And then analyze whether the labour market was separated by gender, industry sector and geographic area through a single index function.

4.2. Pecking Order Analysis —this section and equation (5.9)-(5.14) was developed by Prof. D. J. Mayston

Signalling theory (e.g. Spence, 1973) considers the case where individuals have an incentive to invest, and over-invest, in education when their educational level is used as a signal to convey information to their potential employers about their future productivity, even though the education may not itself directly raise their productivity. Individuals with high ability and/or high productivity will then seek to acquire more education in order to signal this fact to potential employers. However, at the same time, some low ability individuals will also seek to receive more education in order to give a good signal to the employer and acquire high wages. Spence (1973) said, 'Systematic overinvestment in education is a distinct possibility because of the element of arbitrariness in the equilibrium configuration of the market.' He also showed some observable, but unalterable factors (e.g. gender, race, nationality) may drive individuals to be overeducated in order to try to compensate for these unalterable factors.

Spence (1973) explained the reasons that may cause individuals to be overeducated from personal characteristics independently of the availability of jobs. In other words, he assumed the return to education is fixed by the long run supply behaviour of individuals, which cannot explain the empirical observations that the return to high school graduates have declined relative to the return to college graduates. Sattinger (1993), on the other hand, sought to explain the distribution of earnings as arising from the market economy's solution to the problem of assigning workers to jobs. This chapter applies the assignment approach using a pecking order theory to explain the reasons that cause some individuals to face the risk of being overeducated. The pecking order model examines a logical process of sorting individuals into the available jobs that we later seek to estimate empirically using an ordered probit model. Sorting in the employment decision, however, needs to be distinguished from modelling the education process as itself involving a sorting process according to non-educational characteristics of the different individual involved. As noted by Weiss (1995, p. 134) "In sorting models, schooling is correlated with differences among workers that were present before the schooling choices were made; firms make inferences about these productivity differences from schooling choices, and students respond to this inference process by going to school longer" (see also Johnes, 1998). The following pecking order model combines all factors which employers can assess as being related to individuals' productivity, whether they are directly related to such productivity or are more indirectly related to them via other underlying factors.

The demand for graduates is derived demand. It is derived from the demand by employers for individuals whose skills and other characteristics can complement the specifications of the jobs they are seeking to fill. Each job j is assumed to have a set of job specifications at time t given by the vector $Z_{jt}=(Z_{j1t},\ldots, Z_{jmt})$, which specify the work which the job entails. How well that job is carried out depends upon the skills and other characteristics of the individual who occupies the post associated with the job. Each individual i is assumed to possess a vector of characteristics $X_{it}=(X_{i1t},\ldots, X_{int})$ at time t that includes n-1 objective characteristics, such as their educational qualifications, as its first n-1 elements. Its last element is a stochastic term $X_{int} \in u_{it}$ that reflects other less objective characteristics of the individual, such as their enthusiasm, which the employer can assess by less formal means (e.g. interviews), and which also contribute to the individual's ability to make an enhanced contribution.

The value of the output from the job will depend also upon the demand in the product market for the output of goods or services which the job produces. The main drivers of the level of demand at time t for such output across all jobs for whom graduates may be candidates include economy-wide factors $Y_t=(Y_{t1},\ldots, Y_{ts})$, such as the country's level of GDP, its growth rate, its foreign exchange rates with its major trading partners, and its rate of population growth.

We assume that if individual i occupies job j at time t the value of their output is given by the Cobb-Douglas function:

$$V_{ijt} = \omega_t \gamma_j \prod_{\tau=1}^{s} Y_{\tau t}^{c_\tau} \prod_{h=1}^{n} X_{iht}^{a_h} \prod_{k=1}^{m} Z_{jkt}^{b_k} \tag{4.1}$$

where the a_h, b_k and c_τ are positive constants, with $a_h=1$, and the ω_t and γ_j are positive stochastic terms that vary across each t and j respectively according to independent standardised lognormal distributions.

Each employer is assumed to face a wage function of the form:

$$w_{it} = w_{it}(X_{i1t}, \ldots, X_{int}) \tag{4.2}$$

that specifies the wage that must be paid at time t to recruit individual i with characteristics $X_{it}=(X_{i1t},\ldots, X_{iht})$. The employer for job j is assumed to select the individual i who will occupy the post according to the individual's characteristics X_{it} in order to maximise the net value $V_{ijt} - w_{it}$ to the employer of having such an individual perform job j. Since for each $j \in J$, the individual characteristics X_{iht} influence V_{ijt} in (4.1) via the index

$$C_{it} \equiv \prod_{h=1}^{n} X_{iht}^{a_h} \text{ with } \partial V_{jit} / \partial C_{it} > 0 \tag{4.3}$$

employers will evaluate each individual according to their overall value of C_{it}, and are willing to offer a higher wage to individuals whose overall value of C_{it} is greater. For each small increase in C_{it}, the employer for job j would be willing to pay an additional wage premium up to an amount equal to $\partial V_{ijt} / \partial C_{it}$. In a competitive labour market, the wage w_{it} will be bid up to be an increasing function of C_{it}, with

$$(\partial w_{it} / \partial C_{it}) = (\partial V_{ijt} / \partial C_{it}) = v_t J_{jt} \text{ where } v_t \equiv \omega_t \prod_{\tau=1}^{s} Y_{\tau t}^{c_\tau},\ J_{jt} \equiv \gamma_j \prod_{k=1}^{m} Z_{jkt}^{b_k}$$

(4.4)

for the job j which individual i performs. Moreover, since $(\partial^2 V_{ijt} / \partial C_{it} \partial J_{jt}) > 0$ employers with jobs whose specification level is higher according to the index J_{jt} will be willing to offer a greater additional premium to individuals who possess superior characteristics according to the index C_{it} than employers with jobs whose specification level is lower according to the index J_{jt}. In a competitive labour market, employers with jobs whose specification level is higher according to the index J_{jt} will succeed in recruiting individuals with superior characteristics according to the index C_{it}. The top z individuals according to the index of individual characteristics C_{it} are then recruited to the top z jobs according to the index J_{jt} of job specifications for all $0 < z < \Omega_t$, where Ω_t is the total number of jobs in the economy at time t. We will assume for simplicity that the total number of jobs, including those in subsistence agriculture, at each time t, with all individuals assumed to have access to at least a subsistence job. We then have

$$\phi_t(C_{it}) = \varphi_t(J_{jt}(C_{it}))\ for\ all\ C_{it} > 0\ \ and\ hence\ J_{jt}(C_{it}) = \varphi_t^{-1}\phi_t(C_{it}) \tag{4.5}$$

where $J_{jt}(C_{it})$ is the level of specification for the job to which an individual with a level of characteristics C_{it} is recruited, ϕ_t is the distribution function at time t for C_{it} across the population of individuals of working age, and φ_t is the distribution function at time t for J_{jt} across jobs in the economy. If ϕ_t and φ_t are both lognormal distribution functions, (4.5) implies:

$$J_{jt}(C_{it}) = C_{it}^{\sigma_{Jt}/\sigma_{Ct}} A_t\ \ where\ A_t \equiv \exp(\theta_{Jt} - \theta_{Ct}(\sigma_{Jt} / \sigma_{Ct})) \tag{4.6}$$

where θ_{Jt} and θ_{Ct} are the mean values of $\ln J_{jt}$ and $\ln C_{it}$, and σ_{Jt} and σ_{Ct} are their respective standard deviations, at time t.

Equation (4.3), (4.4) and (4.6) imply a wage function of the form:

$$w_{it}(X_{it}) = B_t \prod_{h=1}^{n} X_{iht}^{\beta_{ht}} + \zeta \quad where\ \beta_{ht} \equiv a_h \varpi_t,\ \varpi_t \equiv 1 + (\sigma_{Jt} / \sigma_{Ct}),\ B_t \equiv v_t A_t / \varpi_t \tag{4.7}$$

where ζ is a constant of integration, which equals zero if the reservation wage given by the employment benefit rate is zero and subsistence jobs in agriculture are available to all (see Sattinger, 1993). With unemployment benefits in China less than one per cent of the wage level of many new graduates, we will assume that these conditions hold, and hence set $\zeta = 0$.

In a pecking order model, individual characteristics $X_{it}=(X_{i1t},\ldots, X_{int})$ play an important part in wage determination, as in equation (4.7). However, so too do the parameters, θ_{Jt} and σ_{Jt}, of the distribution of job characteristics, alongside the parameters, θ_{Ct} and σ_{Ct}, of the overall distribution of individual characteristics in the population at large, as in equations (4.6) and (4.7). In previous chapters, we discussed the uncertainty of θ_{Jit} and σ_{Jt} through macro economic variables and θ_{Ct} through individuals' characteristics, whereas the uncertainty of σ_{Ct} is not mentioned. One direct factor affecting σ_{Ct} is the extent of overeducation, which will be analyzed in the next section in detail.

Pecking order theory also has an important role to play in the analysis of the incidence of overeducation and undereducation across different individuals. In analysing the incidence of overeducation and undereducation of graduates with different levels of degree qualifications, we will assume that jobs can be categorised into one of five levels. Level $l = 4$ corresponds to those requiring PhD-level skills, $l = 3$ to those requiring Masters-level skills, $l = 2$ to those requiring undergraduate degree-level skills, $l = 1$ to those requiring college-level skills, and $l = 0$ to those requiring none of these skills. Higher level jobs at time t will be assumed to involve higher levels of specification according to the index J_{jt}. J_t^{ol} will denote the minimum level of the job specifications index J_{jt} for which skills of level l are required, and below which only a lower level of skills is required.

Since from (4.4), (4.5) and (4.7) for any job j for which individual i is selected:

$$(dC_{it} / dJ_{jt}) = (\partial^2 V_{ijt} / \partial C_{it} \partial J_{jt}) / (\partial^2 w_{it} / \partial C_{it}^2) > 0 \qquad (4.8)$$

a pecking-order process will apply to the selection of individuals according to their overall quality given by the linear function:

$$q_{it} = \ln C_{it} = x_{it} a + \varepsilon_i \qquad (4.9)$$

where $x_{it} \equiv (x_{i1t}, ..., x_{i(n-1)t}) \equiv (\ln X_{i1t}, ..., \ln X_{i(n-1)t})$, $a \equiv (a_1, ..., a_{n-1})$, $\varepsilon_i \equiv \ln \mu_i$. Individuals with a higher quality according to their q_{it} rating in the pecking order of individuals in the population will attain a higher value to their C_{it} index of individual characteristics that enhance their employability, and hence secure a higher specification job according to the index J_{jt} in (4.8). In particular, in order to secure a job of level $l > 0$ or above, individual i must have a quality level

$$q_{it} > q_t^{ol} \equiv (\sigma_{Ct} / \sigma_{Jt})(\ln J_t^{ol} - \theta_{Jt}) + \theta_{Ct} \qquad (4.10)$$

Employers thus set a 'hurdle' level q_t^{ol} at time t for the minimum quality q_{it} of individual i to whom an offer is made of a level $l > 0$ job. This hurdle level, moreover, depends upon the parameters of the distribution of job specifications in the economy and the distribution of individual characteristics in the population, both of which may change over time.

In the probit model form of the pecking order analysis, we will assume that ε_i, is a stochastic variable that is normally distributed with a mean of zero and a standard deviation of one across the population, independently of x_{it}. The probability of any individual i with objective characteristics given by x_{it} being offered employment at time t at level $l > 0$ or above is then given by:

$$p_{\ell t}(x_{it}) \equiv \Pr(q_{it} > q_t^{o\ell} | x_{it}) = \Pr(x_{it}a + \varepsilon_i > q_t^{o\ell}) \qquad (4.11)$$

$$= 1 - \Pr(\varepsilon_i \leq -x_{it}a + q_t^{o\ell}) = 1 - N(-x_{it}a + q_t^{o\ell}) = N(x_{it}a - q_t^{o\ell})$$

where N is the cumulative standardised normal distribution function.

From (4.5), (4.6) and (4.10), the total supply of individuals who satisfy the minimum quality hurdle q_t^{ol} is given by:

$$S_t(q_t^{o\ell}) = \eta_t(1 - N((q_t^{o\ell} - \theta_{Ct}) / \sigma_{Ct})) \ \ with \ \partial S_t / \partial q_t^{o\ell} < 0 \qquad (4.12)$$

From (4.5), the total demand D_t^{ol}, by employers at time t for individuals to fill jobs of level $l > 0$ or above is given by:

$$D_{lt} = \eta_t(1 - \varphi_t(J_t^{o\ell})) \qquad (4.13)$$

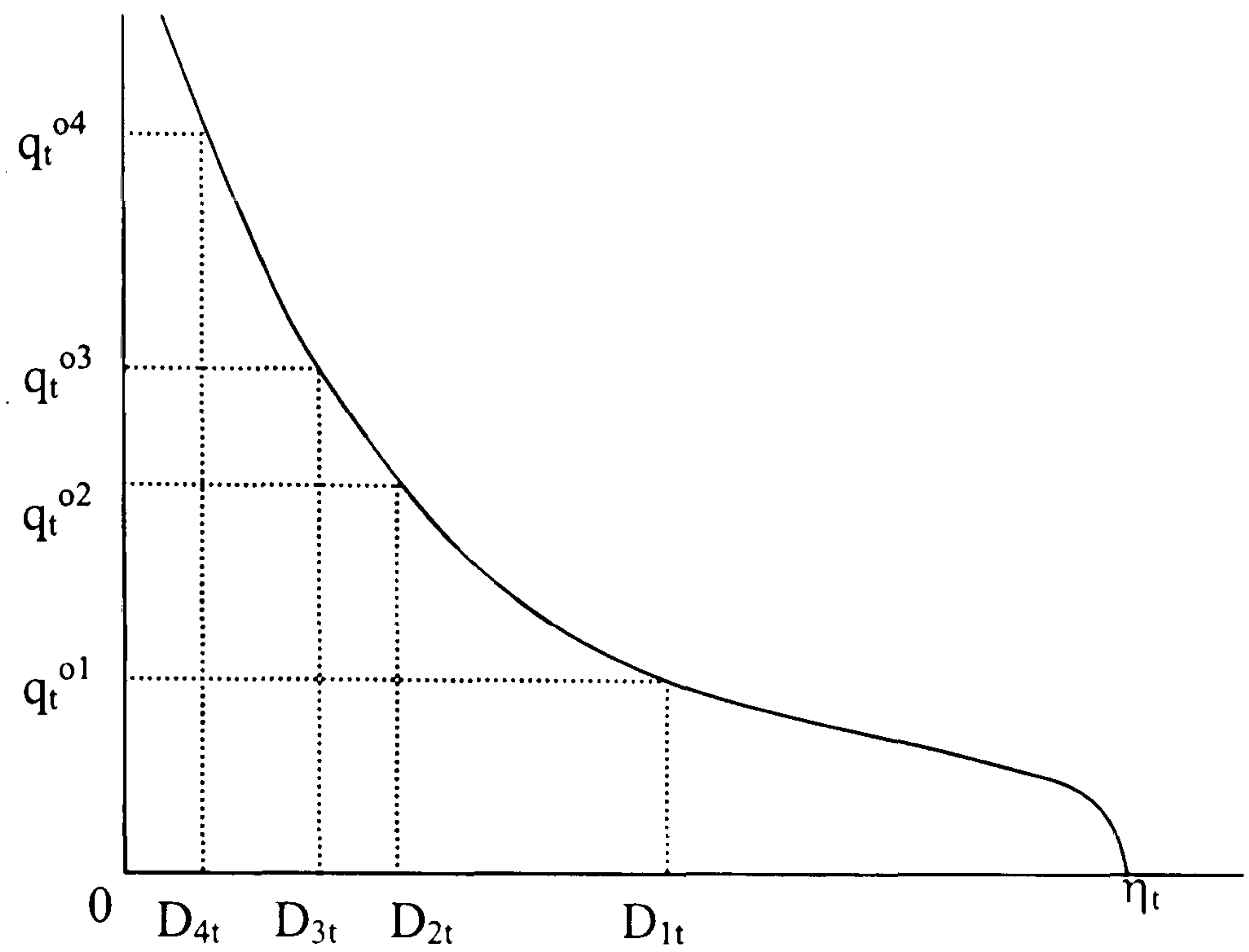
with the minimum quality hurdle q_t^{ol} set in (4.10) to equate the supply, $S_t(q_t^{ol})$ to demand D_{lt}, at each level $l > 0$ and above at each time t. Such equality implies also that

$$S_{lt}^{"} \equiv S_t(q_t^{ol}) - S_t(q_t^{ol+1}) = D_{lt} - D_{l+1,t} \equiv D_{lt}^{"} \text{ for } l = 1, 2, 3 \text{ and}$$

$$S_{4t}^{"} \equiv S_t(q_t^{o4}) = D_{4t} \equiv D_{4t}^{*} \qquad (4.14)$$

i.e. an equality between the number of individuals S_{lt}^{*}, who are available within each quality range and the number of jobs employers are seeking to fill at each level of employment.

Figure 4.1 The critical job level in pecking order theory



The analysis of the incidence of overeducation by pecking order theory can be examined also through Figure 4.1. On the horizontal axis is the level of demand D_{lt}, which exists in the economy at time t for employers to fill vacancies of level $l > 0$ and above, for each job level $\ell > 1,...,4$. The curve in Figure 4.1 is the supply of labour curve corresponding to equation (4.12) above, that shows on the horizontal axis the quantity of labour that will be available in the labour market of quality q_t^{ol} or above at time t, for each such quality level. The curve is downward-sloping because the quantity of labour that is available declines with a rise in its minimum quality level. The quality levels $q_t^{o1},......,q_t^{o4}$ are the maximum quality hurdle levels which can be set in the labour market by employers for all graduates to whom job offers are made, whilst still ensuring that the supply of graduates who meet these quality levels is sufficient to satisfy the corresponding levels of demand $D_{1t},......,D_{4t}$ by employers to fill their vacancies.

Overeducation can arise in the above pecking order analysis from the possibility that an individual will have graduated with a qualification of level $l > 0$ (which we will denote by $\delta_{itl} = 1$), but still have an overall level of quality q_{it} that falls short of the minimum hurdle level q_t^{ol}, for being offered a job of level ℓ. The associated probability of overeducation is given by:

$$p_{\ell t}^{o}(x_{it}^{\ell}) \equiv \Pr(q_{it} < q_t^{o\ell} \,|x_{it}\, \&\, \delta_{it\ell} = 1) = \Pr(x_{it}^{\ell} a + \varepsilon_i < q_t^{o\ell}) = N(q_t^{o\ell} - x_{it}^{\ell} a) \tag{4.15}$$

where x_{it}^{ℓ} is a vector of individual characteristics x_{it} that includes the individual having graduated with a qualification of level $\ell > 0$. Similarly undereducation can arise in the above pecking order analysis if an individual has an overall level of quality q_{it}, that exceeds the minimum hurdle level, $q_t^{o\ell}$, for being offered a job of level ℓ, even though they have not graduated with a qualification of level $\ell > 0$ (which we will denote by $\delta_{it\ell} = 0$). The associated probability of undereducation is given by:

$$p_{\ell t}^{u}(x_{it}^{o\ell}) \equiv \Pr(q_{it} > q_t^{o\ell} \,|x_{it}\, \&\, \delta_{it\ell} = 0) = \Pr(x_{it}^{o\ell} a + \varepsilon_i > q_t^{o\ell}) = N(x_{it}^{o\ell} a - q_t^{o\ell}) \tag{4.16}$$

where $x_{it}^{o\ell}$ is a vector of individual characteristics x_{it} that includes the individual having graduated without a qualification of level ℓ.

The hurdle level q^{ol} in equation (4.10) also depends on the standard deviation of job specifications σ_{Jt}, that itself depends upon the extent and nature of the variety of jobs on offer in the economy, and upon the standard deviation of individuals' characteristics σ_{Ct}, as dependent upon the distribution of levels of qualifications and other attributes amongst potential employees.

An alternative distribution function to the cumulative standardised normal distribution function N in equations (4.11) – (4.16) is provided by the logistic function:

$$G(x_{it}a) = \exp(x_{it}a)/(1+\exp(x_{it}a)) \qquad (4.17)$$

where the corresponding logistic, or Fisk (see Cramer, 2001, p.15), density function is:

$$g(x_{it}a) = \exp(x_{it}a)/(1+\exp(x_{it}a))^2 \qquad (4.18)$$

with a mean of zero. When a is rescaled so that (4.18) has a variance of unity, the logistic and normal density functions, and their corresponding distribution functions, lie close together (see Cramer, 2001, p.16; Davidson and McKinnon, 1993, p. 516). The logit distribution can be shown to result from a similar individual quality index as (4.3), but with the log of the latent variable $\varepsilon_i \equiv \ln \mu_i$ distributed according to the type I extreme value distribution function in standard form (see Cramer (2001, p. 51):

$$H(\ln \mu_i) = \exp(-\exp(-\ln \mu_i)) \qquad (4.19)$$

4.3. The Incidence of Overeducation and Undereducation

4.3.1 Data description

This chapter employs the survey data conducted by a project[19] in June 2003, just before students' graduating in order to control the response rate.

[19] The project was conducted by Peking University under the leader of Weifang Min. The data was provided by the associate prof. of Changjun Yue. The translated copy of the survey was attached in the appendix.

The survey was designed to use sample selection method and chose three provinces in each representative economic development area (the East, the Middle and the West). In each province, 6 higher education institutions were chosen, where there are 2 elite universities, 2 common universities and 2 polytechnic colleges. However due to the inequality of universities in each province and other practical problems, the investigated provinces and higher education institutions are as follows: Beijing (5), Shangdong (6), Guangdong(6), Hunan(6), Shannxi (4), Yunnan (17), Guangxi (1). Altogether, 7 provinces and 45 universities took part in the survey and the total sample number is 18722.[20] Among these samples, 39.3 percent individuals acquired college or equivalent qualifications and graduates who obtained bachelor, master, doctorate occupies 57.1, 3.0 and 0.6 percent respectively. The proportions of males for these four levels of degrees are 52.7, 65.4, 59.2 and 73.9 percent individually. The gender ratio is almost balanced for each level of qualification, except doctoral level.

As the investigation took place just before individuals' graduation, only 40.7 percent respondents have found a job, 4 percent individuals plan to be self-employed, 15.1 percent candidates will continue study, 20 percent have other plans and 20.2 percent have not found a job. As the survey took place just before the graduation, the efficient sample of those who got offers may have selection bias. Heckman two-step selection method was employed in this chapter to control the selection bias.

The most important variable overeducation in this chapter was designed by two questions. In the questionnaire students were asked, "What is your current qualification? (Four possible selections are given from college graduate to doctorate) and what is the minimum formal qualification required in your contracted job?" (Six-point scale from junior school to PhD can be chosen by each respondent). Through matching the two

[20] Related questionnaire was attached in the Appendix A.

groups of answers, we get the statistics results of overeducation, which listed in table 4.2.

Table 4.2 shows that there are around 20 percent graduates are believed they are overeducated in China's labour market, with corresponding percentages for college graduate, bachelor, master and doctor of 12.9, 21, 36 and 42 per cent respectively[21]. Thus, overeducation in China is more frequent among higher degree than among lower degree. This is consistent with Groot (1996), which shows the incidence of overeducation increases with the years of education required for the job, but in contrast with the studies of Frenette (2004) in Canadian labour markets graduates, which shows master degrees are the most likely to be overeducated and followed by college graduates. The overeducation studies[22] on the role of gender are mixed, with findings that vary across countries and survey data. In our data, the ratio of males to females for the bachelor's graduate is 1.89:1, but the gender ratio for the overeducated graduates in the same qualification becomes 2.1:1, implying that there are more men than women are overeducated in the undergraduate level. But the role of gender to overeducation needs to be examined in the later section.

The evidence on the effect of field of study on overeducation differs significantly from the findings in developed countries, where there is a high variation in the rates across fields[23]. The distribution of overeducated (undereducated) graduates in China among major of subjects is almost equal, except for the case of agriculture, which is the easiest subject to be either overeducated or undereducated. Besides agriculture, graduates majoring in economics are more likely to be overeducated and law graduates have the highest probability of being undereducated according to the statistics.

[21] Our sample size for doctors isvery small, which may bias our estimation.

[22] Duncan and Hoffman (1981) shows there is no gender difference between overeducation; Rumberger (1981) explores women are more likely to be overeducated and Groot (1996), Sloane et at(1999) demonstrate men are more frequently to be overeducated.

[23] The evidence from western countries shows that graduates from science, engineering, medicine and law are less likely to be overeducated than other graduates. See appendix B for detail.

4.3.2 The determinants of securing a higher level job

As our dataset is a cross-section data, we cannot observe the duration of overeducation and whether the overeducation will decline with an increasing length of working experience. However, our data records each individual's comprehensive information on personal characteristics, academic achievement and employment information that allows us to examine what kind of factors may affect the level of job in which an individual finds employment, and their associated probability of over or undereducation.

One approach to incorporating the level of an individual's qualification is through assigning 1 to indicate a college diploma or equivalent, 2 to indicate a Bachelor's degree, 3 to indicate a Master's degree and 4 to indicate a PhD. However, this introduces too simple a metric measure into the comparison of the effects of different levels of individual qualification, compared to the use of dummies for each level. Similar remarks apply to the variables corresponding to parents' education, parents' career, university ranking, qualifications in English, class of degree, and location of employment. The use of dummies can also tell which qualification levels have a significant effect on getting a job level l. The results of log likelihood ratio test and associated chi-squared distribution were compared for these two kinds of variables measurement: the first involving adding dummies and the second involving a pre-specified rank order, such as 1, 2, 3, 4. As expected, we found that the use of dummies improves the goodness of fit, and so have listed our empirical evidence based upon using them.

In line with the pecking order analysis above, we also wish to allow the probability of over or undereducation to vary with the level of job involved, as well as with individual characteristics, in a way that depends upon the underlying supply and demand parameters in the labour market. An estimation model that ties in directly with the pecking order analysis and with these requirements is the ordered probit model. This yields threshold hurdle levels of the single quality index for individual characteristics that vary with each job level in a way that can in turn depend upon the level of demand by employers in the economy to fill vacancies at each such job level. As we shall discuss in more detail in chapter 5, another advantage of the ordered probit model is that it can be tied in well with an analysis of individual choice under uncertainty for investment in education, by looking at the impact of additional education on the probability of achieving a job offer at each level.

The ordered probit model enables us to cut the range of the individual quality index at four threshold hurdle points, in line with Figure 4.1 above, with level 4 again corresponding to jobs requiring a doctorate, level 3 to a job requiring a master's degree, level 2 to a job requiring a bachelor's degree, and level 1 to a job requiring a college diploma. The precise list of variables for individual characteristics on which data have been collected and which will be included are listed in table 4.3.

We will initially include these as dummy variables alongside the vector of individual characteristics as determinants of the threshold hurdle levels for securing a job at each job level. This has the effect of including shift parameters for each geographical location and industrial sector alongside the same index for individual characteristics. An alternative approach we shall investigate later is to estimate potentially different single indices of individual characteristics for different geographical location and industrial sector within China.

Since our data sample involves candidates who have already been offered and signed an employment contract, there may exist sample selection bias due to the respondents in our sample potentially having a biased distribution of the latent variable μ_i compared to the wider population of graduates. We therefore include the associated inverse Mills ratio based upon the two-step Heckman estimation procedure (Heckman, 1979) to correct for such potential sample selection bias. The first step probit selection model was listed after table 4.4.

Also of interest is whether or not there is an equal return on additional years of education, in terms of raising the probability of securing a job of a given level, throughout the range from college diploma to PhD. In order to test whether this return is constant or varies across this range, we form a nested model in which the variable lnschooling (abbr. *lnsch*) is defined as the logarithm of the total number of years spent in higher education years minus the logarithm of the number of years required for a college diploma. This can be shown to be a special case, and associated null hypothesis, where equal weights are placed upon the variables of lnbacheloryears (abbr. *lnba*) , lnmasteryears (abbr. *lnma*) and lnphdyears (abbr. *lnphd*), so that the sum of the coefficient of *lnba* plus *lnma* plus *lnphd* equals the coefficient of *lnsch*, where *lnba*, *lnma* and *lnphd* are defined as follows: the logarithm of the total years to acquire a bachelor degree in higher education period over the total years to acquire a college diploma (i.e. *lnba*); the logarithm of the total years to acquire a master degree in higher education period over the total years to acquire a bachelor degree (i.e. *lnma*) and the logarithm of the total years to acquire a PhD in higher education period over the total years to acquire a master degree (i.e. *lnphd*).

Many of the variables have the same significance level in both models. However, gender, university rank, university grade, and the inverse Mills ratio are all significant in the *lnsch* model, but not in the model that allows different coefficients on the additional years' variables at each level of higher education. The significance of inverse Mills ratio indicates selection bias is more serious in the former model. For the *lnsch* model, the sign of gender is negative and significant, which means that females score significantly less highly than males in the pecking order for higher level jobs in this version of the model. In addition, being in a top 100 university in China has a very significant effect on an individual's assessed quality index in the *lnsch* model. This is in contrast to the findings in Battu et al (1999) that graduating from a less selective university plays no part in explaining overeducation in UK. In the *lnsch* model, obtaining a lower class of degree, as indicated by grade 2 and 3, has a significant adverse effect on an individual's position in the pecking order for obtaining a higher level job.

However, once the coefficients on additional years' education at each separate qualification level are allowed to vary, gender, university rank and grade, and the inverse Mills ratio are no longer significant. Whilst the coefficient on the combined *lnsch* variable is highly significant for both genders together and separately. Disaggregation of the schooling variable reveals that it is the coefficients on additional years at the masters and PhD levels, particularly for males, which have the most significant effect on boosting an individual's position in the pecking order.

We may use the log likelihood ratio (LR) test to test whether or not the model with different possible coefficients on the variables for each level of education does offer a significant improvement in fitting over the model using the *lnsch* variable (c.f. Greene, 2000 pp.152). For the combined gender model, the LR test yields the test statistic of -2 (lnR - lnU) = -2(-5445.9856+5437.9439) =16.083. The associated chi-squared statistic with 3 degrees of freedom is significant at any conventional level, so that we would reject the null hypothesis that the model using *lnsch* does not offer a significantly better fit than one involving a different coefficient on additional years at each level of graduate education. That is to say the return to an additional year of schooling is different in terms of increasing the probability to obtain a qualification matched job for different level of qualification. However, the result does not apply to the male and female models separately, in other words, we cannot reject the null hypothesis at 5 percent significant level for the models of male and female separately. The diverse results of integrated sample to the separate sample for men and women only may due to the sample size limitation. If we separate qualification level as well as gender, the sample size for some cells may be smaller than 30, which may influence the correctness of estimation.

The variable *Pmem* associated with being a Communist Party member stands out from the personal characteristics' variables as being very significant in both models in increasing an individual's position in the pecking order for securing a higher level job. The roles of the variables related to parental career and parental qualifications are also very interesting. As one might expect, parents from the lowest social background are less able to contribute any information, wealth or other help for their children in finding a higher level job. Parents belonging to the lowest social rank (i.e. peasants or unemployed) are here found to have a significant negative impact on their children's position in the pecking order, other things being equal. However, so too are parents with qualifications beyond the master's level. One possible explanation of this interesting negative effect is that well-educated parents have a strong desire to let their children have a very good education. This may therefore cause their children to become overeducated compared to their underlying intelligence and abilities which are important factors in selecting individuals for job offers. A positive general coefficient on additional years in higher education may then be partially or wholly offset by a negative coefficient for children of those coming from families with high levels of education. In addition, the variable of parental career and parental qualifications may have the problem of multicollinearity, which in part may result some bias.

The determinant factors may also depend on registration status and whether be a cadre. Respondents coming from countryside and not a cadre are less likely to find a matched job. Cadre in China is quite similar to a student representative or the student union president in UK, who generally organizes some art or sports activities to entertain the students' life. Through organizing some activities or unions, they may develop their leadership ability, which is appreciated by the recruiting companies. The candidates from countryside may be in a disadvantaged position owing to a lack of outlooking and presentation skills, and which become more and more important in interview process.

English skills are particularly important in job hiring decisions in the separate qualification level model, though not for female graduates. The counterintuitive effect of grad2 and grad3 to the probability to get a graduate level of job is because the employers in China pay little attention to the average mark in the university, since 50 percent courses (such as Marxism, Deng's theory, mental healthy, behaviour of a good university student) have nothing to do with individuals' specialties and abilities. As long as you work hard and can memorize all the necessary knowledge points, you can acquire a high score. Generally girls work harder than boys, but in a disadvantage position in finding a good job resulting the grade may have a counter intuitive effect. The effects of the degree subject of study on an individual's chances of a higher level job are also very interesting. In Western countries, students graduating in Physics, biology or mathematics are more likely to find a graduate a job, due to a relative scarcity of such graduates[24]. However, in China, students have less freedom of choices in selecting their subjects of study, with the numbers of graduates from all subjects more evenly balanced due to the centrally planned nature of their provision. In addition, employers may view the courses in these three subjects as too theoretical for their needs, with negative estimated coefficients on these subjects. In contrast, whilst graduating in the subject of Language tends to increase the probability of being overeducated in US, UK and Canada (Frenette, 2004), it contributes positively to find a graduate job in China, especially for females. In Western countries, graduates in arts subjects, such as languages, tend to have lower mathematical abilities than graduates in science subjects. However, in China mathematics is a compulsory subject in the entrance examination for all degree courses, so that this differential effect in favour of scientific subjects is less. The coefficients on law and medicine in

[24] See Dolton and Silles(2001), and Battu et al(1999) for detail

increasing the probability of securing a higher level job are positive, though only significantly so for medicine and for both genders combined.

As well as individual characteristics, a number of characteristics of the employment itself have been included to assess whether or not these influence the probability of securing a higher level job. These include the geographical location of the employment, distinguished by whether it is in cities, towns, villages, in the middle, east, west or elsewhere in China. Work location is found to be very significant in influencing the chances of securing a higher level job, especially for the cities in the Workcity3 group which have a very significant positive effect upon the chances of securing a higher level job. This is in accord with our expectation that large cities generally offer competitive jobs and salaries. The influence of the sector of the economy in which the job is located was also examined, with employment in an educational institution significantly raising the probability of the job being at a higher level.

4.3.3 The ordered logit model

We can compare the goodness of fit of the *lnsch* ordered probit and the *lnsch* ordered logit models (see Davidson & Mackinnon, 2003, p452). The associated LR test yields a test statistic of: $-2(\ln\lambda_R - \ln\lambda_U) = -2(-5380.3222+5445.9856) = 131.3268$. The corresponding chi-squared statistic with 1 degree of freedom is significant at any conventional level, so that the ordered logit model is found to give a better fit here than the ordered probit model.

We can also compare the goodness of fit of the two versions of the ordered logit model, the first with the *lnsch* variable and the second with the more general formulation that permits different coefficients on years of education at different levels of higher education. The associated LR test yields a test statistics of : $-2(\ln\lambda R - \ln\lambda U) = -2(-5380.3222 + 5371.2414) = 18.1616$. The corresponding chi-squared statistic with 3 degree of freedom is significant at any conventional level, so that the more general formulation is found to give a significantly better fit than the special case using only the *lnsch* variable.

In addition, we can compare the ordered probit and ordered logit models using the separate *lnba* etc variables, as in the second set of columns in Tables 4.3 and 4.4 respectively. The associated LR test yields a test statistic of : -2 ($\ln\lambda_R$ - $\ln\lambda_U$) = -2 (-5371.2414 + 5437.9439) = 133.405. The corresponding chi-squared statistic with 1 degree of freedom is significant at any conventional level, so that the ordered logit model is again found to give a better fit here than the ordered probit model.

As can be seen from Tables 4.3 and 4.4, education is clearly amongst columns the most important variable in determining an individual's chance of securing a higher level job. The different coefficients on the influence for the three separate qualification levels also raise the question of how this model compares with inserting an increasing trend for the variable *lnsch*. Table 4.5 shows the results for both *lnsch* and separate qualification (e.g. *lnba* etc.) of the ordered logit model from inserting a quadratic version of the *lnschooling* variable. The highly significant positive coefficient on the square of the *lnsch* variable suggests that additional years of education have an increasing, rather than constant, return in terms of their relative impact on the probability of securing a higher level job.

Then we compare the goodness of fit of table 4.6 with table 4.5 within the more general model with the separate education variables plus the square of *lnsch*. First we compare the first column of table 4.6 with the first column of table 4.5 under the restriction of the coefficient of $(\text{lnsch})^2$ equals to 0 in table 4.5. The statistics of LR test is -2(lnR-lnU)=-2(-5380.322+5371.3059)=18.032. The corresponding chi-squared statistic with 1 degree of freedom is significant at any conventional level, so that the improvement from inserting the additional squared term is highly significant. The *lnsch* model for the male only satisfies the same attribute as the combined group. While the statistics of LR test for female model is not significant. A quadratic lnsch model is not significantly better for female group.

The goodness of fit is quite slim no matter when we compare the results within table 4.6 or compare the last three column of table 4.6 with the last three column of table 4.5. In other words, inserting additional squared term does not improve the fitness of the separate education level model (i.e. using *lnba*, etc) no matter for combined group or male and female separately. The ordered logit model using the years of education disaggregated by level of higher education in Table 4.5 therefore provides the best fit when compared to the above alternatives

4.3.4 Gender and sectoral analysis

Since a main focus for the pecking order model is a single index of individual characteristics, we can first investigate whether a significantly better fit can be obtained by using different indices of their remaining individual characteristics for male and for female graduates.

We examined the log likelihood of both the ordered probit model and the ordered logit model for both *lnsch* case and separate *lnba* etc. variables case of how the combined gender model compares with the male and female split. According to the results, we could reject the hypotheses that male and female share the same coefficients under any circumstances. Take ordered logit model for example: the log likelihood ratio statistics are 125.2 and 117.59 for *lnsch* case and separate *lnba* etc. variables case, respectively, the chi-squared statistic with 47 and 49 degrees of freedom separately are significant at any conventional level.

It is also of interest to examine whether job characteristics relating to geographical location and industrial sector are best modelled as simply having an additive effect alongside a common index of individual characteristics for each gender in determining the chances of securing a higher level job offer, as in Tables 4.4-4.6, or whether they are better modelled as requiring a different index of individual characteristics for each gender for each different geographical location and each different industrial sector, with different possible relative weightings placed on these different individual characteristics in each case. Since ordered logit model gives a better fit, we only list the results from ordered logit regressions.

Disaggregating the ordered logit analysis into different geographical areas in Table 4.7 provides an unrestricted model of which the model in Table 4.9 with the same coefficients across all geographical areas is a restricted nested version. Due to the sample limitation, we use the logarithm level of dummy variable for each individual characteristics as in Table 4.9 instead of each dummy variable for each level as in table 4.4 and table 4.5 to guarantee the numbers in each cell larger than 10. We can therefore apply the corresponding LR test with a test statistic for males of: -2 (lnR - lnU) = -2 (-3591.9563 - (-2221.661 - 345.874 -967.513) = 113.816. Since the restrictions for the restricted model are the coefficients of all the variables for each geographic area are the same that makes 3 times 30 variables (all together 90) restrictions all together. This chi-squared statistic of 113.816 with 90 degrees of freedom is not significant at 5% significant level, which means we cannot reject the single index function version of Table 4.9 for males.

Similarly for females, the corresponding test statistic is: -2 (lnR - lnU) = -2 (-1774.9539 - (-1080.116 - 128.690 - 516.765) = 98.766. This chi-squared statistic with 90 degrees of freedom is not significant at any conventional level, which means we cannot reject the single index function version of Table 20 for females. The results of disaggregating the ordered logit model by industrial sector are shown in Table 4.8.

Disaggregating the ordered logit analysis into different industrial sectors in Table 4.8 provides an unrestricted model of which the model in Table

4.9 with the same coefficients across all industrial sectors is a restricted nested version. We can therefore apply the corresponding LR test, with a test statistic for males of: -2 (lnR - lnU) = -2 (-5419.527 - (-673.879 - 1754.061 - 429.750 - 1479.321 - 989.859)) = 185.314. This Chi-squared statistic with 145 degrees of freedom is not significant at any conventional level, so that we cannot reject a common index on individual characteristics for all industrial sectors. This means though the important of different factors is different in each geographic area and in each industry sector, they generally satisfy the same rule. Labour market in China is not segmented by geographic area and industry sector in terms of overeducation.

4.4. Conclusion

The results reported in a large postgraduate survey suggest there are around 20 percent graduates are overeducated among all the fresh graduates from higher education and the percentage is increasing with qualification. The mismatches across genders are significantly different from the results by Frenette (2004) and Dolton & Vignoles (2000), with men are more easily to be overeducated. The distribution of overeducated graduates is almost equal in all the subjects, except agriculture.

The most important facts that may prompt individuals into a high pecking order position are Party membership, university rank, English skills and be a cadre. That is to say, though the rapid expansion of higher education in the recent years, graduates from top ranked universities with a good English skills are less likely to find a mismatched job. Similar to the findings by Frenette (2004) that graduates from computer science, electronics have an advantage in finding a fit job, but so does the field of language, which contributes negatively in Frenette (2004). Parents' background plays a very interesting role in getting a higher level of job that Parents belonging to the lowest social rank (i.e. peasants or unemployed) are here found to have a significant negative impact on their children's position in the pecking order, other things being equal. However, so too are parents with qualifications beyond the master's level.

Gender, geographic area and industry sector play a role in explaining overeduation as well. Female graduates who are working in the east part of China and state-owned company are easily to be overeducated. However, the labour market in China is not separated by geographic areas and industry sectors, which examined by LR test in the final part of this chapter.

This chapter also discusses the goodness of fits of several estimating methods on securing a high level of job. The results show the logit model gives a better fit than probit model and inserting a quadratic item of schooling does not improve the goodness of fit. In addition, there is not an equal rate of return on additional years of education, in terms of raising the probability of securing a job of a given level, therefore using separate *lnba* etc. instead of *lnsch* can improve the estimation results.

Table 4.1 The growth rate of GDP and higher education in China

Year	Total number of undergraduates (Units: thousand)	Total number of post graduates (Units: thousand)	New entrants of undergraduates (Units: thousand)	New entrants of postgraduates (Units: thousand)	Percentage increase of new entrants in higher education (%)	GDP (Units: Yuan billion) based on the price of 1989	Percentage annual increase of GDP (%)
1990	1970	93	579	30	-8.8*	1799.02	-8.93
1995	1956	145	875	51	11.5*	3093.00	14.4
1996	2858	163	907	59	4.3	3315.35	7.19
1997	2998	176	1064	64	3.5	3537.56	6.70
1998	3211	199	1011	73	8.4	3752.03	6.06
1999	3901	234	1505	92	47.3	3986.10	6.24
2000	5260	301	2078	128	38.1	4328.24	8.58
2001	6797	393	2518	165	21.6	4675.12	8.01
2002	8533	501	3002	203	19.5	5093.35	8.95
2003	10435	651	3553	269	19.3	5611.02	10.16
2004	13335	820	4473	326	16.7	6304.74	12.36
2005	15618	979	5044	365	12.7	7219.88	14.52

Source: China's education statistics bulletin and China's labour statistics bulletin

Table 4.2 Incidence of mismatch across education and subjects (%)

Education level	combined			Males			Females		
	under	acquired	Over	under	acquired	Over	under	acquired	Over
diploma[25]	41.1	46.0	12.9	21.8	23.8	7.1	19.3	22.2	5.8
Bachelor	12.4	66.5	21.1	8.7	42.4	14.3	3.7	24.1	6.8
Master	7.3	57.0	35.8	3.8	33.5	21.8	3.5	23.5	14.0
PhD	---	58.0	42.0	---	43.5	30.4	---	14.5	11.6
Total	17.4	62.1	20.5	10.8	38.5	13.5	6.6	23.6	7.0
Econ	18.9	58.9	22.2	10.3	34.5	11.6	8.6	24.4	10.6
Law	22.0	58.3	19.6	12.9	27.9	11.0	9.1	30.4	8.6
Art	18.6	64.1	17.2	6.9	25.0	9.0	11.7	39.2	8.1
Medicine	17.2	65.5	17.2	7.6	31.7	5.5	9.6	33.8	11.7
Science	12.2	69.6	18.3	8.5	37.8	12.8	3.7	31.8	5.5
Engineering	16.8	64.7	18.5	13.1	50.0	14.9	3.7	14.7	3.7
agriculture	8,3	63.5	28.1	7.3	43.8	21.9	1.0	19.7	6.2

[25] Also includes graduates from polytechnic college

Table 4.3 Variable definition

Variable	Description
Gender	female
cadre	be a student representative in the class or in the university
pmem	be a member of Communist Party
ethic	belongs to a Minority
qualification-- base case college	
lnsch	log or total years of higher education over college study
lnba	log of 4 years' bachelor study over 2 years college study
lnma	log of 7 years' master study over 4 years bachelor study
lnphd	log of 10 years' PhD study over 7 years master study
father's career-- base case unemployment, retired or peasant	
pcar1	father is a worker, retailer or a shop assistant
pcar2	father is a professional technician or clerk
pcar3	father is a manager, officer or government official
father's qualification-- base case below junior school	
pqu1	father's highest qualification is Junior school graduate
pqu2	father's highest qualification is senior school graduate
pqu3	father's highest qualification is college graduate
pqu4	father's highest qualification is bachelor
pqu5	father's highest qualification is master or above
registration status-- base case village	
reg1	be born in a small town
reg2	be born in a small city
reg3	be born in a large city
university rank-- base case polytechnic college	
rank1	all the other university
rank2	top 100 university
English qualification-- base case no English qualificaiton	
Eng1	acquired the qualification of CET-4
Eng2	acquired the qualification of CET-6

Table 4.3 variable definition-Continue

Class of degree-- base case below 75%	
grad1	class of degree is from top 50% to top 75%
grad2	class of degree is top 50%-25%
grad3	class of degree is top 25%
Working location-- base case the West	
East	include Beijing, Tianjin, Shanghai, Liaoning,
	Fujian, Shandong, Guangdong, Jiangsu
	Hebei, Zhejiang, Hainan
Mid	include Hei longjiang, Jilin, Anhui, Henan
	Jiangxi, Huibei, Hunan, Chongqing
Sector-- base case "other"	
gov	work in the government or related bureau
jon	work in a joint-venture company
stat	work in a state-owned company
edui	work in an education institution
Work place-- base case village	
wor1	work in a small town
wor2	work in a small city
wor3	work in a large city
Degree faculty-- base case "other"	
bio	major in biology
math	major in mathematics
phy	major in physics
elec	major in electronic and computer
cons	major in construction
soc	major in sociology
pol	major in politics
lan	major in language
lit	major in literature
art	major in art
chem	major in chemistry

Table 4.3 variable definition-Continue

econ	major in economics
law	major in economics
med	major in medicine
man	major in management
lev1	Threshold level1, the job require college graduate qualification
lev2	Threshold level2, the job require first degree qualification
lev3	Threshold level3, the job require master qualification
lev4	Threshold level4, the job require PhD qualification

Table 4.4 Determinants of the probability of a job offer at different levels using the ordered probit model

Var	Ordered probit model			Ordered probit model		
	comb	**male**	**female**	**comb**	**male**	**female**
gender	-0.251*** 0.045			-0.003 0.079		
lnsch	2.648*** 0.187	2.514*** 0.227	3.404*** 0.370			
lnba				0.602 0.561	0.703 0.887	1.383 1.321
lnma				1.910*** 0.267	1.775*** 0.371	2.591*** 0.620
lnphd				3.152*** 0.504	3.212*** 0.605	3.405*** 0.995
Pcar1	-0.194*** 0.048	-0.165*** 0.061	-0.254*** 0.082	-0.210*** 0.049	-0.176*** 0.061	-0.278*** 0.083
Pcar2	0.020 0.048	0.039 0.060	-0.007 0.082	0.018 0.049	0.039 0.060	-0.016 0.082
Pcar3	-0.043 0.060	-0.038 0.076	-0.073 0.101	-0.035 0.060	-0.031 0.076	-0.071 0.101
Pqu1	0.029 0.048	-0.008 0.057	0.085 0.089	0.018 0.048	-0.017 0.058	0.072 0.089
Pqu2	0.083* 0.050	0.059 0.060	0.131 0.092	0.068 0.050	0.046 0.060	0.112 0.093
Pqu3	0.101 0.067	0.097 0.082	0.105 0.119	0.083 0.067	0.083 0.082	0.082 0.120
Pqu4	-0.018 0.069	-0.068 0.085	0.066 0.121	-0.043 0.069	-0.084 0.086	0.034 0.123
Pqu5	-0.581*** 0.130	-0.431*** 0.166	-0.816*** 0.217	-0.617*** 0.130	-0.460*** 0.167	-0.850*** 0.218
cadre	0.070* 0.039	0.067 0.048	0.124* 0.071	-0.097* 0.058	-0.076 0.083	-0.049 0.129
Reg1	-0.125** 0.058	-0.102 0.071	-0.168 0.105	-0.115** 0.059	-0.093 0.071	-0.152* 0.105
Reg2	-0.078* 0.046	-0.143*** 0.055	0.071 0.086	-0.076* 0.046	-0.139*** 0.055	0.084 0.086

Table 4.4 Determinants of the probability of a job offer at different levels using the ordered probit model –continued

Reg3	-0.077	-0.086	-0.012	-0.081*	-0.088	-0.004
	0.049	0.060	0.089	0.049	0.060	0.089
pmem	0.261***	0.236***	0.318***	0.233***	0.213***	0.289***
	0.039	0.050	0.066	0.040	0.052	0.068
rank1	0.492***	0.470***	0.582***	0.127	0.141	0.231
	0.083	0.107	0.141	0.125	0.187	0.261
rank2	0.630***	0.577***	0.797***	0.115	0.118	0.292
	0.096	0.123	0.169	0.164	0.247	0.359
grad1	-0.099	-0.100	-0.216	-0.030	-0.035	-0.137
	0.073	0.082	0.170	0.075	0.088	0.177
grad2	-0.184***	-0.152*	-0.420***	-0.074	-0.050	-0.293
	0.071	0.081	0.163	0.077	0.094	0.181
grad3	-0.180**	-0.139*	-0.426***	-0.036	-0.008	-0.266
	0.074	0.085	0.166	0.083	0.105	0.193
Eng1	0.102***	0.149***	0.002	0.206***	0.232***	0.110
	0.040	0.049	0.074	0.049	0.063	0.101
Eng2	0.058	0.083	-0.024	0.201***	0.204**	0.135
	0.055	0.070	0.093	0.066	0.089	0.136
Wor1	-0.060	-0.088	-0.081	-0.070	-0.107	-0.082
	0.164	0.201	0.291	0.164	0.202	0.291
Wor2	0.172	0.212	-0.020	0.162	0.202	-0.023
	0.155	0.187	0.284	0.155	0.187	0.284
Wor3	0.440***	0.506***	0.205	0.429***	0.496***	0.205
	0.153	0.184	0.282	0.153	0.184	0.282
east	-0.031	0.010	-0.125*	-0.037	0.007	-0.125
	0.039	0.049	0.067	0.039	0.049	0.067
mid	0.031	-0.055	0.295***	0.025	-0.060	0.295***
	0.060	0.074	0.110	0.060	0.074	0.110
Gov	0.022	-0.102	0.326	0.019	-0.103	0.327***
	0.059	0.073	0.104	0.059	0.073	0.104
Stat	-0.046	-0.058	-0.042	-0.043	-0.057	-0.043
	0.045	0.056	0.082	0.045	0.056	0.082

Table 4.4 Determinants of the probability of a job offer at different levels using the ordered probit model –continued

Jon	0.120*	0.116	0.135	0.120*	0.117	0.131
	0.066	0.083	0.110	0.066	0.083	0.110
edui	0.310***	0.299***	0.335***	0.305***	0.294***	0.334***
	0.051	0.067	0.082	0.051	0.067	0.082
Bio	-0.306**	-0.382**	-0.204	-0.290**	-0.380**	-0.181
	0.144	0.188	0.233	0.144	0.188	0.234
Math	-0.076	-0.265*	0.062	-0.068	-0.252*	0.076
	0.097	0.140	0.146	0.097	0.140	0.146
Phy	-0.131	-0.061	-0.266	-0.119	-0.051	-0.256
	0.097	0.115	0.191	0.097	0.115	0.191
Elec	0.179***	0.195***	0.065	0.185***	0.197***	0.073
	0.048	0.055	0.100	0.048	0.055	0.101
Cons	0.082	0.058	0.233	0.086	0.057	0.228
	0.080	0.091	0.167	0.080	0.091	0.167
Soc	0.269	-0.009	0.890**	0.263	-0.023	0.877**
	0.238	0.297	0.407	0.238	0.297	0.407
Pol	-0.111	-0.293	0.014	-0.092	-0.291	0.032
	0.136	0.194	0.200	0.136	0.194	0.200
Lan	0.180**	-0.062	0.270***	0.203***	-0.059	0.286***
	0.078	0.147	0.106	0.078	0.147	0.107
Lit	-0.009	-0.006	0.021	-0.010	-0.007	0.020
	0.074	0.103	0.117	0.075	0.103	0.117
Art	0.163	0.185	0.167	0.140	0.158	0.167
	0.147	0.197	0.227	0.147	0.197	0.227
Chem	0.032	-0.044	0.411	0.047	-0.039	0.437
	0.135	0.154	0.289	0.135	0.154	0.289
econ	-0.027	-0.035	0.016	-0.016	-0.034	0.030
	0.081	0.104	0.137	0.082	0.104	0.137
Law	0.098	0.194	-0.053	0.104	0.199	-0.047
	0.108	0.149	0.164	0.108	0.149	0.164
med	0.169	0.227	0.144	0.175*	0.232	0.157
	0.106	0.155	0.156	0.107	0.155	0.157
man	0.016	0.068	-0.046	0.020	0.068	-0.040
	0.053	0.065	0.097	0.053	0.065	0.097

Table 4.4 Determinants of the probability of a job offer at different levels using the ordered probit model –continued

Mill	-0.329***	-0.328***	-0.412***	0.032	0.008	-0.073
	0.039	0.050	0.072	0.101	0.166	0.224
Lev1	1.463	1.491	1.943	-0.798	-0.551	-0.453
	0.299	(0.364)	(0.619)	0.655	1.028	1.625
Lev2	2.481	2.460	3.097	0.219	0.417	0.700
	0.298	(0.364)	(0.617)	0.655	1.028	1.625
Lev3	4.410	4.320	5.237	2.149	2.276	2.840
	0.301	(0.367)	(0.622)	0.656	1.029	1.626
Lev4	5.620	5.477	6.671	3.373	3.442	4.285
	0.309	(0.375)	(0.646)	0.656	1.029	1.629
Loglihood	-5445.99	-3608.95	-1773.26	-5437.94	-3605.72	-1771.92
Pseudo R2	0.100	0.091	0.145	0.102	0.092	0.146
Obs	5289	3393	1896	5289	3393	1896

Note: the results were estimated by probit model in Stata 8. * means the coefficient is significant at the10 per cent level, ** means the coefficient is significant at the 5 per cent level and *** means the coefficient is significant at the1 per cent level.

The first step probit estimation of table 4.4

Var	comb	male	female	comb	male	female
gender	-0. 137***			-0. 138***		
	0. 027			0. 026		
lnsch	0. 343***	0. 251***	0. 424***			
	0. 055	0. 073	0. 087			
lnba				0. 220***	0. 306***	0. 398***
				0. 030	0. 035	0. 042
lnma				0. 263***	0. 344***	0. 457***
				0. 052	0. 060	0. 071
lnphd				-0. 105	0. 145*	-0. 022
				0. 075	0. 079	0. 112
Pcar1	-0. 080**	-0. 026	-0. 146***	-0. 082**	0. 000	-0. 172**
	0. 038	0. 051	0. 057	0. 038	0. 045	0. 052
Pcar2	0. 076**	0. 082	0. 065	0. 072**	0. 066	0. 091*
	0. 037	0. 051	0. 056	0. 037	0. 046	0. 051
Pcar3	0. 228***	0. 254***	0. 193***	0. 223***	0. 216***	0. 216***
	0. 048	0. 066	0. 072	0. 048	0. 058	0. 066
Pqu1	0. 131*	0. 075	0. 192*	0. 124*	0. 025	0. 064
	0. 073	0. 097	0. 111	0. 072	0. 086	0. 101
Pqu2	0. 132*	0. 132	0. 120	0. 129*	0. 085	0. 010
	0. 072	0. 097	0. 110	0. 072	0. 086	0. 099
Pqu3	0. 196***	0. 258***	0. 115	0. 193***	0. 221***	0. 044
	0. 067	0. 088	0. 102	0. 066	0. 078	0. 092
Pqu4	0. 026	0. 074	-0. 046	0. 026	0. 056	-0. 048
	0. 067	0. 088	0. 103	0. 067	0. 078	0. 093
Pqu5	-0. 065	-0. 000	-0. 166	-0. 070	-0. 007	-0. 147
	0. 069	0. 091	0. 109	0. 069	0. 080	0. 098
cadre	0. 101***	0. 093***	0. 111***	0. 102***	0. 141***	0. 140***
	0. 026	0. 035	0. 039	0. 026	0. 031	0. 036
Reg1	-0. 202***	-0. 154***	-0. 250***	-0. 210***	-0. 067	-0. 200***
	0. 034	0. 047	0. 050	0. 034	0. 042	0. 045

The first step probit estimation of table 4.4-continued

Reg2	-0. 220***	-0. 180***	-0. 263***	-0. 225***	-0. 093*	-0. 261***
	0. 045	0. 061	0. 065	0. 044	0. 054	0. 059
Reg3	0. 207***	0. 131***	0. 309***	0. 213***	0. 046	0. 255***
	0. 037	0. 049	0. 059	0. 037	0. 044	0. 053
pmem	-0. 051	-0. 047	-0. 061	-0. 049	0. 011	-0. 013
	0. 033	0. 045	0. 048	0. 033	0. 040	0. 043
rank1	0. 454***	0. 472***	0. 413***	0. 407***	0. 648***	0. 417***
	0. 039	0. 053	0. 060	0. 040	0. 048	0. 055
rank2	0. 462***	0. 495***	0. 420***	0. 424***	0. 587***	0. 439***
	0. 048	0. 063	0. 075	0. 049	0. 056	0. 067
grad1	-0. 192***	-0. 175***	-0. 220***	-0. 186***	-0. 100***	-0. 138***
	0. 029	0. 040	0. 042	0. 028	0. 035	0. 038
grad2	-0. 232***	-0. 203***	-0. 279***	-0. 234***	-0. 135***	-0. 158***
	0. 038	0. 048	0. 063	0. 038	0. 042	0. 056
grad3	-0. 288***	-0. 232***	-0. 417***	-0. 281***	-0. 190***	-0. 216**
	0. 059	0. 068	0. 126	0. 058	0. 060	0. 110
Eng1	0. 709***	0. 750***	0. 669***	0. 706***	0. 543***	0. 596***
	0. 040	0. 053	0. 060	0. 039	0. 047	0. 054
Eng2	0. 385***	0. 327***	0. 474***	0. 400***	0. 139***	0. 316***
	0. 037	0. 049	0. 056	0. 037	0. 044	0. 051
Bio	-0. 102	-0. 121	-0. 012	-0. 142*	-0. 334***	-0. 105
	0. 088	0. 123	0. 128	0. 087	0. 111	0. 119
Math	-0. 274***	-0. 545***	0. 026	-0. 290***	-0. 474***	0. 089
	0. 067	0. 097	0. 099	0. 067	0. 084	0. 088
Phy	-0. 441***	-0. 481***	-0. 336**	-0. 451***	-0. 245***	-0. 159
	0. 080	0. 096	0. 148	0. 080	0. 080	0. 125
Elec	0. 013	0. 017	0. 019	0. 027	0. 056	0. 010
	0. 039	0. 047	0. 071	0. 039	0. 042	0. 064
Cons	0. 484***	0. 492***	0. 480***	0. 464***	0. 447***	0. 643***
	0. 093	0. 108	0. 181	0. 092	0. 095	0. 158

The first step probit estimation of table 4.4–continued

Soc	-0. 400***	-0. 401**	-0. 383*	-0. 407***	-0. 509***	-0. 503**
	0. 152	0. 204	0. 228	0. 152	0. 186	0. 230
Pol	-0. 469***	-0. 579***	-0. 336***	-0. 494***	-0. 597***	-0. 300***
	0. 095	0. 158	0. 123	0. 095	0. 136	0. 114
Lan	-0. 089	-0. 256**	0. 024	-0. 102*	-0. 188**	0. 094
	0. 056	0. 107	0. 074	0. 056	0. 096	0. 068
Lit	-0. 260***	-0. 415***	-0. 107	-0. 275***	-0. 353***	-0. 132*
	0. 054	0. 080	0. 078	0. 054	0. 071	0. 072
Art	-0. 267***	-0. 357***	-0. 137	-0. 270***	-0. 385***	-0. 247**
	0. 078	0. 113	0. 111	0. 078	0. 103	0. 105
Chem	-0. 361***	-0. 411***	-0. 280*	-0. 373***	-0. 315***	-0. 429***
	0. 089	0. 113	0. 148	0. 089	0. 101	0. 139
econ	-0. 223***	-0. 284***	-0. 124	-0. 237***	-0. 251***	-0. 128
	0. 063	0. 086	0. 094	0. 063	0. 077	0. 087
Law	-0. 403	-0. 467***	-0. 317***	-0. 428***	-0. 542***	-0. 449***
	0. 078	0. 110	0. 114	0. 078	0. 010	0. 105
med	-0. 405***	-0. 286**	-0. 419***	-0. 428***	-0. 458***	-0. 488***
	0. 086	0. 126	0. 120	0. 086	0. 110	0. 106
man	0. 092**	0. 143	0. 003	0. 095**	0. 090*	0. 016
	0. 042	0. 055	0. 070	0. 043	0. 049	0. 063
const	-0. 121	-0. 091***	-0. 286*	0. 125	-0. 013	0. 06500
	0. 101	0. 131	0. 156	0. 085	0. 098	. 118
Loglihood	- 8677. 40 82	- 4932. 75 47	- 3698. 01 76	- 8729. 71 19	- 4975. 59 32	- 3707. 48 59
Pseudo R2	0. 2024	0. 1977	0. 2052	0. 2038	0. 1974	0. 2087
Obs	16005	9235	6770	16005	9235	6770

Table 4.5 Determinants of the probability of a job offer at different levels using the ordered logit model

	Ordered logit model			Ordered logit model		
	comb	**male**	**female**	**comb**	**male**	**female**
gender	-0.495*** 0.080			-0.017 0.140		
lnsch	5.279*** 0.344	5.021*** 0.414	6.683*** 0.702			
lnba				1.469 0.981	1.192 1.541	3.576 2.386
lnma				3.959*** 0.473	3.548*** 0.651	5.444*** 1.137
lnphd				6.190*** 0.915	6.266*** 1.087	6.661*** 1.860
Pcar1	-0.313*** 0.087	-0.271*** 0.109	-0.384*** 0.148	-0.344*** 0.087	-0.294*** 0.109	-0.422*** 0.150
Pcar2	0.036 0.087	0.034 0.108	0.068 0.151	0.033 0.087	0.037 0.108	0.053 0.151
Pcar3	-0.061 0.110	-0.063 0.139	-0.047 0.185	-0.047 0.110	-0.050 0.139	-0.044 0.185
Pqu1	0.048 0.086	0.003 0.102	0.124 0.160	0.032 0.086	-0.015 0.103	0.105 0.161
Pqu2	0.150 0.089	0.132 0.107	0.196 0.167	0.124 0.090	0.105 0.108	0.170 0.168
Pqu3	0.191 0.120	0.208 0.146	0.179 0.216	0.164 0.120	0.179 0.147	0.147 0.218
Pqu4	-0.012 0.125	-0.066 0.154	0.108 0.220	-0.056 0.125	-0.100 0.155	0.061 0.223
Pqu5	-1.292*** 0.250	-0.869*** 0.314	-2.057*** 0.416	-1.358*** 0.250	-0.929*** 0.315	-2.097*** 0.416
cadre	0.129* 0.071	0.123 0.086	0.211 0.131	-0.184* 0.103	-0.182 0.145	-0.058 0.237
Reg1	-0.224** 0.105	-0.186 0.128	-0.324* 0.189	-0.205** 0.105	-0.170 0.128	-0.299 0.189
Reg2	-0.184** 0.083	-0.304*** 0.099	0.052 0.158	-0.184** 0.083	-0.299*** 0.100	0.070 0.158

Table 4.5 Determinants of the probability of a job offer at different levels using the ordered logit model—continued

Reg3	-0.164*	-0.182*	-0.063	-0.173**	-0.186*	-0.052
	0.089	0.109	0.162	0.089	0.109	0.162
pme	0.505***	0.457***	0.631***	0.452***	0.409***	0.587***
m	0.072	0.091	0.121	0.073	0.094	0.126
rank1	0.908***	0.897***	1.031***	0.224	0.194	0.490
	0.146	0.187	0.252	0.219	0.327	0.470
rank2	1.193***	1.120***	1.475***	0.231	0.142	0.697
	0.171	0.216	0.307	0.287	0.433	0.648
grad1	-0.185	-0.216	-0.203	-0.060	-0.081	-0.079
	0.136	0.150	0.329	0.139	0.158	0.341
grad2	-0.342***	-0.318**	-0.577*	-0.140	-0.104	-0.378
	0.132	0.147	0.318	0.141	0.168	0.349
grad3	-0.321**	-0.286*	-0.554*	-0.058	-0.014	-0.307
	0.138	0.156	0.323	0.152	0.188	0.369
Eng1	0.185***	0.284***	-0.031	0.382***	0.459***	0.138
	0.073	0.088	0.134	0.087	0.112	0.184
Eng2	0.082	0.153	-0.131	0.349***	0.410***	0.113
	0.101	0.128	0.172	0.119	0.159	0.248
Wor1	0.348	0.369	0.320	0.323	0.323	0.312
	0.299	0.378	0.506	0.299	0.378	0.507
Wor2	0.776***	0.897***	0.463	0.749***	0.873***	0.453
	0.286	0.358	0.496	0.287	0.357	0.497
Wor3	1.238***	1.403***	0.854*	1.207***	1.379***	0.845*
	0.284	0.353	0.493	0.284	0.353	0.494
east	-0.038	0.022	-0.190	-0.049	0.015	-0.187
	0.070	0.087	0.123	0.070	0.088	0.123
mid	0.074	-0.087	0.562***	0.061	-0.100	0.560***
	0.110	0.133	0.203	0.110	0.134	0.203
Gov	0.097	-0.138	0.670***	0.091	-0.143	0.673***
	0.107	0.130	0.191	0.107	0.131	0.192
Stat	-0.090	-0.126	-0.037	-0.086	-0.123	-0.038
	0.081	0.098	0.147	0.081	0.098	0.147
Jon	0.225*	0.227	0.213	0.224*	0.227	0.212
	0.118	0.149	0.199	0.118	0.149	0.199

Table 4.5 Determinants of the probability of a job offer at different levels using the ordered logit model—continued

edui	0.607***	0.573***	0.667***	0.591***	0.552***	0.663***
	0.093	0.121	0.151	0.093	0.122	0.151
Bio	-0.462*	-0.567*	-0.381	-0.428*	-0.555*	-0.344
	0.258	0.341	0.416	0.259	0.341	0.417
Math	-0.050	-0.35	0.069	-0.035	-0.311	0.095
	0.176	0.263	0.264	0.176	0.264	0.264
Phy	-0.157	-0.007	-0.499	-0.132	0.022	-0.480
	0.171	0.202	0.339	0.171	0.202	0.339
Elec	0.373***	0.422***	0.078	0.384***	0.424***	0.091
	0.087	0.100	0.184	0.087	0.100	0.185
Cons	0.160	0.125	0.398	0.169	0.124	0.388
	0.142	0.160	0.305	0.142	0.160	0.305
Soc	0.596	0.138	1.565**	0.589	0.117	1.552**
	0.427	0.519	0.716	0.428	0.519	0.716
Pol	-0.190	-0.491	-0.070	-0.150	-0.482	-0.043
	0.235	0.336	0.352	0.235	0.336	0.352
Lan	0.360***	-0.086	0.459**	0.405***	-0.085	0.485***
	0.138	0.259	0.191	0.139	0.259	0.192
Lit	0.000	0.048	-0.063	-0.002	0.044	-0.062
	0.134	0.186	0.210	0.134	0.186	0.210
Art	0.299	0.377	0.196	0.259	0.335	0.199
	0.251	0.333	0.400	0.250	0.333	0.399
Chem	0.115	0.018	0.555	0.142	0.031	0.596
	0.243	0.274	0.542	0.243	0.274	0.542
econ	-0.020	-0.019	-0.054	0.003	-0.016	-0.031
	0.146	0.185	0.247	0.146	0.186	0.248
Law	0.171	0.390	-0.220	0.177	0.399	-0.211
	0.206	0.286	0.313	0.206	0.286	0.312
med	0.319	0.450*	0.201	0.327*	0.460*	0.221
	0.193	0.275	0.292	0.193	0.275	0.293
man	0.039	0.178	-0.196	0.048	0.179	-0.186
	0.095	0.117	0.174	0.095	0.117	0.174
Mill	-0.655***	-0.658***	-0.801***	0.015	0.053	-0.281
	0.070	0.089	0.134	0.176	0.289	0.404

Table 4.5 Determinants of the probability of a job offer at different levels using the ordered logit model—continued

Lev1	3.573	3.697 (0.671)	4.563 1.146	-0.646 1.153	-0.632 1.799	0.881 (2.933)
Lev2	5.491	5.508 (0.670)	6.746 1.140	1.265 1.153	1.176 1.799	3.061 (2.933)
Lev3	8.840	8.723 (0.678)	10.525 1.156	4.616 1.155	4.388 1.801	6.839 (2.938)
Lev4	11.390	11.151 (0.713)	13.505 1.231	7.228 1.160	6.863 1.804	9.855 (2.950)
Loglihood	-5380.322	-3569.866	-1747.856	-5371.241	-3565.555	-1746.893
Pseudo R2	0.1111	0.1010	0.1575	0.1126	0.1021	0.1579
obs	5289	3393	1896	5289	3393	1896

Table 4.6 A Quadratic *lnsch* model for the determinants of the probability of a higher level job offer

	Ordered logit model			Ordered logit model		
	comb	**male**	**female**	**comb**	**male**	**female**
gender	-0.012 0.139			-0.017 0.140		
lnsch	-2.863 1.951	-3.535 3.030	0.578 4.465			
lnba				-1.560 1.035	-1.875 1.553	0.316 2.298
lnma				-0.895 0.883	-1.366 1.088	0.220 1.665
lnphd						
$(lnsch)^2$	2.060*** 0.487	2.147*** 0.754	1.460 1.056	1.457*** 0.215	1.475*** 0.256	1.568*** 0.438
Pcar1	-0.343*** 0.087	-0.292*** 0.109	-0.422*** 0.150	-0.344*** 0.087	-0.294*** 0.109	-0.422*** 0.150
Pcar2	0.033 0.087	0.038 0.108	0.053 0.151	0.033 0.087	0.037 0.108	0.053 0.151
Pcar3	-0.046 0.110	-0.045 0.139	-0.044 0.185	-0.047 0.110	-0.050 0.139	-0.044 0.185
Pqu1	0.032 0.086	-0.017 0.103	0.106 0.161	0.032 0.086	-0.015 0.103	0.105 0.161
Pqu2	0.123 0.090	0.103 0.108	0.171 0.168	0.124 0.090	0.105 0.108	0.170 0.168
Pqu3	0.163 0.120	0.175 0.147	0.147 0.218	0.164 0.120	0.179 0.147	0.147 0.218
Pqu4	-0.058 0.125	-0.107 0.155	0.062 0.223	-0.056 0.125	-0.100 0.155	0.061 0.223
Pqu5	-1.359*** 0.250	-0.936*** 0.315	-2.096*** 0.416	-1.358*** 0.250	-0.929*** 0.315	-2.097*** 0.416
cadre	-0.187* 0.103	-0.202 0.143	-0.055 0.233	-0.184* 0.103	-0.182 0.145	-0.058 0.237
Reg1	-0.205** 0.105	-0.170 0.128	-0.300 0.189	-0.205** 0.105	-0.170 0.128	-0.299 0.189
Reg2	-0.185** 0.083	-0.299*** 0.099	0.069 0.158	-0.184** 0.083	-0.300*** 0.099	0.070 0.158

Table 4.6 A Quadratic *lnsch* model for the determinants of the probability of a higher level job offer—continued

Reg3	-0.173** 0.089	-0.184* 0.109	-0.052 0.162	-0.173** 0.089	-0.186* 0.109	-0.052 0.162
pme m	0.450*** 0.073	0.402*** 0.093	0.587*** 0.125	0.452*** 0.073	0.409*** 0.094	0.587*** 0.126
rank1	0.218 0.219	0.152 0.321	0.496 0.461	0.224 0.219	0.194 0.327	0.490 0.470
rank2	0.221 0.286	0.081 0.423	0.706 0.634	0.231 0.287	0.142 0.433	0.697 0.648
grad1	-0.059 0.139	-0.074 0.158	-0.080 0.340	-0.060 0.139	-0.081 0.158	-0.079 0.341
grad2	-0.139 0.141	-0.091 0.167	-0.379 0.348	-0.140 0.141	-0.104 0.168	-0.378 0.349
grad3	-0.055 0.152	0.005 0.186	-0.310 0.367	-0.058 0.152	-0.014 0.188	-0.307 0.369
Eng1	0.385*** 0.087	0.471*** 0.110	0.135 0.180	0.382*** 0.087	0.459*** 0.112	0.138 0.184
Eng2	0.350*** 0.119	0.420*** 0.158	0.110 0.245	0.349*** 0.119	0.410*** 0.159	0.113 0.248
Wor1	0.323 0.299	0.322 0.378	0.312 0.507	0.323 0.299	0.323 0.378	0.312 0.507
Wor2	0.749*** 0.287	0.874*** 0.357	0.453 0.497	0.749*** 0.287	0.873*** 0.357	0.453 0.497
Wor3	1.206*** 0.284	1.377*** 0.352	0.845* 0.494	1.207*** 0.284	1.379*** 0.353	0.845* 0.494
east	-0.049 0.070	0.014 0.088	-0.187 0.123	-0.049 0.070	0.015 0.088	-0.187 0.123
mid	0.062 0.110	-0.097 0.133	0.560*** 0.203	0.061 0.110	-0.100 0.134	0.560*** 0.203
Gov	0.091 0.107	-0.144 0.130	0.673*** 0.192	0.091 0.107	-0.143 0.131	0.673*** 0.192
Stat	-0.086 0.081	-0.122 0.098	-0.038 0.147	-0.086 0.081	-0.123 0.098	-0.038 0.147
Jon	0.224* 0.118	0.227 0.149	0.212 0.199	0.224* 0.118	0.227 0.149	0.212 0.199

Table 4.6 A Quadratic *lnsch* model for the determinants of the probability of a higher level job offer—continued

edui	0.592***	0.554***	0.663***	0.591***	0.552***	0.663***
	0.093	0.122	0.151	0.093	0.122	0.151
Bio	-0.425*	-0.551	-0.345	-0.428*	-0.555*	-0.344
	0.259	0.341	0.417	0.259	0.341	0.417
Math	-0.033	-0.305	0.094	-0.035	-0.311	0.095
	0.176	0.264	0.264	0.176	0.264	0.264
Phy	-0.131	0.025	-0.481	-0.132	0.021	-0.480
	0.171	0.202	0.339	0.171	0.202	0.339
Elec	0.386***	0.430***	0.090	0.384***	0.424***	0.091
	0.087	0.100	0.185	0.087	0.100	0.185
Cons	0.171	0.128	0.388	0.169	0.124	0.388
	0.142	0.160	0.305	0.142	0.160	0.305
Soc	0.592	0.121	1.551**	0.589	0.117	1.552**
	0.428	0.519	0.716	0.428	0.519	0.716
Pol	-0.147	-0.475	-0.044	-0.150	-0.482	-0.043
	0.235	0.336	0.352	0.235	0.336	0.352
Lan	0.407***	-0.085	0.484***	0.405***	-0.085	0.485***
	0.139	0.259	0.192	0.139	0.259	0.192
Lit	-0.001	0.048	-0.063	-0.002	0.044	-0.062
	0.134	0.186	0.210	0.134	0.186	0.210
Art	0.259	0.333	0.199	0.259	0.335	0.199
	0.250	0.333	0.399	0.250	0.333	0.399
Chem	0.143	0.035	0.596	0.142	0.031	0.596
	0.243	0.274	0.542	0.243	0.274	0.542
econ	0.005	-0.011	-0.032	0.003	-0.016	-0.031
	0.146	0.185	0.248	0.146	0.186	0.248
Law	0.179	0.407	-0.211	0.177	0.399	-0.211
	0.206	0.286	0.312	0.206	0.286	0.313
med	0.330*	0.466*	0.220	0.327*	0.460*	0.221
	0.193	0.275	0.293	0.193	0.275	0.293
man	0.051	0.184	-0.187	0.048	0.179	-0.186
	0.095	0.117	0.174	0.095	0.117	0.174
Mill	0.023	0.101	-0.287	0.015	0.053	-0.281
	0.175	0.281	0.395	0.176	0.289	0.404

Table 4.6 A Quadratic *lnsch* model for the determinants of the probability of a higher level job offer—continued

Lev1	-1.687 2.231	-2.334 3.458	2.028 5.407	0.054 1.168	0.077 1.821	1.635 2.999
Lev2	0.223 2.232	-0.527 3.458	4.207 5.407	1.965 1.168	1.885 1.821	3.814 2.999
Lev3	3.576 2.233	2.688 3.460	7.985 5.412	5.316 1.170	5.097 1.822	7.593 3.000
Lev4	6.184 2.231	5.155 3.456	11.003 5.408	7.928 1.177	7.572 1.828	10.608 3.019
Loglihood	-5371.3059	-3565.7984	-1746.8954	-5371.2414	-3565.5548	-1746.893
Pseudo R2	0.1126	0.1020	0.1579	0.1126	0.1021	0.1579
obs	5289	3393	1896	5289	3393	1896

Table 4.7 Determinants of the probability of a job offer at different levels using ordered logit model by geographical area

Var	east		middle		west	
	male	female	male	female	male	female
lnsch	4.934***	6.902***	6.891***	9.763***	1.861*	8.227***
	0.435	0.777	1.854	4.869	1.097	1.721
lnpca	-0.026	-0.046	0.168	0.164	-0.065	-0.240
	0.104	0.144	0.320	0.480	0.158	0.227
lnpqu	0.047	0.183	-0.450	-0.105	0.071	-0.105
	0.101	0.152	0.310	0.589	0.156	0.234
cadre	0.049	0.234	0.028	-0.136	-0.098	0.443*
	0.105	0.159	0.289	0.661	0.174	0.270
lnregis	-0.238***	-0.134	-0.027	-0.662	-0.275**	0.071
	0.091	0.141	0.275	0.455	0.140	0.221
pmem	0.437***	0.597***	0.620*	0.308	0.398**	0.536**
	0.114	0.156	0.326	0.495	0.176	0.224
lnrank	0.754***	1.394***	1.018	3.549**	0.756*	1.886***
	0.210	0.321	0.678	1.681	0.443	0.648
lngrad	-0.213*	-0.447**	0.262	0.172	-0.306	-0.823**
	0.129	0.223	0.352	0.752	0.209	0.359
lnEng	0.155	-0.334*	0.484	-0.083	0.677***	0.067
	0.133	0.190	0.369	0.720	0.237	0.328
lnwor	1.765***	0.829***	0.135	3.302**	1.258***	0.670
	0.249	0.309	0.721	1.535	0.267	0.448
Gov	-0.027	0.650***	-0.147	2.474***	-0.104	0.216
	0.165	0.235	0.480	0.880	0.248	0.358
Stat	-0.121	0.139	-0.118	-0.133	0.002	-0.280
	0.123	0.181	0.305	0.599	0.202	0.294
Jon	0.172	0.312	1.305**	0.926	0.019	-0.198
	0.172	0.229	0.550	1.021	0.400	0.461
edui	0.633***	0.775***	0.309	0.459	0.681***	0.353
	0.157	0.187	0.408	0.673	0.233	0.299
Bio	-0.115	-0.542			-0.964**	-0.164
	0.506	0.608			0.484	0.595
Math	-0.842**	0.435	-0.575	0.242	0.363	0.075
	0.346	0.420	1.007	1.022	0.443	0.418

Table 4.7 Determinants of the probability of a job offer at different levels using ordered logit model by geographical area—continued

Phy	0.149	-0.150			-0.102	-1.344**
	0.334	0.460			0.288	0.549
Elec	0.537***	-0.098	0.298	1.232	0.042	0.277
	0.123	0.234	0.346	0.777	0.198	0.363
Cons	0.288	0.296	0.557	-0.151	-0.402	1.031
	0.196	0.358	0.498	1.440	0.374	0.663
Soc	0.787	2.063***	-1.982*	-0.372	0.481	
	0.634	0.774	1.054	2.581	1.323	
Pol	-0.614	0.332	-0.892	0.923	1.260	-1.278
	0.393	0.490	0.829	1.062	1.065	0.811
Lan	0.255	0.682***	-0.239	2.712***	-1.014**	-0.196
	0.340	0.242	0.933	1.033	0.498	0.364
Lit	-0.472**	0.051	-1.067	1.495	1.270***	-0.403
	0.216	0.242	0.970	1.267	0.382	0.506
Art	0.202	0.579	6.159***	1.075	0.375	-0.403
	0.473	0.681	2.003	2.738	0.493	0.571
Chem	0.150	1.217	0.008		-0.691	-0.120
	0.352	0.837	0.780		0.566	0.723
econ	0.180	-0.113	-0.949**	0.965	-0.423	0.051
	0.246	0.285	0.440	1.069	0.403	0.577
Law	0.700**	-0.206	-2.556***	1.260	0.146	-0.696
	0.349	0.368	0.948	1.214	0.543	0.700
med	-0.645	-0.958***	0.207	0.867	0.640**	0.936**
	0.575	0.390	1.961	1.305	0.325	0.448
man	0.145	0.044	0.545	0.347	0.098	-0.510
	0.144	0.217	0.341	0.765	0.259	0.353
Mills ratio	-0.575***	-0.827***	-0.913***	-1.562*	-0.110	-1.096***
	0.095	0.148	0.359	0.880	0.216	0.314
Lev1	4.196	5.400	4.484	13.554	0.483	5.728
	0.614	1.138	2.362	6.787	1.363	2.345
Lev2	5.923	7.413	6.296	15.829	2.537	8.162
	0.610	1.129	2.353	6.779	1.357	2.334
Lev3	9.240	11.267	9.585	20.241	5.641	11.864
	0.628	1.162	2.371	6.873	1.371	2.354

Table 4.7 Determinants of the probability of a job offer at different levels using ordered logit model by geographical area—continued

Lev4	11.768 0.685	14.795 1.307	12.606 2.633	23.154 7.015	7.751 1.402	14.413 2.515
Loglihood	-2221.661	-1080.116	-345.874	-128.690	-967.513	-516.765
Pseu--doR2	0.1102	0.1520	0.1177	0.2206	0.0965	0.1717
obs	2159	1191	336	159	898	546

Table 4.8 Determinants of the probability of a job offer at different levels using ordered logit model by industrial sector area

variables	gov	Stat	Jon	Edui	others
gender	0.371*	-0.460***	-0.407	-0.515***	-0.587***
	0.212	0.147	0.303	0.138	0.209
lnsch	4.420***	5.145***	3.682***	5.362***	4.444***
	0.769	0.629	1.153	0.517	0.985
lnpca	-0.114	-0.066	0.037	0.022	0.088
	0.095	0.059	0.109	0.066	0.080
lnpqu	-0.027	-0.035	0.034	-0.046	0.135**
	0.082	0.048	0.100	0.055	0.069
cadre	-0.122	0.000	0.577**	-0.063	0.027
	0.200	0.118	0.247	0.131	0.161
lnregis	-0.127	-0.077	0.005	0.017	-0.179***
	0.086	0.049	0.098	0.056	0.066
pmem	0.597***	0.431***	0.938***	0.436***	0.549***
	0.190	0.123	0.285	0.138	0.180
lnrank	-0.120	0.542***	0.842***	0.400***	0.308*
	0.200	0.123	0.271	0.123	0.178
lngrad	-0.040	-0.002	-0.084	-0.075	-0.142*
	0.105	0.063	0.134	0.069	0.083
lnEng	-0.279	0.032	-0.420	-0.049	0.090
	0.228	0.143	0.280	0.148	0.217
lnwor	0.489***	0.366***	0.148	0.598***	0.442***
	0.170	0.103	0.227	0.083	0.116
east	0.252	0.023	0.190	-0.203	-0.144
	0.200	0.115	0.311	0.134	0.167
Mid	0.489	0.082	0.551	-0.078	-0.043
	0.365	0.174	0.505	0.223	0.260
Bio	0.886	-1.570*	0.953	-0.407	-0.079
	1.374	0.826	2.016	0.318	0.711
Math	-0.027	-0.475	--	0.018	1.025
	0.551	0.589		0.219	1.047
Phy	0.448	0.290	0.673	-0.200	-0.477
	0.831	0.702	2.008	0.216	0.806
Elec	0.432	0.584***	0.028	0.102	0.180
	0.272	0.138	0.300	0.185	0.200

Table 4.8 Determinants of the probability of a job offer at different levels using ordered logit model by industrial sector area—continued

Cons	0.516 0.467	0.263 0.200	0.211 0.834	-0.071 0.420	0.295 0.307
Soc	1.268* 0.773	-0.689 0.996	-0.658 1.203	0.974 0.956	-0.091 1.039
Pol	-0.977* 0.516	-0.160 0.627	-0.630 2.030	0.048 0.322	0.762 0.834
Lan	0.624 0.430	0.426 0.312	-0.179 0.394	0.338 0.234	0.623* 0.339
Lit	-0.014 0.320	-0.007 0.271	-0.561 0.681	-0.115 0.220	0.298 0.347
Art	0.283 1.113	0.828* 0.478	-0.177 1.299	-0.154 0.512	0.540 0.427
Chem	-0.597 0.923	0.361 0.475	0.978 0.630	-0.613 0.392	0.188 0.641
econ	0.055 0.396	0.109 0.236	-0.520 0.409	0.585 0.548	-0.507* 0.309
Law	0.202 0.351	0.487 0.420	-0.205 0.669	0.718 0.679	0.192 0.503
med	0.291 0.480	0.776** 0.347	-0.866 0.803	0.536 0.479	-0.057 0.368
man	0.198 0.282	-0.037 0.150	0.110 0.323	0.283 0.254	-0.007 0.212
Mills ratio	-0.415*** 0.154	-0.560*** 0.121	-0.416* 0.222	-0.597*** 0.105	-0.425** 0.176
Lev1	2.288 1.285	3.531 0.876	2.483 1.902	4.019 0.807	2.408 1.280
Lev2	3.724 1.279	5.706 0.871	4.599 1.892	5.619 0.803	4.556 1.278
Lev3	6.969 1.301	9.160 0.892	8.029 1.928	8.924 0.816	7.997 1.302
Lev4	9.554 1.369	11.968 0.958	10.478 1.995	11.419 0.878	10.676 1.376
Loglihood	-673.879	-1754.061	-429.750	-1479.321	-989.859
PseudoR2	0.0950	0.0911	0.1136	0.1462	0.0965
obs	636	1805	428	1428	992

Table 4.9 Determinants of the probability of a job offer at different levels using the Logarithm of level variables

	Ordered probit model			Ordered logit model		
	comb	**male**	**female**	**comb**	**male**	**female**
gender	-0.201***			-0.396***		
	0.043			0.076		
lnsch	2.336***	2.167***	3.045***	4.674***	4.291***	6.044***
	0.165	0.200	0.326	0.303	0.363	0.612
lnpcar	-0.007	-0.002	-0.014	-0.012	-0.011	-0.003
	0.018	0.023	0.031	0.033	0.042	0.057
lnpqu	-0.007	-0.007	-0.009	-0.007	-0.002	-0.018
	0.015	0.019	0.026	0.028	0.034	0.049
cadre	-0.001	-0.002	0.043	0.004	-0.005	0.076
	0.037	0.045	0.067	0.066	0.081	0.122
lnreg	-0.035**	-0.046**	-0.005	-0.072***	-0.094***	-0.020***
	0.016	0.019	0.028	0.028	0.034	0.051
pmem	0.264***	0.248***	0.312***	0.497***	0.470***	0.598***
	0.039	0.050	0.065	0.071	0.091	0.120
lnrank	0.183***	0.143***	0.284***	0.372***	0.297***	0.560*
	0.037	0.045	0.069	0.068	0.081	0.129
lngrad	-0.041**	-0.028	-0.086**	-0.066*	-0.051	-0.125
	0.020	0.024	0.036	0.036	0.042	0.066
lnEng	0.039	0.061*	-0.007	0.058	0.112*	-0.054
	0.027	0.034	0.045	0.049	0.061	0.084
lnwor	0.253***	0.282***	0.191***	0.483***	0.538***	0.367***
	0.028	0.034	0.047	0.049	0.062	0.086
east	-0.045	-0.002	-0.134**	-0.053	0.018	-0.199*
	0.038	0.048	0.066	0.069	0.086	0.120
mid	0.015	-0.066	0.269***	0.049	-0.102	0.516***
	0.060	0.074	0.108	0.109	0.133	0.198
Gov	0.031	-0.090	0.334	0.117	-0.110	0.663***
	0.059	0.072	0.103	0.106	0.129	0.189
Stat	-0.032	-0.045	-0.030	-0.069	-0.104	-0.024
	0.045	0.055	0.081	0.080	0.098	0.145
Jon	0.126**	0.128	0.132	0.218*	0.226	0.208
	0.065	0.082	0.109	0.117	0.148	0.196
edui	0.297***	0.297***	0.298***	0.581***	0.565***	0.603***
	0.051	0.067	0.082	0.092	0.121	0.149

Table 4.9 Determinants of the probability of a job offer at different levels using the Logarithm of level variables—continued

Bio	-0.259*	-0.352*	-0.100	-0.370	-0.520	-0.203
	0.143	0.187	0.230	0.256	0.338	0.410
Math	-0.058	-0.269**	0.112	-0.012	-0.360	0.173
	0.097	0.140	0.144	0.175	0.261	0.259
Phy	-0.119	-0.044	-0.337*	-0.133	0.029	-0.599*
	0.096	0.113	0.188	0.170	0.200	0.334
Elec	0.163***	0.180***	0.053	0.337***	0.380***	0.076
	0.047	0.053	0.098	0.085	0.097	0.181
Cons	0.123	0.095	0.258	0.225	0.173	0.450
	0.079	0.090	0.165	0.141	0.159	0.301
Soc	0.293	-0.011	0.969	0.610	0.119	1.634**
	0.237	0.296	0.400	0.424	0.516	0.722
Pol	-0.117	-0.300	0.000	-0.194	-0.500	-0.102
	0.135	0.194	0.197	0.234	0.335	0.344
Lan	0.177**	-0.088	0.294***	0.365***	-0.099	0.511***
	0.077	0.146	0.105	0.137	0.258	0.188
Lit	-0.042	-0.053	-0.001	-0.066	-0.062	-0.073
	0.073	0.101	0.114	0.132	0.183	0.207
Art	0.198	0.250	0.137	0.367	0.489	0.169
	0.146	0.195	0.225	0.250	0.331	0.394
Chem	-0.020	-0.090	0.338	0.020	-0.074	0.528
	0.134	0.153	0.285	0.241	0.272	0.530
econ	-0.070	-0.083	-0.020	-0.119	-0.117	-0.130
	0.081	0.103	0.135	0.144	0.184	0.242
Law	0.064	0.143	-0.099	0.124	0.320	-0.280
	0.107	0.147	0.161	0.204	0.283	0.308
med	0.152	0.207	0.122	0.289	0.434	0.148
	0.106	0.153	0.154	0.193	0.274	0.291
man	0.014	0.072	-0.067	0.033	0.169	-0.208
	0.052	0.064	0.096	0.094	0.115	0.172
Mill	-0.238***	-0.225***	-0.324***	-0.487***	-0.452***	-0.653***
	0.032	0.040	0.059	0.057	0.072	0.110
Lev1	1.536	1.549	2.082	3.394	3.365	4.462
	0.235	0.280	0.505	0.429	0.505	0.942

Table 4.9 Determinants of the probability of a job offer at different levels using the Logarithm of level variables—continued

Lev2	2.542 0.235	2.508 0.279	3.210 0.502	5.291 0.427	5.162 0.502	6.597 0.935
Lev3	4.455 0.238	4.353 0.284	5.318 0.509	8.608 0.437	8.347 0.515	10.307 0.956
Lev4	5.651 0.248	5.495 0.295	6.735 0.537	11.126 0.472	10.741 0.554	13.228 1.034
Loglihood	-5489.919	-3635.136	-1799.591	-5419.527	-3635.136	-1799.591
Pseudo R2	0.0930	0.0846	0.1325	0.1039	0.0946	0.1436
obs	5289	3393	1896	5289	3393	1896

Appendix A Questionnaire

Basic questions

1. Email address
2. Graduate university
3. University belongings(province)
4. Highest qualification
5. Specialty
6. Age
7. Gender
8. Race
9. Registration
10. Score for entrance examination to university
11. father's career
12. mother's career
13. father's highest qualification
14. mother's highest qualification
15. your current situation

Questions about university

16. Are you a party number?
17. Have you passed College English Test(CET) 4?
18. Have you passed CET-6?
19. Other qualifications
20. Are you interested in your specialty
21. Are you satisfied with the teaching level?
22. Do you want to transfer to other specialty?
23. Class rank
24. Working experience
25. Have you got scholarship during the four year study?

Questions about employment

1. How much do you know about your current employed company
2. Which province does your company belongs to?
3. Where do you work?

A. Large or middle city B. small city C. town D. village

4. Monthly glossary payoff
5. Are you satisfied with your current company
6. Whether your current job is related with what you have learned in the university
7. What is the minimum formal qualification required in your contracted job?
8. How many people in your company
9. Which industry sector does your company belong to?

A. state-owned B. foreign C. school or university D. government

E. joined venture F. others

Chapter 5 Chinese Graduates' Education Choices: considering overeducation uncertainty

5.1. Introduction

In the previous chapter, we learned about 20% individuals are overeducated to their current jobs and individuals characteristics, geographic area and industry sector may all affect individuals positions on getting a matched job. But what is interesting to individuals is whether overeducation will affect their earnings, what kinds of factors are determining their payment and what are the optimal education choices.

According to the pecking order theory in chapter 4, individuals' wages were determined by individuals' other characteristics, not their qualification level or job level. However for individuals who have the same characteristics may have different levels of qualification and for individuals with the same level of qualification may get varied level of jobs. It is very instructive to find out the impact of job level and surplus (deficit) schooling on wages in order to give a guideline for individuals to make education decisions.

In this chapter we will first examine the determinants of graduation wages by the same dataset we used in chapter 4 to see whether individuals' characteristics, geographic area and industry sector still play a role in wages. And then the effect of job level, surplus schooling and deficit schooling on individuals wages. According to these regression results and pecking order theory, we evaluate individuals' optimal education choices under uncertainty. Finally we discuss the impact of changes over time on individuals' education decisions.

5.2. The Determinants of Graduation Wages

In the Pecking Order theory of Section 2 in chapter 4, wages were determined only by individual characteristics, as in equations (4.2) and (4.7). In this section, we want to examine what kind of factors may affect wages empirically and to what extent. Table 5.1 below shows the results of regressing the natural log of wages on graduation on individual characteristics, including firstly lnschooling years, and secondly on each separate additional qualification level compared to being a basic college graduate. In order to control for selection bias on the determination of wages by including only those graduates who have found a job, we include the relevant Inverse Mills ratio as well. The value of F-test against the hypotheses that all the coefficients except the constant term are zero in each case proves to be strongly significantly different from zero.

Table 5.1 shows the estimated percentage increase in graduation wages from each subsequent percentage in years spent in higher education is highly significantly different from zero at 9.5 per cent for both genders and 7.9 per cent for male graduates and 9.9 per cent for female graduates. Decomposing the years spent in higher education significantly improves the goodness of fit, with variations in the estimated percentage increase in graduation wages from subsequent percentages increases in years spent in higher education at bachelors, Masters, and PhD level respectively of 3.7 per cent, 17.1 per cent and 15 per cent for both genders combined. Hence investing in a Masters degree yields the highest rewards, with investment in a PhD yielding an estimated negative return (though one which is not significantly different from zero). The results accord to our expectation, with a recent survey by a graduate recruiting website[26] in China displaying mean and median wages for new Masters that are all higher than the wages for those with new Doctorates. This is also consistent with Flenette (2004)'s Canadian graduate labour market study that the return to master degree is the highest. Whilst being a female in the labour market itself

[26] www.chinahr.com

makes the baseline wage more than 4.8 per cent lower, the marginal return to female undertaking additional years of higher education through bachelors and Masters degrees exceeds that for males.

In line with the finding of Rumberger & Thomas (1993) that college quality had a significant impact on earnings, we find that the marginal impact on wages from graduating from a top 100 university in China is 14.4 per cent. In China, English skills can also have an important impact on graduation wages, with having a qualification of College English Test level 6 (CET-6) yielding a marginal return on graduation wages of 15.9 per cent for men, though not a significant effect for women. In contrast to its effect in some western countries (see Dolton & Vignoles, 2000), the class or grade of degree does not have a significant positive effect on graduation wages in China. However, being a student leader (cadre) or coming from a large city does have a significant positive effect on wages. So does parents' education and career for the separate qualification logit model. It is also interesting to note that whilst in China party membership plays on important role in raising the probability of securing a better job, it has no direct effect on graduation wages.

Starting salaries for graduates in China differ also across the type of employer involved, especially for men. Relative to working in other sectors, working in the government, state-owned companies and education institutions tends to imply a lower wage level, and with starting salaries in joint ventures significantly higher. However, since employees in government, state-owned firms and education institutions have higher welfare benefits, more stable employment and a less pressurised workload, lower wages may be being traded-off by graduates against these benefits, with a tendency towards equalisation of net advantages in attracting graduate employees. The region where the employment is located also has an impact on starting salaries for graduates, with working in the West of China implying less favourable payments than working in the East or Middle of China. This is in line with the findings elsewhere by Griffin &

Edwards (1993) that regional differences in earnings are not simply due to differences in individual characteristics.

The returns to the subject of the degree study are quite different to those found in studies in several other countries[27]. Studying language or literature generates a significant positive return in China, but a negative one in Western countries. On the other hand, a subject with a high positive return in many Western countries, namely medicine, has a significantly negative effect on wages in China overall, and particularly for women. One reason for the low initial return to studying medicine is the high requirement for experience in medical jobs before wages rise after lower initial graduation wages. Another is that the quality of medical schools in China varies greatly, with no strong entry barriers to entering lower quality medical schools in China, in contrast to the strong competition for medical school places in countries such as the UK and USA.

It is also notable that studying mathematics at university does not yield a significant positive return in China. In China, studying mathematics is compulsory up until university entrance, so mathematical skills are not in such short supply as they are in some Western countries, such as the UK, where there is evidence of a strongly positive rate of return to studying mathematics at school beyond the lower compulsory age (see Dolton & Vignoles, 2002). However, there is a significant positive return to studying electronics and computing (elec&comp), which play an important part in China's economic growth. There is also, but to a lesser extent, a significant positive return in China to studying physics. However, studying construction yields a significant negative return, despite China's current construction boom.

An interesting question which arises in the context of the Pecking Order model is whether the single index of individual characteristics that determines the probability of an individual receiving the offer of a higher

[27] Walker and Zhu(2003), Dolton and Vignoles(2000) for example

level job can also explain the graduation wage. Under the Pecking Order model in chapter 4, equations (4.7) and (4.9) imply that:

$$\ln w_{it} = \alpha_t + \varpi_t q'_{it} + \varepsilon'_{it} \quad \text{where } q'_{it} \equiv \sum_{h=1}^{n-1} x_{iht} a_h,\ \alpha_t \equiv \ln B_t,\ \varepsilon'_{it} \equiv \varpi_t \varepsilon_{it} \qquad (5.1)$$

where q'_{it} is the single index of observable individual characteristics that together with the stochastic term ε_{it} determines the probability of individual i receiving a job offer at a given level. In order to test empirically whether the lnwages equation can also be expressed simply as a linear function of the same single index q'_{it}, as in (5.1), we first regress lnwages on the single index of observable individual characteristics implied by the ordered probit model of Table 4.4, and test whether it gives a significantly lower goodness of fit to the model of Table 5.1 above, where the coefficients in the lnwages regression are free to differ from those in the single index model. To test this idea, we run the wages equations on individual character by weighted coefficients from ordered probit model in table 5.2.

We find that the significance and adjusted R^2 of table 5.1 and table 5.2 are almost the same, while some variables' signs are different, which due to the opposite effect of some variables in getting a high level job and receiving a large amount of wages. For example, parents belong to the lowest social background will play a negative role in getting a high level of job, but contribute positively to good salaries[28]. It is not hard to understand that parents belong to the lowest social background cannot provide sufficient information or help on finding a job for their kids resulting in a negative role in getting a high level job. At the same time, graduates from lowest social background may earn higher wages, since they are less likely to be overeducated and also pay little concern on working conditions. Parents with the highest qualification, working in the east, major in physics and literature all have opposite effects on getting a high level job and receiving a large amount of wages. By comparing table 4.4 (also table 4.5) in chapter 4 and table 5.1 in this chapter, we find that

[28] One can observe the results by comparing table 4 or table 5 in chapter 5 and table 1 in chapter 6.

parents with the highest qualification play a negative role in getting a high level job, but contribute positively on wages. This may due to the subjective bias on reporting overeducation and undereducation. Individuals whose parents acquire the highest rank of qualification may have a higher expectation on their current job. They may receive a comparatively high wages and report overeducated to their current jobs in the same time. The high wages for working in the east can be explained by the high consumer price index in the east. Overeducated graduates in physics and literature subjects earn a comparatively high wages can attribute to the subject specialty. Students who are major in physics and literature will learn a lot of theoretic knowledge on these subjects, which may not be used on work. Therefore, graduates from these two subjects may think they are overeducated. All these phenomena suggest job level and wages are not highly correlated, which inspires us to examine whether job level plays a role in wages determination.

In the full version of the pecking order model, wages are determined by individual characteristics X_{it}, as in equation (4.7) in chapter 4, as influenced by their marginal productivity in the underlying production function. If these individual characteristics, and the parameters of the wage function in (4.7), are known to the individual, there is no uncertainty for the individual concerning their graduation wage, save for the stochastic term ω_t that affects v_t and B_t in equations (4.4) and (4.7). This term reflects general macroeconomic uncertainty that affects the overall level of wages at time t in equation (4.7). We will abstract from such macroeconomic uncertainty for the present by assuming that ω_t is a known constant. There will then be a known variation in wages with individual characteristics, irrespective of the level of job which the individual succeeds in securing on graduation. Whether an individual gets a doctoral or a master level job does not affect their wages in the full version of the pecking order model, since it is their individual characteristics, including their education, which determine their productivity.

However, particularly in a country such as China where market forces have only recently been allowed to have greater influence, wage rigidities due to institutional factors may limit the extent to which wages vary with marginal productivity and associated individual factors, such as ability and degree qualifications. Instead, the designated job level may also play an important role in determining wages. In the following Table 5.3, we therefore investigate the effect of adding job level to the list of variables which may influence graduation wages.

Table 5.3 shows the goodness of fit of the lnwage equation increases significantly after adding the job level, particularly for men, with a 3.5 per cent increase in wages overall with job level. Table 5.3 used 1,2,3,4 as the job level variable, which assumes an equal influence as one goes from Diploma to Batchelor to Masters to PhD job level. But there may be a negative return on the PhD job level as a separate variable. Then we add each of these job levels as a separate dummy variable in table 5.4 and find the overall goodness of fit significantly improved compared to the single job level variable in Table 5.3 by F-test. That is to say, the wages include a reward to not only qualifications and other individual characteristics but also the job level.

A related approach here is to examine how far wages vary with the extent of *matching* between an individual's education and the level of the job they secure on graduation. Following Sicherman (1991), we may then investigate two stylized hypotheses: i) the earnings of individuals in occupations that require more schooling than they actually have (i.e. the undereducated) are more than the earnings of workers with the same qualification but in a less demanding job that just requires their existing level of schooling. However, at the same time, undereducated individuals receive less earnings than the workers who have the qualifications needed for the job; ii) the earnings of individuals in occupations that require less schooling than they actually have (i.e. the overeducated) are less than the earnings of workers with the same level of education as themselves who are in a job whose level matches their qualifications. However, overeducated individuals earn more than individuals with similar jobs but with qualifications that just match the level of the job concerned.

Equation (5.2) examined the two hypotheses from mathematics form, where Q_{it}^{r} means the years of schooling required to match the level of the job individual i has at time t, $Q_{it}^{s} \geq 0$ denotes years of surplus schooling which individual i has in excess of the level required by their job, $Q_{it}^{u} \geq 0$ denotes the shortfall in years between the education individual i has and that required for their job, Ψ_{it} represents all the other variables that may affect individuals' wages including personal characteristics and academic achievement[29], and ς_{it} is a disturbance term. $\rho_t\lambda_t$ is the term to overcome the selection bias (Heckman, 1979), and λ_t is the relevant inverse Mills ratio.

$$\log w_{it} = \alpha_{1t} + \alpha_{2t}Q_{it}^{r} + \alpha_{3t}Q_{it}^{s} + \alpha_{4t}Q_{it}^{u} + \alpha_{5t}\Phi\Psi_{it} + \rho_t\lambda_t + \varsigma_{it} \qquad (5.2)$$

We expect the absolute value of the return to Q_{it}^{r} to be larger than that for Q_{it}^{s} and Q_{it}^{u}, with the coefficient of Q_{it}^{s} having a positive sign and the coefficient of Q_{it}^{u} having a negative sign. If this holds, then the above two hypotheses will be confirmed.

The results show the returns to required years of schooling, surplus years of schooling and a shortfall of years of schooling are significant in the directions expected, with the return on required years of schooling greater than that on surplus years of schooling for males, but not for females. Therefore we think two stylized facts hold in China. Comparing to the findings by Rumberger (1987), Verdugo & Verdugo(1989) and Dolton & Vignoles (2000), our regression results to surplus years of schooling are comparatively high, which may explain the large number of overeducated individuals.

[29] Again non-dummy variables were regressed by their natural logarithm due to better fit.

5.3. Educational Choices under Uncertainty and Pecking Order Theory

5.3.1 Individuals optimal educational choices under uncertainty

After analyzing the pecking order theory, individuals' education decisions not only depend on individuals' characteristics, but also economy wide factors (e.g. GDP, growth rate, exchange rate, population growth rate) and job specification. That is to say, individuals' wages were affected by their own characteristics as well as how many graduate jobs the economy can offer and their ranks within the same qualification candidates.

If job level does not play a role in education decisions, individuals' wages will be determined by individuals' characteristics only, which has been analyzed in chapter 3 and chapter 4. However, our above empirical findings suggest that not only are individual characteristics significant determinants of wages on graduation, but so too is the level of the job offered to the individual on graduation. In the next section, we will analyze what the optimal education choice is if job level does affect individuals' wages.

5.3.2 When Job Level Does Affect Wages

In such a case that job level does affect wages, the probabilities of receiving a job offer at each different level should be added into the analysis of optimal education choices under uncertainty. The expected beginning wages in next period (e.g. equation (2.11)) will be

$$EY_{t+1} = p^o(J_{it}, C_{it}) p^g_{nm}(C_{it}) \cdot Y^o_{nm} + [1 - p^o(J_{it}, C_{it})] p^g_{nm}(C_{it}) \cdot Y^m_{nm} + [1 - p^g_{nm}(C_{it})] \cdot Y_{n(m-1)} \tag{5.3}$$

where p^o denotes the probability to be overeducated, which is a function of the sum of individual characteristics C and job specification J. The probability to graduate $p^g(C_{it})$ equals to one given the China's drop out rate in higher education is less than one per cent. Y_{nm}^o denotes the graduation wages in a job for which the individual is overeducated with qualification m and Y_{nm}^m represents the graduation wages for a job that matched their qualifications of level m.

Then, as in Chapter 2, individuals are assumed to maximise their expected utility in order to find their optimal choices for investment in their education. Specifically, we define:

$$V(C_i,t) = \max E[\sum_{r=t}^{T}(1+\theta)^{-rc}U(C_{ir}) \mid I(t)] \qquad (5.4)$$

subject to the life-time earnings $\sum_{r=t}^{T} C_{ir} = \sum_{r=t}^{T} EY_{im}(1+g_m)^r$. c is the inter-temporal discount rate. In the following empirical analysis, we only consider academic choices, since technical education system is undeveloped in China and there is not any data on technical education.

Due to the limitations of our micro dataset, we do not have data on the detailed nature of individual utility functions in (2.6). Here assume utility is determined by peculiar return only, which implies that they will not invest in education unless the expected net present value of their future earnings are larger than not investing. In other words, the net benefits of their investment in additional education is positive. If we examine for simplicity the case where there is a risk simply of the individual being either overeducated or exactly matched in their education with the requirements of the job they secure on graduation, we may write the expected earnings in the next period as:

$$E(Y_m) = p_m^o Y_m^o(1+g_m) + (1-p_m^o)Y_m^m(1+g_m) \qquad (5.5)$$

where g_m is the average wage growth rate for qualification m. The expected present value $E(PV_j)$ of life-time earnings, using the discount rate θ, may then be expressed as:

$$E(PV_m) = \sum_{t=1}^{T_m} (1+\theta)^{-t} [p_m^o Y_m^o (1+g_m)^{t-1} + (1-p_m^o) Y_m^m (1+g_m)^{t-1}] \qquad (5.6)$$

T_m is the number of years of working life that the individual with qualification m has before retirement. Consistent with the analysis in chapter 2 and 3, we assume the retirement age is 60 for both male and female and the total numbers of years of working life are 38, 35 and 32 respectively for bachelor, master and doctorate degree.

The expected net present value $E(NPV_m)$ of life-time earnings is given by:

$$E(NPV_m) = E(PV_m) - C_m \qquad (5.7)$$

where C_m is the total costs of acquiring schooling to level m, including tuition fees. The increase in the expected net present value of the future earnings that result from investing in achieving a level m qualification, rather than a level m-1 qualification, net of the cost I_m of this additional educational investment is then given by:

$$\Delta E(NPV_m) = E(PV_m) - E(PV_{m-1}) - I_m \text{ for } I_m \equiv C_m - C_{m-1} \qquad (5.8)$$

In view of the importance of the increase in the expected net present value of the future earnings that results from their investment in additional education being positive if risk-averse individuals are to invest in such additional education, it is instructive to examine the threshold case of $\Delta E(NPV_m) = 0$ in (5.8). In this threshold case, there exists a positive trade-off relationship between the probability p_m of an individual with qualifications m being overeducated (with the negative consequences this has for their graduation wages once job level affects their graduation wages) and the growth rate g in their future earnings after graduation (that *ceteris paribus* will boost the expected net present value of their future life-time earnings), for a given value of p^0_{m-1} and other relevant variables. The critical probability $p^c_m(g, Y^o_m, Y^m_m, p^o_{m-1}, Y^o_{m-1}, Y^m_{j-1})$ is the value of the probability p^o_m of being overeducated with qualifications m at which of $\Delta E(NPV_m) = 0$ in (5.8) for a given value of g, $Y^o_m, Y^m_m, Y^o_{m-1}, Y^m_{m-1}$ and p^0_{m-1}. This can to some extent provide a guideline for individual decision-making on the extent of the risks involved in investing in being educated to level m. The following figure draws the relationship between average wages growth rate g and the critical probability p^o_m for each degree qualification level, given the mean levels of p^0_{m-1} found in our survey and for values of Y^o_m, Y^m_m, Y^o_{m-1} and Y^m_{m-1} equal to their mean level found in our survey.

In the following figures, we map out the trade-off between p^o_m and g from setting $\Delta E(NPV_m) = 0$ in (5.8), for the mean levels of p^0_{m-1} found in our survey and for values of Y^o_m, Y^m_m, Y^o_{m-1} and Y^m_{m-1} equal to their mean level found in our survey.

Figure 5.1 The relationship between average wages growth rate and the critical probability

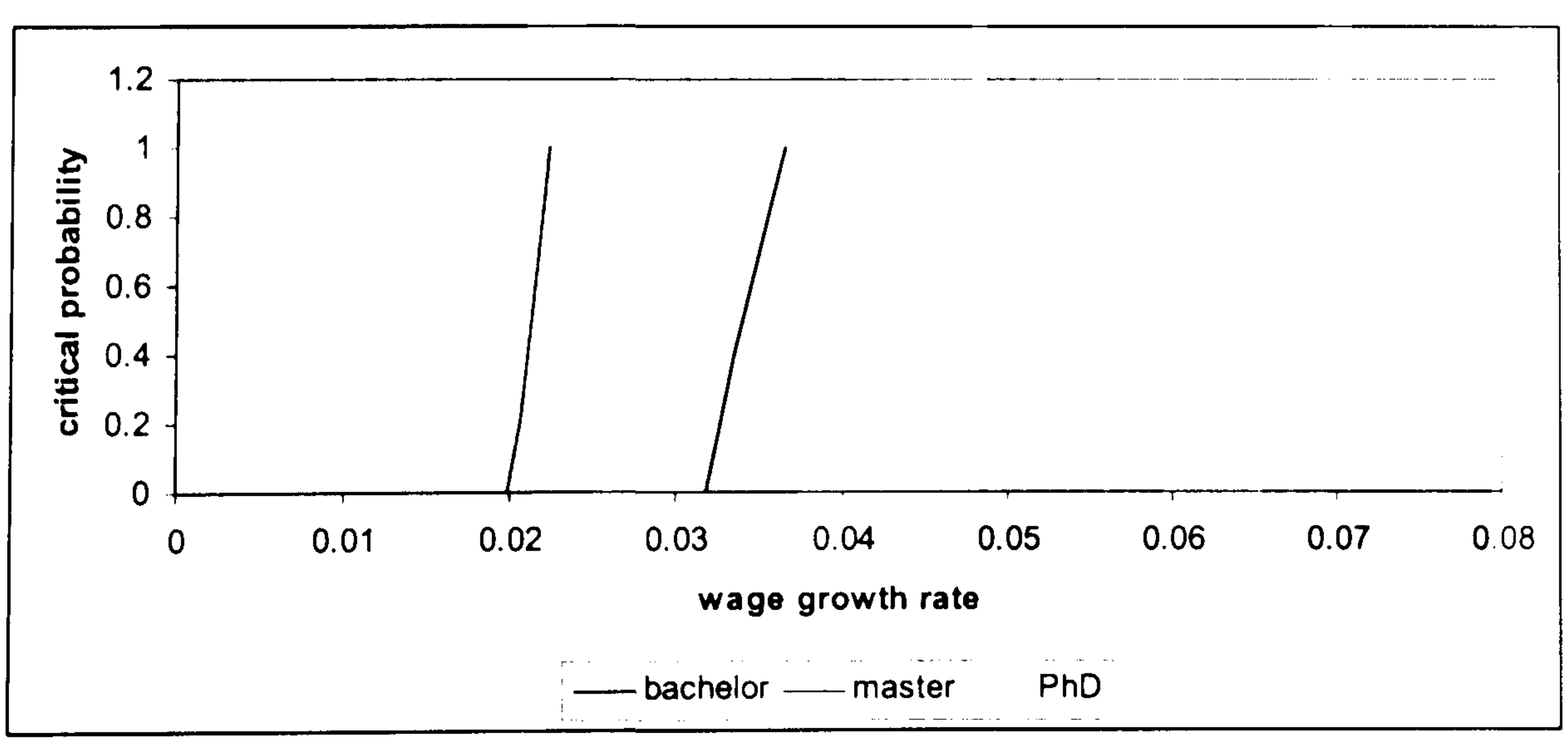


Figure 5.1 shows the range of overeducation rate and wages growth rate when NPV=0. Since all these lines are quite steep, we list the range of

average wages growth rate g_m in table 5.6 when the overeducation probability p^o varies from 0 to 1. The results show investing in Masters education appears to be the least risky. As long as the wages growth rate is larger than 2.2 per cent per year, individuals will never be overeducated. This result is consistent with the findings in table 5.1 that the return to master degree is the highest, so does the overeducation rate for master degree.

Table 5.6 The contrast of critical growth rate with empirical evidence on growth rate

	bachelor	master	PhD
g	0.0308-0.0352	0.0196-0.0221	0.0650- 0.0675
g_0	0.032	0.045	0.054

We can also get an idea of the current rate of growth, g_0, of earnings of graduates as they progress through their careers from a Chinese recruiting company[30] that selected a one per cent sample of the workforce in 30 Chinese cities for different levels of degree qualifications, as shown in Table 5.6. This suggests that the rate of growth of individual earnings after graduation does depend upon the qualification level m. Contrast two growth rates (g and g_0), we can again get the conclusion that investing in master degree is very profitable.

5.3.3 The impact of changes over time

One important feature of investment in university education is the significant lag that exists between the date when a decision is made by an individual to invest in seeking to acquire an additional university qualification and the date at which they enter the labour market with such a qualification. Another important feature of the current situation in China is the very high rates of expansion of the number of university graduates passing through the university system in recent and future years. These

[30] www.chinahr.com

two features mean that an individual who is currently considering whether or not to invest in additional university education needs to take into account in their calculations of the expected net present value of such an additional investment under uncertainty not simply the current graduation wage levels and current probability of being overeducated or securing a job at a given level, but instead their projected future levels when the individual graduates with the additional qualification. In addition, a rational individual making these important investments in their human capital under uncertainty needs to make predictions about the future growth rate in their earnings after graduation, since this is also an important element in the calculation of the expected present value of the enhanced future earnings which their additional qualification will yield.

While the above calculation of the critical probability $p_m^c(g, Y_m^o, Y_m^m, p_{m-1}^o, Y_{m-1}^o, Y_{j-1}^m)$ of being overeducated on graduation may shed some light *ceteris paribus* on the calculation of whether or not the change in the expected net present value of future earnings as a result of the additional qualification is positive, this calculation does not itself take into account the changes which may take place in the levels of the graduation wages Y_m^o and Y_m^m, and of the probability, p_{m-1}^0, of being overeducated even if the additional qualification is not taken, by the time the individual does graduate. The extent to which individuals can, and do, anticipate these changes will itself form an important part of any market adjustment mechanism by which any future over-supply of university graduates results in a potential fall in graduation wages and rise in the probability of overeducation that feed back into the investment decisions of those seeking additional university qualifications.

Moreover, since the change in the net present value of future earnings in equations (5.6) and (5.8) above depends upon a comparison of future earnings with and without the additional qualification, a rational investor in human capital must here seek to project the extent of any changes over time not just in the graduation wages Y_m^o and Y_m^m, but also in Y_{m-1}^m and

Y^{m}_{m-1}, that the individual may receive if they do not undertake the additional education and are not currently in employment. Growing rates of overeducation, and falling wages on graduation, of graduates with lower qualifications than those being considered by the individual may indeed themselves act as a spur to encourage the individual to invest in a further qualification in order to boost their competitive position in the labour market, in the way the pecking order model suggests, and improve their wages relative to being without the qualification, even if graduate wages are falling overall.

The above pecking order model itself provides a framework in which graduation wages and the probability of receiving a job offer at a given level on graduation may be related to the underlying parameters of the distribution of individual characteristics, including their educational qualifications, and to the parameters of the distribution of job characteristics in the economy. Each of these parameters may indeed change over time with increases in the supply of graduates with different educational qualifications and with the rate of economic growth and other macro-economic factors that affect the distribution of available job characteristics in the economy at large. These parameters are the mean (θ_{Ct}) and the standard deviation (σ_{Ct}) of the distribution of individual characteristics, and the mean (θ_{mt}) and the standard deviation (σ_{mt}) of the distribution of job characteristics, all of which may change over time.

For the reason of simplicity, we will assume that the ratio (σ_{Ct}/σ_{mt}) between the respective standard deviations for the distributions of the indices of individual characteristics and job specifications remains the same over time, with the two distributions remaining lognormal over time. Equations (4.10) and (4.11) then imply that the change in the overall probability of receiving a job offer at level ℓ for an individual with a given set, x_{it}, of the first n-1 characteristics is given by:

$$dp_{lt}(x_{it})/dt = \xi(x_{it}a - q_t^{ol})((\sigma_{Ct}/\sigma_{mt})(d\theta_{mt}/dt) - (d\theta_{Ct}/dt)) \qquad (5.9)$$

where ξ is the standardised normal density function. The condition

$$(d\theta_{Ct} / dt) > (d\theta_{mt} / dt)(\sigma_{Ct} / \sigma_{mt}) \tag{5.10}$$

is then necessary and sufficient for a decrease over time in the overall probability $p_{\ell t}(x_{it})$ of the offer of employment at level $\ell > 0$ or above for an individual with characteristics given by X_{it} . It should be noted here that the overall probability $p_{\ell t}(x_{it})$ is that the first nth individual characteristic can get a level l job. ε_i is a stochastic latent variable, such as individual enthusiasm, that cannot be directly measured but which nevertheless affects the individual's position in the pecking order. It is important to note here that if the individual does have further information on the value of this variable for their own individual case within the overall distribution across all individuals, they may be able to further refine their estimate of their own probability of securing a job at a given level beyond that implied by the overall probability $p_{lt}(x_{it})$.

Macro-economic factors, such as growth in GDP, are likely to impact not just upon the value of output V_{imt} produced by any job j with given job specifications in equation (4.1) of chapter 4 in terms of the tasks which the job involves. They are also likely to change the distribution of job specifications across jobs in the economy. In particular, economic growth may be associated with a change in the average level of the complexity of the tasks which jobs in the economy involve, and in the associated desirable degree of education and training for those who carry out these tasks. If we assume that

$$\theta_{mt} \equiv E(\ln J_{mt}) = \sum_{r=1}^{s} \alpha_r \ln M_{rt} \tag{5.11}$$

where each α_r is a constant, it follows from equations (4.3) and (4.9) of chapter 4 that condition (5.11) is equivalent to

$$\sum_{h=1}^{n} a_h (d\bar{x}_{ht} / dt) > (\sigma_{Ct} / \sigma_{mt}) \sum_{r=1}^{s} \alpha_r \kappa_{rt} \tag{5.12}$$

where $\bar{x}_{ht}$ is the mean level of x_{iht} at time t across the population of individuals and $\kappa_{\tau t}$ is the proportionate growth rate at time t in the macroeconomic factor $M_{\tau t}$. The change in wages over time for an individual with any given vector of individual characteristics is given by:

$$dw_{it}(X_{it})/dt = (w_{it}(X_{it}) - \zeta)(\sum_{\tau=1}^{s}(c_\tau + \alpha_\tau)\kappa_{\tau t} - (\sigma_{mt}/\sigma_{Ct})\sum_{h=1}^{n} a_h(d\bar{x}_{ht}/dt)) \tag{5.13}$$

with

$$\sum_{h=1}^{n} a_h(d\bar{x}_{ht}/dt) < (\sigma_{Ct}/\sigma_{mt})\sum_{\tau=1}^{s}(c_\tau + \alpha_\tau)\kappa_{\tau t} \tag{5.14}$$

being a necessary and sufficient condition for $dw_{it}(X_{it})/dt > 0$ in (4.19). Clearly (5.12) and (5.14) occurring together requires that the growth rates in the macroeconomic factors have a sufficiently large impact on each V_{imt} in equation (4.1) via the coefficients c_τ to more than offset the imbalance in the growth rates in (5.12). However, if condition (5.14) does not hold and each $c_\tau > 0$ and $\kappa_{\tau t} > 0$, the growth in the population mean values of the individual characteristics will be sufficiently large overall to imply a decline both in the probability of any individual i receiving an offer of employment at level $\ell > 0$ or above and in their wage rate.

Though we do not have panel micro data to examine the impact of changes over time, Macro statistics do show the consistency with our theoretic prediction. The GDP growth rate $\kappa_{\tau t}$ in equation (5.14) is around 10 percent in 2003, nevertheless the increase of new graduates in the labour market is more than 40 percent implying equation (5.13) and (5.14) does not hold in China. According to our theoretical prediction, the probability p_{lt} to get an offer of employment at level 1 will decrease significantly and the average wages will decline accordingly. Our empirical survey took place in 2003, when the first group of students graduated after higher education expansion. Until June 2003, there is about 30 percent graduates did not find a satisfactory job with an average wage of 1351 yuan for undergraduates. The unemployment rate for the university graduates increases to 50 percent in June 2005 with an average wage of 1000 yuan for undergraduates.

Table 5.1A The determinants of wages on graduation

	Including schooling			Including each qualification level		
	comb	male	female	comb	male	female
gender	-0. 033* 0. 018			-0. 048 0. 013		
sch	0. 095*** 0. 023	0. 079*** 0. 027	0. 099** 0. 048			
ba				0. 037*** 0. 005	0. 035 0. 058	0. 149* 0. 081
ma				0. 171*** 0. 011	0. 171*** 0. 026	0. 223*** 0. 042
phd				0. 150*** 0. 020	0. 169*** 0. 034	0. 152*** 0. 064
Pcar1	0. 043** 0. 019	0. 047* 0. 026	0. 048 0. 035	0. 029* 0. 017	0. 004 0. 021	0. 074*** 0. 027
Pcar2	0. 018 0. 019	-0. 003 0. 024	0. 058* 0. 034	0. 031* 0. 018	-0. 008 0. 022	0. 106*** 0. 029
Pcar3	0. 092*** 0. 023	0. 077** 0. 032	0. 125*** 0. 043	0. 093*** 0. 021	0. 050* 0. 027	0. 182*** 0. 035
Pqu1	-0. 018 0. 037	0. 003 0. 047	-0. 086 0. 070	-0. 022 0. 017	-0. 009 0. 020	-0. 051* 0. 030
Pqu2	0. 002 0. 039	0. 021 0. 052	-0. 078 0. 067	-0. 009 0. 017	0. 024 0. 022	-0. 064** 0. 033
Pqu3	-0. 087** 0. 041	-0. 068 0. 050	-0. 192*** 0. 083	0. 053** 0. 024	0. 090*** 0. 030	-0. 022 0. 042
Pqu4	-0. 103*** 0. 040	-0. 091** 0. 047	-0. 194** 0. 082	0. 058** 0. 025	0. 081*** 0. 031	0. 001 0. 043
Pqu5	-0. 082** 0. 039	-0. 085** 0. 048	-0. 130* 0. 080	0. 150*** 0. 051	0. 164*** 0. 065	0. 114 0. 085

Table 5.1A The determinants of wages on graduation--continued

cadre	0.020 0.017	-0.003 0.022	0.039 0.033	0.091*** 0.021	0.026 0.030	0.159*** 0.045
Reg1	0.058*** 0.017	0.048** 0.024	0.075*** 0.030	0.033 0.021	0.047* 0.025	0.020 0.037
Reg2	0.038* 0.022	0.034 0.030	0.036 0.045	0.034** 0.016	0.050*** 0.019	0.002 0.029
Reg3	0.090*** 0.020	0.096*** 0.028	0.093*** 0.035	0.080*** 0.017	0.080*** 0.021	0.068** 0.030
pmem	0.003 0.017	0.023 0.021	-0.056* 0.031	0.004 0.014	0.006 0.018	-0.005 0.023
rank1	-0.360 0.268	-0.445 0.322	-0.362 0.264	-0.080* 0.046	-0.186*** 0.069	-0.010 0.092
rank2	0.144*** 0.054	0.203** 0.095	0.161*** 0.059	0.151** 0.060	0.013 0.091	0.226* 0.126
grad1	-0.010 0.018	-0.021 0.023	-0.024 0.031	0.002 0.027	0.038 0.031	-0.109* 0.064
grad2	0.022 0.022	0.016 0.025	-0.017 0.053	-0.038 0.028	0.007 0.034	-0.154** 0.066
grad3	0.023 0.029	-0.001 0.032	0.127* 0.073	-0.031 0.031	0.022 0.039	-0.148** 0.070
Eng1	-0.220*** 0.022	-0.277*** 0.042	-0.151*** 0.036	0.024 0.017	0.081*** 0.023	-0.054 0.035
Eng2	0.159*** 0.040	0.207*** 0.060	0.171*** 0.052	0.071*** 0.024	0.160*** 0.031	-0.030 0.048
Wor1	-0.116*** 0.017	-0.099*** 0.021	-0.140*** 0.028	0.072 0.061	0.095 0.076	-0.002 0.108
Wor2	-0.095*** 0.028	-0.096*** 0.036	-0.101** 0.043	0.028 0.058	0.051 0.070	-0.017 0.104
Wor3	0.192*** 0.059	0.213*** 0.070	0.139 0.107	0.120** 0.057	0.135** 0.069	0.076 0.105

Table 5.1A The determinants of wages on graduation—continued

east	0.009	0.012	0.065*	0.281***	0.219***	0.392***
	0.023	0.028	0.040	0.013	0.017	0.023
mid	-0.062	-0.048	-0.089	0.151***	0.164***	0.115***
	0.043	0.053	0.074	0.021	0.026	0.038
Gov	-0.065***	-0.085***	-0.014	-0.074***	-0.111***	0.005
	0.022	0.026	0.039	0.022	0.026	0.037
Stat	-0.027*	-0.018	-0.053*	0.000	-0.015	0.026
	0.016	0.020	0.030	0.017	0.021	0.029
Jon	0.171***	0.170***	0.168***	0.133***	0.130***	0.137***
	0.024	0.029	0.039	0.023	0.029	0.037
edui	-0.048***	-0.063***	-0.026	-0.030	-0.073***	0.033
	0.019	0.024	0.030	0.019	0.025	0.029
Bio	-0.064	-0.024	-0.117	-0.043	-0.062	-0.049
	0.064	0.101	0.084	0.052	0.071	0.078
Math	-0.025	0.110*	-0.187***	0.013	0.074	-0.026
	0.037	0.061	0.065	0.035	0.051	0.049
Phy	0.029***	0.013	0.075	0.083**	0.100**	0.055
	0.008	0.045	0.090	0.034	0.041	0.065
Elec	0.055***	0.070***	-0.029	0.112***	0.126***	0.088**
	0.020	0.025	0.048	0.018	0.020	0.036
Cons	-0.177***	-0.186***	-0.335**	-0.139***	-0.127***	-0.160***
	0.053	0.055	0.169	0.029	0.033	0.060
Soc	0.052	0.116	-0.043	-0.021	0.015	-0.125
	0.093	0.129	0.150	0.084	0.109	0.131
Pol	0.002	0.000	-0.010	0.027	0.063	0.005
	0.054	0.074	0.087	0.050	0.070	0.073

Table 5.1A The determinants of wages on graduation–continued

Lan	0. 119***	0. 154***	0. 012	0. 160***	0. 177***	0. 110***
	0. 036	0. 056	0. 076	0. 028	0. 054	0. 037
Lit	0. 042**	0. 024	0. 013	0. 079***	0. 060	0. 088**
	0. 018	0. 043	0. 047	0. 027	0. 037	0. 041
Art	-0. 007	0. 089	-0. 159*	-0. 031	0. 058	-0. 150*
	0. 055	0. 074	0. 087	0. 052	0. 070	0. 078
Chem	-0. 024	0. 011	-0. 053	-0. 005	0. 000	0. 018
	0. 054	0. 064	0. 111	0. 050	0. 058	0. 096
econ	-0. 037	-0. 067	-0. 071	0. 030	0. 024	0. 014
	0. 035	0. 047	0. 065	0. 031	0. 039	0. 050
Law	-0. 018	-0. 025	-0. 027	-0. 022	-0. 047	-0. 013
	0. 045	0. 068	0. 067	0. 040	0. 056	0. 058
med	-0. 179***	-0. 123**	-0. 287***	-0. 060	-0. 015	-0. 108**
	0. 039	0. 058	0. 068	0. 038	0. 054	0. 054
man	-0. 030	0. 014	-0. 207***	0. 026	0. 053**	-0. 024
	0. 035	0. 038	0. 085	0. 019	0. 024	0. 034
cons	6. 341***	6. 612***	7. 463***	6. 1490***	6. 725***	5. 900***
	0. 615	0. 788	1. 184	0. 244	0. 381	0. 580
Mill	-0. 126	-0. 116	-0. 151	-0. 088**	0. 012	-0. 127
	0. 151	0. 203	0. 264	0. 038	0. 062	0. 080
Obs	16005	9281	6724	16005	9281	6724

Table 5.1B First step probit estimation of table 5.1

Var	Ordered probit model			Ordered probit model		
	comb	**male**	**female**	**comb**	**male**	**female**
gender	-0.125***			-0.124***		
	0.024			0.024		
lnsch	0.753***	0.696***	0.819***			
	0.048	0.062	0.077			
lnba				0.408***	0.362***	0.447***
				0.028	0.036	0.044
lnma				0.454***	0.441***	0.473***
				0.038	0.050	0.060
lnphd				0.429***	0.454***	0.373***
				0.057	0.068	0.108
Pcar1	-0.017	-0.078	0.063	-0.024	-0.080	0.053
	0.034	0.044	0.053	0.034	0.044	0.053
Pcar2	0.003	-0.033	0.049	-0.000	-0.036	0.047
	0.033	0.044	0.052	0.033	0.044	0.052
Pcar3	0.022	0.080	0.067	0.027	0.081	0.059
	0.042	0.055	0.065	0.042	0.055	0.065
Pqu1	0.077	0.047	0.121	0.084	0.051	0.123
	0.064	0.083	0.101	0.064	0.083	0.101
Pqu2	0.129**	0.174	0.054	0.136**	0.184**	0.049
	0.064	0.084	0.102	0.064	0.083	0.101
Pqu3	0.235***	0.196***	0.277***	0.241***	0.205***	0.271***
	0.059	0.076	0.094	0.059	0.076	0.094
Pqu4	0.201***	0.146***	0.259***	0.211***	0.157***	0.258***
	0.060	0.077	0.096	0.059	0.076	0.096
Pqu5	0.181***	0.139***	0.222**	0.186***	0.145**	0.217**
	0.062	0.079	0.102	0.062	0.078	0.101
cadre	0.105***	0.102***	0.118***	0.105***	0.105***	0.115***
	0.023	0.030	0.037	0.023	0.030	0.036
Reg1	0.010	0.069	-0.058	0.009	0.070	-0.059
	0.030	0.040	0.045	0.030	0.040	0.045
Reg2	-0.015	0.053	-0.104**	-0.014	0.054	-0.100
	0.040	0.053	0.062	0.040	0.053	0.062
Reg3	0.059*	0.107*	0.011	0.055*	0.108	0.018
	0.033	0.043	0.054	0.033	0.043	0.054
pmem	0.080	0.060	0.096**	0.084***	0.059	0.104**
	0.028	0.037	0.043	0.028	0.038	0.043
rank1	0.509***	0.722***	0.209***	0.484***	0.707***	0.185***
	0.036	0.048	0.056	0.038	0.051	0.059
rank2	0.342***	0.504***	0.155**	0.323***	0.488***	0.142**
	0.042	0.056	0.067	0.043	0.057	0.068
grad1	0.103***	0.102***	0.103***	0.106***	0.107***	0.103***
	0.025	0.034	0.039	0.025	0.034	0.039
grad2	0.120***	0.074**	0.209***	0.118***	0.075*	0.206***
	0.033	0.041	0.057	0.033	0.041	0.057

Table 5.1B The first step probit estimation of table 5.1–continued

grad3	0. 078*	0. 027	0. 135	0. 075	0. 031	0. 132
	0. 051	0. 059	0. 112	0. 051	0. 059	0. 112
Engl	0. 130***	0. 257***	-0. 059	0. 126***	0. 253***	-0. 059
	0. 034	0. 045	0. 054	0. 034	0. 045	0. 054
Eng2	0. 342***	0. 408***	0. 227***	0. 324***	0. 400***	0. 202***
	0. 030	0. 039	0. 047	0. 030	0. 040	0. 048
Bio	-0. 379***	-0. 524***	-0. 133	-0. 410***	-0. 548***	-0. 181
	0. 084	0. 118	0. 121	0. 084	0. 118	0. 122
Math	-0. 063***	-0. 191***	0. 151*	-0. 078	-0. 193**	0. 115
	0. 062	0. 090	0. 088	0. 062	0. 090	0. 089
Phy	0. 095***	0. 034	0. 234*	0. 093	0. 051	0. 189
	0. 066	0. 079	0. 124	0. 066	0. 078	0. 124
Elec	0. 076*	0. 090**	0. 114*	0. 061*	0. 085**	0. 084
	0. 033	0. 040	0. 064	0. 034	0. 040	0. 065
Cons	0. 463***	0. 337***	0. 842***	0. 445***	0. 333***	0. 794***
	0. 068	0. 078	0. 138	0. 067	0. 078	0. 136
Soc	-0. 215	-0. 299	-0. 093	-0. 230	-0. 300	-0. 138
	0. 148	0. 199	0. 222	0. 147	0. 199	0. 222
Pol	-0. 168*	-0. 089	-0. 185	-0. 194**	-0. 104	-0. 229*
	0. 088	0. 134	0. 120	0. 088	0. 134	0. 121
Lan	0. 188***	0. 006	0. 307***	0. 182***	0. 024	0. 282***
	0. 052	0. 098	0. 068	0. 051	0. 097	0. 069
Lit	0. 020	0. 100	0. 001	0. 002	0. 087	-0. 030
	0. 049	0. 070	0. 072	0. 049	0. 070	0. 072
Art	-0. 094	-0. 054	-0. 057	-0. 093	-0. 039	-0. 085
	0. 081	0. 115	0. 118	0. 081	0. 114	0. 118
Chem	-0. 158*	-0. 127*	-0. 176	-0. 178**	-0. 140	-0. 206
	0. 086	0. 105	0. 151	0. 086	0. 105	0. 150
econ	0. 154***	0. 186***	0. 159*	0. 139***	0. 183***	0. 123
	0. 055	0. 074	0. 086	0. 055	0. 074	0. 086
Law	-0. 147**	-0. 208***	-0. 082	-0. 171***	-0. 217**	-0. 130
	0. 071	0. 101	0. 104	0. 072	0. 101	0. 105
med	-0. 036	-0. 097***	0. 090	-0. 063	-0. 107	0. 038
	0. 073	0. 107	0. 104	0. 074	0. 107	0. 105
man	0. 274***	0. 221***	0. 375***	0. 261***	0. 216***	0. 345***
	0. 037	0. 048	0. 062	0. 037	0. 048	0. 063
const	-2. 157***	-2. 260***	-2. 191***	-1. 628***	-1. 781***	-1. 591***
	0. 091	0. 118	0. 145	0. 077	0. 098	0. 121
Loglihood	-8729. 7119	-5395. 6448	-3559. 0403	-9071. 9474	-5430. 9161	-3579. 6953
Pseudo R2	0. 2038	0. 1152	0. 0969	0. 1091	0. 1153	0. 0973
Obs	16005	9235	6770	16005	9235	6770

Table 5.2 Weighted wages equation by single index function

	combined		male		female	
	Coeff.	Std. Err	Coeff.	Std. Err.	Coeff.	Std. Err.
gender	-44. 685***	5. 467				
sch	0. 175***	0. 006	0. 782***	0. 197	0. 948***	0. 336
Pcar1	-0. 145	0. 092	-0. 018	0. 138	-0. 289***	0. 114
Pcar2	1. 502*	0. 891	-0. 273	0. 575	-14. 867***	4. 161
Pcar3	-2. 099***	0. 524	-1. 125	0. 758	-2. 534***	0. 492
Pqu1	-0. 631	0. 601	0. 490	2. 622	-0. 564	0. 369
Pqu2	-0. 052	0. 220	0. 473	0. 373	-0. 449*	0. 249
Pqu3	0. 579**	0. 244	1. 028***	0. 314	-0. 217	0. 403
Pqu4	-3. 850***	1. 428	-1. 418***	0. 473	0. 122	0. 652
Pqu5	-0. 268***	0. 090	-0. 416***	0. 154	-0. 132	0. 105
cadre	1. 526***	0. 205	1. 197***	0. 264	1. 353***	0. 204
Reg1	-0. 283*	0. 173	-0. 474*	0. 259	-0. 131	0. 224
Reg2	-0. 460**	0. 217	-0. 351**	0. 143	0. 068	0. 427
Reg3	-1. 016***	0. 235	-0. 905***	0. 261	-5. 495**	2. 602
pmem	0. 044	0. 056	0. 083	0. 079	-0. 002	0. 074
rank1	-0. 102	0. 063	-0. 153*	0. 086	0. 016	0. 088
rank2	0. 319***	0. 057	0. 313***	0. 080	0. 332***	0. 077
grad1	0. 015	0. 273	-0. 174	0. 302	0. 495*	0. 290
grad2	0. 249*	0. 143	0. 194	0. 195	0. 355**	0. 144
grad3	0. 248*	0. 152	0. 192	0. 226	0. 348**	0. 143
Eng1	0. 118	0. 147	0. 336***	0. 122	-34. 521***	13. 273
Eng2	1. 100***	0. 349	1. 517***	0. 312	1. 730	1. 382
Wor1	-1. 219	1. 044	-1. 159	0. 875	0. 042	1. 337
Wor2	0. 166	0. 345	0. 252	0. 338	0. 872	5. 306

Table 5.2 Weighted wages equation by single index function—continued

Wor3	0. 283**	0. 133	0. 281**	0. 139	0. 379	0. 514
east	-9. 112***	0. 461	21. 969***	1. 792	-3. 173**	0. 190
mid	4. 783***	0. 697	-2. 928***	0. 482	0. 392**	0. 128
Gov	-3. 382***	0. 996	1. 079***	0. 267	0. 006	0. 113
Stat	0. 033	0. 364	0. 279	0. 355	-0. 613	0. 694
Jon	1. 128***	0. 198	1. 157***	0. 259	1. 010***	0. 284
edui	-0. 104*	0. 061	-0. 245***	0. 084	0. 093	0. 088
Bio	0. 163	0. 174	0. 175	0. 189	0. 271	0. 387
Math	-0. 107	0. 459	-0. 239	0. 194	-0. 466	0. 793
Phy	-0. 604**	0. 266	-1. 552**	0. 684	-0. 183	0. 248
Elec	0. 574***	0. 099	0. 601***	0. 105	1. 192**	0. 552
Cons	-1. 782***	0. 354	-2. 302***	0. 573	-0. 720***	0. 257
Soc	-0. 111	0. 314	-1. 183	12. 157	-0. 151	0. 148
Pol	-0. 152	0. 456	-0. 176	0. 241	-0. 334	5. 223
Lan	0. 854***	0. 159	-2. 889***	0. 892	0. 385***	0. 140
Lit	-8. 043***	3. 092	-9. 044	6. 346	3. 840*	2. 019
Art	-0. 196	0. 327	0. 369	0. 385	-0. 937**	0. 478
Chem	-0. 272	1. 560	0. 141	1. 331	0. 049	0. 234
econ	-0. 705	1. 135	-0. 431	1. 116	0. 138	3. 126
Law	-0. 315	0. 412	-0. 307	0. 291	0. 392	1. 100
med	-0. 422*	0. 225	-0. 112	0. 239	-0. 834**	0. 380
man	1. 079	1. 221	0. 663*	0. 355	0. 716	0. 750
Mill	-0. 126***	0. 014	-0. 116***	0. 018	-0. 151***	0. 026
cons	5. 268***	0. 149	5. 322***	0. 181	5. 011***	0. 305
Adj. R2	0.3853		0.3447		0. 3920	
Obs.	4632		2999		1633	

Table 5.3 Wages determinants after adding job level

	combined		male		female	
	Coeff.	Std. Err	Coeff.	Std. Err.	Coeff.	Std. Err.
gender	-0. 129***	0. 016				
sch	0. 097***	0. 007	0. 080***	0. 008	0. 101***	0. 006
Joblev	0. 035***	0. 008	0. 044***	0. 010	0. 023	0. 015
Pcar1	0. 021*	0. 012	0. 005	0. 023	0. 071***	0. 023
Pcar2	0. 028*	0. 018	-0. 012	0. 022	0. 095***	0. 029
Pcar3	0. 085***	0. 019	0. 041	0. 029	0. 106***	0. 031
Pqu1	-0. 021	0. 017	-0. 003	0. 021	-0. 047	0. 031
Pqu2	-0. 005	0. 018	0. 028	0. 022	-0. 059*	0. 033
Pqu3	0. 052**	0. 021	0. 094***	0. 030	-0. 025	0. 042
Pqu4	0. 068***	0. 025	0. 179***	0. 032	0. 006	0. 043
Pqu5	0. 164***	0. 053	0. 189***	0. 066	0. 111	0. 084
cadre	0. 104***	0. 013	0. 030***	0. 008	0. 166***	0. 025
Reg1	0. 035*	0. 021	0. 055**	0. 026	0. 023	0. 038
Reg2	0. 033**	0. 015	0. 081**	0. 020	0. 004	0. 030
Reg3	0. 076***	0. 013	0. 079***	0. 022	0. 067**	0. 031
pmem	0. 006	0. 015	0. 013	0. 019	-0. 004	0. 024
rank1	-0. 058**	0. 029	-0. 087**	0. 040	0. 001	0. 051
rank2	0. 183***	0. 034	0. 161***	0. 046	0. 253***	0. 061
grad1	0. 000	0. 027	0. 020	0. 030	-0. 109*	0. 063
grad2	-0. 042	0. 026	-0. 027	0. 030	-0. 141**	0. 060
grad3	0. 011	0. 007	-0. 024	0. 031	0. 142**	0. 061
Eng1	0. 010	0. 015	0. 066***	0. 018	-0. 069***	0. 027
Eng2	0. 053***	0. 019	0. 163***	0. 026	-0. 041	0. 033
Wor1	0. 076	0. 063	0. 055	0. 077	0. 001	0. 108
Wor2	0. 026	0. 059	0. 046	0. 071	-0. 014	0. 106
Wo3	0. 110**	0. 059	0. 130*	0. 070	0. 074	0. 105

Table 5.3 Wages determinants after adding job level--continued

east	0. 279***	0. 010	0. 216***	0. 017	0. 392***	0. 019
mid	0. 141***	0. 017	0. 159***	0. 024	0. 104***	0. 031
Gov	-0. 072***	0. 021	-0. 109***	0. 027	-0. 002	0. 037
Stat	-0. 002	0. 016	-0. 015	0. 021	0. 027	0. 029
Jon	0. 131***	0. 021	0. 130***	0. 030	0. 131***	0. 030
edui	-0. 042**	0. 022	-0. 072***	0. 025	0. 027	0. 030
Bio	-0. 043	0. 0539	-0. 057	0. 072	-0. 057	0. 079
Math	0. 013	0. 035	0. 076	0. 051	-0. 029	0. 049
Phy	0. 081**	0. 032	0. 098**	0. 048	0. 051	0. 066
Elec	0. 008***	0. 018	0. 115***	0. 020	0. 072**	0. 032
Cons	-0. 141***	0. 021	-0. 132***	0. 031	-0. 177***	0. 060
Soc	-0. 036	0. 084	0. 010	0. 109	-0. 144	0. 132
Pol	0. 019	0. 041	0. 063	0. 070	-0. 002	0. 073
Lan	0. 132***	0. 009	0. 183***	0. 052	0. 009***	0. 002
Lit	0. 060***	0. 018	0. 058	0. 038	0. 071**	0. 032
Art	-0. 028	0. 043	0. 064	0. 071	-0. 151**	0. 060
Chem	-0. 010	0. 050	-0. 006	0. 058	0. 016	0. 096
econ	0. 023	0. 031	0. 021	0. 039	0. 003	0. 050
Law	-0. 020	0. 040	-0. 062	0. 056	-0. 018	0. 058
med	-0. 053**	0. 038	-0. 030	0. 054	-0. 119**	0. 051
man	0. 018	0. 019	0. 044*	0. 024	-0. 030	0. 034
Mill	-0. 119***	0. 014	-0. 108***	0. 018	-0. 146***	0. 026
cons	5. 313***	0. 150	5. 379***	0. 181	5. 050***	0. 306
Adj. R2	0.3861		0.3488		0.3925	
Obs.	4632		2999		1633	

Table 5.4 Wages determinants after adding each job level

	combined		male		female	
	Coeff.	Std. Err	Coeff.	Std. Err.	Coeff.	Std. Err.
gender	-0. 048***	0. 013				
sch	0. 102***	0. 007	0. 094***	0. 006	0. 104***	0. 008
Joblev1	0. 074***	0. 027	0. 077**	0. 031	0. 060	0. 048
Joblev2	0. 108***	0. 026	0. 133***	0. 032	0. 076*	0. 046
Joblev3	0. 158***	0. 030	0. 173***	0. 036	0. 134***	0. 048
Joblev4	0. 016	0. 030	0. 066	0. 059	-0. 051	0. 099
Pcar1	0. 023*	0. 018	0. 005	0. 023	0. 072***	0. 029
Pcar2	0. 025*	0. 018	-0. 010	0. 022	0. 103***	0. 029
Pcar3	0. 090***	0. 022	0. 044	0. 029	0. 181***	0. 036
Pqu1	-0. 020	0. 017	-0. 004	0. 021	-0. 058*	0. 031
Pqu2	-0. 006	0. 018	0. 027	0. 022	-0. 066**	0. 033
Pqu3	0. 053**	0. 025	0. 094***	0. 030	-0. 031	0. 042
Pqu4	0. 069***	0. 026	0. 097***	0. 032	0. 006	0. 043
Pqu5	0. 179***	0. 052	0. 195***	0. 066	0. 140	0. 086
cadre	0. 113***	0. 015	0. 085***	0. 018	0. 166***	0. 026
Reg1	0. 036*	0. 022	0. 046*	0. 026	0. 025	0. 038
Reg2	0. 036**	0. 017	0. 048***	0. 020	0. 004	0. 030
Reg3	0. 072***	0. 018	0. 085***	0. 022	0. 063**	0. 031
pmem	0. 006	0. 015	0. 011	0. 019	-0. 003	0. 023
rank1	-0. 043	0. 032	-0. 068*	0. 041	0. 019	0. 052
rank2	0. 109***	0. 031	0. 181***	0. 026	0. 107***	0. 051
grad1	-0. 008	0. 027	0. 013	0. 030	-0. 122**	0. 063
grad2	-0. 051**	0. 026	-0. 033	0. 030	-0. 160***	0. 061
grad3	-0. 050*	0. 027	-0. 030	0. 031	-0. 164***	0. 062
Eng1	0. 006	0. 015	0. 042**	0. 018	-0. 068***	0. 027
Eng2	0. 051***	0. 020	0. 110***	0. 026	-0. 043	0. 033
Wor1	0. 050	0. 063	0. 075	0. 077	-0. 014	0. 108
Wor2	-0. 001	0. 059	0. 015	0. 072	-0. 026	0. 106
Wor3	0. 091	0. 059	0. 099	0. 071	0. 064	0. 105

Table 5.4 Wages determinants after adding each job level—continued

east	0. 280***	0. 014	0. 216***	0. 018	0. 390***	0. 024
mid	0. 144***	0. 022	0. 158***	0. 028	0. 112***	0. 038
Gov	-0. 071***	0. 022	-0. 107***	0. 027	0. 001	0. 037
Stat	-0. 002	0. 017	-0. 017	0. 021	0. 028	0. 029
Jon	0. 133***	0. 024	0. 134***	0. 030	0. 131***	0. 038
edui	-0. 033*	0. 019	-0. 075***	0. 024	0. 026	0. 030
Bio	-0. 045	0. 053	-0. 060	0. 072	-0. 045	0. 069
Math	0. 013	0. 035	0. 076	0. 051	-0. 022	0. 049
Phy	0. 078**	0. 035	0. 088**	0. 042	0. 059	0. 066
Elec	0. 093***	0. 017	0. 113***	0. 019	0. 081**	0. 036
Cons	-0. 147***	0. 028	-0. 133***	0. 031	-0. 169***	0. 058
Soc	-0. 041	0. 084	0. 002	0. 109	-0. 153	0. 132
Pol	0. 018	0. 051	0. 062	0. 070	-0. 002	0. 073
Lan	0. 150***	0. 029	0. 181***	0. 051	0. 100***	0. 036
Lit	0. 069***	0. 028	0. 055	0. 038	0. 081**	0. 043
Art	-0. 038	0. 053	0. 058	0. 071	-0. 156**	0. 080
Chem	-0. 010	0. 050	-0. 006	0. 058	0. 005	0. 096
econ	0. 023	0. 031	0. 022	0. 039	0. 001	0. 050
Law	-0. 025	0. 040	-0. 055	0. 056	-0. 019	0. 058
med	-0. 069*	0. 038	-0. 027	0. 054	-0. 121**	0. 055
man	0. 018	0. 019	0. 042*	0. 024	-0. 030	0. 034
Mill	-0. 136***	0. 015	-0. 125***	0. 019	-0. 163***	0. 028
cons	5. 812***	0. 115	5. 854***	0. 139	5. 607***	0. 241
Adj. R2	0.3563		0.3508		0.3956	
Obs.	4632		2999		1633	

Table 5.5 Returns to overeducation and undereducation

	combined		male		Female	
	Coeff.	Std. Err	Coeff.	Std. Err.	Coeff.	Std. Err.
gender	-0.046***	0.015				
reqschooling	0.041***	0.009	0.025***	0.007	0.045***	0.009
surschooling	0.017***	0.003	0.021***	0.008	0.016***	0.003
defschooling	-0.020***	0.004	-0.017***	0.002	-0.016***	0.002
Pcar1	0.017	0.018	-0.006	0.023	0.064**	0.029
Pcar2	0.019	0.018	-0.023	0.023	0.096***	0.029
Pcar3	0.082***	0.023	0.036	0.029	0.177***	0.036
Pqu1	-0.031*	0.018	-0.017	0.021	-0.058*	0.032
Pqu2	-0.016	0.018	0.016	0.022	-0.071**	0.033
Pqu3	0.052**	0.025	0.090***	0.031	-0.027	0.043
Pqu4	0.063**	0.026	0.091***	0.033	0.001	0.043
Pqu5	0.155***	0.053	0.167**	0.067	0.117	0.087
cadre	0.034***	0.013	0.008	0.016	0.092***	0.022
Reg1	0.051**	0.022	0.067**	0.027	0.031	0.038
Reg2	0.046***	0.017	0.067***	0.021	0.006	0.031
Reg3	0.099***	0.018	0.101***	0.023	0.082***	0.031
pmem	0.008	0.015	0.015	0.019	-0.007	0.024
rank1	-0.222***	0.028	-0.248***	0.037	-0.159***	0.043
rank2	-0.007	0.032	-0.028	0.042	0.045	0.050
grad1	0.032	0.027	0.053*	0.030	-0.067	0.063
grad2	0.010	0.026	0.027	0.029	-0.076	0.060
grad3	0.024	0.027	0.043	0.031	-0.064	0.061
Eng1	0.030**	0.015	0.061***	0.018	-0.042	0.026
Eng2	0.133*	0.019	0.200***	0.025	0.025	0.031
Wor1	0.072	0.063	0.085	0.078	0.023	0.109
Wor2	0.022	0.060	0.033	0.072	0.004	0.107
Wor3	0.122**	0.059	0.125*	0.071	0.103	0.106

Table 5.5 Returns to overeducation and undereducation--continued

east	0. 283***	0. 014	0. 224***	0. 018	0. 392***	0. 024
mid	0. 142***	0. 022	0. 161***	0. 027	0. 104***	0. 038
Gov	-0. 060***	0. 022	-0. 089***	0. 027	0. 007	0. 037
Stat	-0. 005	0. 017	-0. 019	0. 021	0. 023	0. 029
Jon	0. 135***	0. 024	0. 132***	0. 030	0. 136***	0. 039
edui	-0. 026	0. 019	-0. 059	0. 025	0. 028	0. 030
Bio	-0. 065	0. 054	-0. 083	0. 073	-0. 078	0. 080
Math	-0. 013	0. 035	0. 053	0. 052	-0. 060	0. 049
Phy	0. 057*	0. 035	0. 069*	0. 042	0. 024	0. 066
Elec	0. 065***	0. 018	0. 079***	0. 020	0. 039	0. 035
Cons	-0. 172***	0. 029	-0. 158***	0. 033	-0. 206***	0. 060
Soc	-0. 086	0. 085	-0. 029	0. 110	-0. 221*	0. 133
Pol	-0. 026	0. 051	0. 015	0. 071	-0. 050	0. 073
Lan	0. 130***	0. 029	0. 162***	0. 056	0. 070**	0. 038
Lit	0. 029	0. 028	0. 013	0. 038	0. 030	0. 042
Art	-0. 057	0. 054	0. 053	0. 072	-0. 196**	0. 080
Chem	-0. 042	0. 050	-0. 048	0. 059	-0. 009	0. 097
econ	-0. 012	0. 031	-0. 013	0. 039	-0. 039	0. 050
Law	-0. 082**	0. 040	-0. 113**	0. 057	-0. 070	0. 058
med	-0. 113***	0. 038	-0. 056	0. 055	-0. 168***	0. 054
man	-0. 015	0. 020	0. 015	0. 024	-0. 074**	0. 034
Mill	0. 009	0. 008	0. 027	0. 010	-0. 017	0. 013
cons	6. 561***	0. 095	6. 646***	0. 115	6. 415***	0. 184
Adj. R2	0.3377		0.3310		0.3821	
Obs.	4632		2999		1633	

Chapter 6 Socially Optimal Investment in Education
---new evidence from China

6.1. Introduction

In the previous two chapters, we considered the effect of individual and job characteristics on individuals' earnings and education choices, which implies the return to education is not fixed as Mincerian equation suggested, but amongst other things varies with the uncertainties associated with labour supply and demand in the long run. Therefore, the determination of an education policy and the supply of qualified labor to the market is quite critical to individuals' education choices. Whether or not graduates will face a high risk of being overeducated for their available job depends in large part upon the direction of government policy on the expansion of higher education. Moreover, the risk of being overeducated for the job secured on graduation becomes less significant if the economy at large is capable in the long run of productively using the expanding quantity of graduates that the government's educational policy determines. In this chapter, we try to derive the social return to human capital in order to investigate its implications for future education policy and the trend of long time labour supply.

In order to improve their relative position in the pecking order and acquire a graduate level of job, individuals have an incentive to accept more education. The analysis of last section in chapter 5 shows more and more individuals in China will be overeducated and have an average lower wages since the speed of higher education expansion is far more than the GDP growth rate[31]. This may result a negative externality on other

[31] However, if the ratio changes over time, such as with more complex production processes etc, it is possible that the demand for graduates may rise faster than the growth rate of GDP. Considering the huge gap between the higher education expansion and GDP growth ratio, it is less likely to happen.

individuals through competitive education theory. However, it is comprehensively believed that education plays a very important role in promoting national economic growth, individuals' skills level, more equitable distribution, national consumption and other wider benefits. In other words, acquiring more education may not be a good education choice to each individual, but overeducated individual will benefit to the whole society because of the good externality of education.

In order to analyze whether higher education expansion benefits the society (here specially refer to China) as a whole or not, we need to find out the impact and the dimension of education on economic growth in China. If education does play an important role in economic growth, the social optimum amount of human capital in China should be estimated so that we can conclude whether China is overeducated from the social point of view and so the future education policy.

This chapter is organized as follows: in the next section, we will discuss whether education can play a role in economic development and in what extent it may promote economic growth. Secondly we develop an endogenous model to estimate the social optimal investment. The policy suggestions and conclusions are drawn in the final part.

6.2. Investment in Education and Economic Growth

In most of the OECD member countries, the public and private expenditure on education accounts for over 6 per cent of the collective GDP. The expenditure in China once criticized by western scholars for its low investment in education has recently risen from 3 per cent of GDP in 1995 to about 5 per cent of GDP in 2002. However the empirical analysis in chapter 4 and 5 suggests that China now faces an overeducation rate of up to 20 per cent. Whether a significant amount of investment from public and private in China is redundant becomes an interesting topic to research.

In this section I will mainly examine the effects of education on economic growth, which in western countries has been widely used to justify a high social rate of investment in education. We have found a positive and significant private return to education in the previous chapter, whereas the relationship between education and the economic growth is still unclear. If the relationship is positive, whether the current investment on education is sufficient and efficient will be examined in the second part.

According to Aghion & Howitt (1998), there are two branches of exogenous methods to test the relationship between human capital and economic growth. One exemplified by Nelson & Phelps (1966) (NL), who assume that the stock of human capital has a level effect on output, whereas the other called augmented Solow model (e.g. Mankiw et al 1992, Romer 1990) (ASM), which states that human capital would affect output's growth rate. The empirical evidence on these two frameworks is quite controversial. Mankiw et al (1992) and De la Fuente & Domenech (2000) demonstrate that the effect of the schooling variable on economic growth is positive and significant. Benhabib & Spiegel (1994) and Pritchett (1996) argue there is no relationship between economic growth and human capital for a large samples of countries, especially in developing countries..

Though the importance of sample quality, especially the proxy for human capital has been emphasized by some scholars, the results are quite controversial among countries. It is our interest in this study to consider whether these two frameworks can be applied to China and whether human capital is a significant factor in determining output growth. Panel data collected by China's statistics bureau will be utilized to realize this assignment. Output in each country in NL and ASM will be replaced by output in each province in China[32] during the period of 1984 to 2004. The logarithm of the average percentage of fixed income investment over GDP in each province was treated as the physical capital investment variable.

[32] Chongqing was separated from Sichuan in 1999, which means we do not have separate data for Chongqing before 1999. Then we combine these two provinces as one unit. Xizhang's data is sparse with poor quality and so we did not include it into our regression. This makes the number of final observation 29.

The logarithm of the average growth rate of the population at age 15-59 is the proxy of changes in the labour force. The logarithm of the average percentage of people with higher education in the work force and the logarithm of the average percentage of people with upper secondary education in the work force act as the stock of human capital separately in the regression.

Using the NL approach, we will assume the Cobb-Douglas production function:

$$Y(t) = A[H(t),t]K(t)^{\alpha} L(t)^{\beta} \tag{6.1}$$

where Y is output, K physical capital, L labour and A is the level of technology depending on the stock of human capital $H(t)$ and time t. The log difference of end and initial period A in province i can be represented as:

$$\log A[H(t),t)_i - \log A[H(0),0)_i = \theta_1 + \theta_2 H_i + \theta_3 H_i[(Y_{\max} - Y_i)/Y_i] \tag{6.2}$$

θ_1 is exogenous technological progress, $\theta_2 H_i$ reflects domestic investment in the stock of human capital and the last term represents technology diffusion from other province, where $Y_{\max}$ is the average output of the richest province during time [0, t], Y_i the average output of province i during time [0, t]. Taking logs of the production function and inserting equation (6.2) yields

$$\begin{aligned}\log Y(t)_i - \log Y(0)_i &= \theta_1 + \theta_2 H_i + \theta_3 H_i[(Y_{\max} - Y_i)/Y_i] \\ &+ \alpha[\log K(t) - \log K(0)] + \beta[\log L(t) - \log L(0)]\end{aligned} \tag{6.3}$$

Table 6.1 shows the regression results of equation (6.3) by OLS using the logarithm of the average percentage of higher education attendees and upper secondary education attendees separately. The coefficients for the stock of human capital and diffusion in both column 2 and column 3 are significant implying human capital does affect the economic growth in China, especially for the contribution of upper secondary attendees.

Besides, the positive role of physical investment on economic growth is quite obvious, which confirms the findings by Nelson & Phelps (1966). Population growth rate may affect the economic growth negatively, albeit not significantly so.

Table 6.1 Regression results from NL approach

	Treat higher education attendees as human capital investment	Treat upper secondary education attendees as human capital investment
Lnschool	$0.0585^{*}(0.041)$	$0.145^{***}(0.060)$
Diffusion	$0.061^{***}(0.034)$	0.048(0.032)
Ln(I/GDP)	$0.285^{***}(0.090)$	$0.303^{***}(0.082)$
lnL	-0.184(0.417)	-0.379(0.405)
Constant	$1.307^{***}(0.218)$	$1.767^{***}(0.318)$
Adjusted R^2	0.307	0.398
observations	29	29

Then we examine the Augmented Solow model (ASM), exemplified by Mankiw et al (1992). Assume production at time t satisfies a Cobb-Douglas function, which can be expressed as:

$$Y(t) = K(t)^{\alpha} H(t)^{\beta} (A(t)L(t))^{1-\alpha-\beta} \qquad (6.4)$$

where Y is output, K physical capital, H the stock of human capital, L labour, and A the level of technology. Solow (1956) assumes the rates of saving or investment rate (s), population growth rate (n) and technological progress rate (g) are exogenous. The time paths of the right-hand side variables are described by the following equations.

$$\begin{aligned}
\dot{k}(t) &= s_k y(t) - (n+g+\delta)k(t) \\
\dot{h}(t) &= s_h y(t) - (n+g+\delta)h(t) \\
\dot{A}(t) &= g(t)A(t) \\
\dot{L}(t) &= n(t)L(t)
\end{aligned} \tag{6.5}$$

where $y = Y/AL$, $k = K/AL$ and $h = H/AL$ is the level of output per effective unit of labour, the stock of physical capital per effective unit of labour and the stock of human capital per effective unit of labour respectively. s_k and s_h are the investment rate in physical and human capital. δ is the depreciation rate. When the economy converges to a steady-state, physical capital and human capital will grow at a constant rate. Hence the exogenous economic growth rate $n+g+\delta$ will have a negative relationship with $\dot{k}(t)$ and $\dot{h}(t)$. Equation (6.5) implies that the optimal physical investment and human capital investment are:

$$\begin{aligned}
k^* &= \left(\frac{s_k^{1-\beta} s_h^{\beta}}{n+g+\delta}\right)^{\frac{1}{1-\alpha-\beta}} \\
h^* &= \left(\frac{s_k^{\alpha} s_h^{1-\alpha}}{n+g+\delta}\right)^{\frac{1}{1-\alpha-\beta}}
\end{aligned} \tag{6.6}$$

substituting (6.6) into production function and taking logs yields the expression for the steady-state output in intensive form.

$$\ln y^* = \ln A(t) + \frac{\alpha}{1-\alpha-\beta}\ln(s_k) + \frac{\beta}{1-\alpha-\beta}\ln(s_h) - \frac{\alpha+\beta}{1-\alpha-\beta}\ln(n+g+\delta)$$

(6.7)

Mankiw et al (1992) also shows the growth of income in the Solow model is a function of the determinants of the ultimate steady state and the initial level of income.

$$\begin{aligned}
\ln(y(t) - \ln(y(0)) &= (1-e^{-\lambda t})\frac{\alpha}{1-\alpha-\beta}\ln(s_k) + (1-e^{-\lambda t})\frac{\beta}{1-\alpha-\beta}\ln(s_h) \\
&-(1-e^{-\lambda t})\frac{\alpha+\beta}{1-\alpha-\beta}\ln(n+t+\delta) - (1-e^{-\lambda t})\ln(y(0))
\end{aligned} \tag{6.8}$$

where $\lambda = (n+g+\delta)(1-\alpha-\beta)$

Equation (6.8) suggests the economic growth increases faster for developing provinces than developed provinces. In other words, the high initial income retains the speed of economic growth. The first column in table 6.2 lists the regression results of the change in the log of income per capital over the period 1984 to 2004 on the log of income per capital in 1984, average investment rate, average population rate during this period for a 29-province sample in China. The negative coefficient on lnGDP84 shows there is a tendency toward convergence in this 29 provinces sample, but this coefficient is not significant. This may suggest China's economic growth has not arrived at a steady state. Column two and three are the regression results after we add the average increasing rate of higher education attendees over the population in each province and the average increasing rate of upper secondary education attendees over the population in each province separately. These two new variables improve the goodness of fit and lower the effect of the initial level of income, which implying human capital investment can explain a certain level of economic growth in China.

Table 6.2 Regression results from augmented Solow model

	Without conditioning on human capital	Conditioning on higher education	Conditioning on upper secondary education
LnGDP84	-0.342 (0.522)	-0.213 (0.203)	-0.209^{*} (0.156)
Ln(I/GDP)	0.288 (0.204)	0.320 (0.215)	0.289 (0.208)
Ln(n+g+δ)	-5.375^{**} (2.445)	-7.013^{**} (3.063)	-6.361^{**} (2.682)
Ln(school)		0.083 (0.091)	0.138^{**} (0.071)
constant	2.767^{***} (0.653)	2.318^{***} (0.131)	2.358^{***} (0.092)
Adj-R^2	0.399	0.417	0.489

observations	29	29	29

6.3. Socially Optimal Investment in Human Capital

Since both NL method and Augmented Solow approach suggest human capital investment does play a role in explaining economic growth, then what is the social optimal amount of human capital investment will generate great interests. Inspired by Lucas (1988), we estimate the efficient amount of human capital stock in each country by maximizing individuals' utility function:

$$U(c,\ \sigma) = \int_0^{\infty} e^{-\rho t} \frac{1}{1-\sigma}[c(t)^{1-\sigma} - 1]N(t)dt \qquad (6.9)$$

where the discount rate ρ and the coefficient of relative risk aversion σ are both positive. The production function is still the Cobb-Douglas format.

$$Y = AK^{\alpha}(uNh)^{1-\alpha} = K + C \qquad (6.10)$$

where A is the level of technology, K physical capital, N the total labour force, u the effort or time devoted to production and h rate of increase of human capital stock. h in Lucas (1988) plays two roles, one contributes to the labour productivity as an endogenous part of uN and the other plays an exogenous factor that may affect the level of technology. Here we only consider the endogenous effect to labour productivity for simplicity.

Rebelo (1991) brings the idea that the increase rate of human capital is determined by the effort devoted to human capital production as well as the amount of physical capital investment. Therefore he considers the level of physical capital investment, when he estimates the effect of human capital stock to economic growth in the endogenous Lucas (1988) model. However the recent report[33] from OECD shows mid income countries (generally with a high economic growth rate) experience sharp increase of student enrolment in upper secondary and higher education.

[33] Education Trends in Perspective--analysis of the world educaiton indicators 2005 edition, UNESCO/OECD, Montreal, 2005

The article indicates 77 per cent increase in tertiary enrolments over the past decade in the middle income countries, compared to a 43 per cent increase in rich countries. Thereafter we assume the time path of h is related with the growth rate of physical capital K[34] (instead of the level of physical capital) and the effort devoted to production u can be represented as

$$\dot{h} = \delta(\frac{\dot{K}}{K} + 1)(1 - u)h \tag{6.11}$$

where δ controls the level of human capital grows and also can be explained as the maximum growth rate that human capital can reach is $\delta(\frac{\dot{K}}{K}+1)$. The current value of Lagrange function for this optimum problem can be written as

$$L(K,h,\theta_1,\theta_2,c,u) = \frac{\mathrm{N}}{1-\sigma}(c^{1-\sigma} - 1) + \theta_1[AK^{\alpha}(uNh)^{1-\alpha} - Nc]$$
$$+\theta_2[\delta(\frac{\dot{K}}{K}+1)(1-u)h] \tag{6.12}$$

with θ_1 and θ_2 can be viewed as the physical capital price and human capital price (Lucas, 1988). The first-order conditions for this equation are:

$$c^{-\sigma} = \theta_1 \tag{6.13}$$

$$Nh(1-\alpha)\theta_1 AK^{\alpha}(uNh)^{-\alpha} = \theta_2\delta(\frac{\dot{K}}{K}+1)h \tag{6.14}$$

and the rate of changes of prices are given by

$$\dot{\theta}_1 = \rho\theta_1 - \frac{\partial H}{\partial K} = \rho\theta_1 - \alpha\theta_1 AK^{\alpha-1}(uNh)^{1-\alpha} \tag{6.15}$$

$$\dot{\theta}_2 = \rho\theta_2 - (1-\alpha)\theta_1 AK^{\alpha}(uNh)^{-\alpha}uN - \theta_2\delta(\frac{\dot{K}}{K}+1)(1-u) \tag{6.16}$$

Let κ denote the growth of per capita consumption $\frac{\dot{c}}{c}$ and λ ($\frac{\dot{N}}{N}$) is the exogenous growth rate of labour force. On a balanced path, physical

[34] Presumably high economic growth rate is related with a high physical capital increase rate.

capital grows at constant rate and so equation (6.10) implies N(t)c(t)/K(t) is constant, by differentiating one gets

$$\frac{\dot{K}}{K}=\frac{\dot{c}}{c}+\frac{\dot{N}}{N}=\kappa+\lambda \tag{6.17}$$

Equation (6.13) yields $\dot{\theta}_1 / \theta_1 = -\sigma K$ and equation (6.15) implies

$$\dot{\theta}_1 / \theta_1 = \rho-\alpha A K^{\alpha-1}(uNh)^{1-\alpha} \tag{6.18}$$

By equating these two equation, one can get

$$\rho+\sigma\kappa=\alpha A K^{\alpha-1}(uNh)^{1-\alpha}h^{1-\alpha-\beta} \tag{6.19}$$

If assuming the growth rate of human capital $v=\dot{h}(t)/h(t)$, then

$$v=\delta(\kappa+\lambda+1)(1-u) \tag{6.20}$$

Differentiating (6.19) with respect to t by assuming consumption and both kinds of capital are growing at constant rates and the time allocation variable u(t) is constant yields the growth rate of consumption

$$\kappa^{*}=v^{*} \tag{6.21}$$

Equation (6.21) means consumption grows at the same rate of human capital, which is also positively related with the growth rate of physical capital. The price change rate of human capital can be derived from (6.16) as

$$\dot{\theta}_2 / \theta_2 = \rho-\delta(\kappa+\lambda+1) \tag{6.22}$$

and from differentiating equation (6.14), we can also get

$$\dot{\theta}_2 / \theta_2 = \lambda-\sigma\kappa \tag{6.23}$$

Differentiate equation (6.21), (6.20) to substitute u and κ with other parameters, then delete $\frac{\dot{\theta}_2}{\theta_2}$ from above two equations, provides the efficient rate of human capital growth:

$$v^{*} = \kappa^{*} = \frac{\rho - \delta\lambda - \delta - \lambda}{\delta - \sigma} \tag{6.24}$$

According to the restriction of v, which cannot exceed $\delta(1+\kappa+\lambda)$ (see equation 6.20), we get

$$\sigma \leq \delta - \frac{(\rho - \delta\lambda - \delta - \lambda)}{\delta(1+\kappa+\lambda)} \tag{6.25}$$

The coefficient of risk aversion σ in Lucas (1988) has a minimum bound, whereas in this model there is a maximum bound for σ (e.g. equation 6.25). In other words, the model only applies for the intertemporal substitutability of consumption is not too low. Since the saving rate in China is extremely high, the assumption is reasonable. In order to work out v and u, we need to figure out the net savings rate s, which according to Lucas (1988) defined as

$$s = \frac{\dot{K}}{Nc + \dot{K}} = \frac{\beta(\kappa+\lambda)}{\rho + \sigma\kappa} \tag{6.26}$$

Based on the data collected by China's statistics yearbook and also the data we estimated above, the average growth rate of labour force λ during the period of 1994-2004 is 0.009, the average growth rate of consumption κ is 0.041, the average growth rate of saving is 0.147 and β equals 0.233. Equation (6.26) derives $\rho+\sigma\kappa$ equals 0.079. Combining equation (6.21), (6.24) and $\rho+\sigma\kappa$ yields an estimate for δ of 0.067. This implies u=0.417.

However substitute these parameters into our efficient human capital growth rate, equation (6.24) and assume $\sigma=1$, then concludes the efficient human capital growth rate is 0.054 and the time allocation to production u is 0.242, which is almost half of the observed value (table 6.3 lists the other possible values). If we devote more effort on human capital accumulation, the economy will enjoy 1.3 percentages higher of consumption growth than it does. The empirical results from our idealistic model imply there are still some room for human capital growth and the whole economy has not been overeducated.

Table 6.3 The optimal human capital investment

σ	v^*	u^*	κ^*
1	0.054	0.242	0.054
2	0.047	0.336	0.047
3	0.045	0.363	0.045

6.4. Conclusion

The regression results in section 2 show that using either NL models or Mankiw et al (1992)'s method, human capital does play an important role in economic growth. The contribution of individuals with upper secondary education to economic growth is around 15 percent in both models, but for the individuals with higher education is not significant for Mankiw et al (1992) model, which accords with the findings of Johnes (2006) for developing countries. The insignificance of individuals with higher education to economic growth may attribute to the size of this amount of people is too small[35]. Empirical evidence by other authors from other countries also uses individuals with secondary education to represent human capital.

The estimation results from Lucas (1988)'s endogenous model show the whole economy has not been overeducated. In other words, the economy can absorb more educated individuals. However, there appears to be a conflict between the 20 per cent of individuals who report themselves as overeducated for their current job that we estimated in chapter 4 with the finding that the whole economy has not been overeducated that we calculated in the last section. The result is similar to the findings in USA. Duncan & Hoffman (1981) found there are around 40 percent individuals are overeducated, but Lucas (1988) estimated the whole economy is

[35] Individuals with higher education account for less than 1 percent of the whole population in the 80s.

insufficient on education investment. Although 20 percent graduates report they are overeducated, there are around 17 percent individuals think they are undereducated. In addition, education may have large social externalities not only in economic production, but also in other wider benefits (such as health, social behaviour, crime). Even if individuals are overeducated, the education investment for the whole society may still be insufficient.

However we must be very careful in referring to the results of section two of the extended Lucas model, since we did not consider other important factors to economic growth. Table 6.1 and table 6.2 show that human capital, physical capital and population can only explain 40 to 50 percent economic growth, and that there exist other important factors which have not been explored by the model. Johnes (2006) explains that besides the factors we considered above, openness, performance of trading partners, political stability, landlock, institutions, etc are also critical to economic growth. A famous Chinese economist Yingqiu Liu said to *The People Daily* (24th, Nov., 2006) that there are three primary forces drive China's economic growth: structural reforms, opening-up policy and cheap labour. Therefore, the results solely in section two are not enough to guide education policy, more factors need to be considered in Lucas's model in future research in order to provide valuable suggestions.

Chapter 7 Conclusion

This thesis is mainly tackling the problem of education choices under uncertainty. Under this frame we discussed the optimal education choices under certainty by multinomial logit model, optimal education choices under uncertainty by B-S option model, the impact of a particular uncertainty (overeducation) on getting a graduate job and labour market wages and finally the socially optimal human capital investment.

In order to evaluate the optimal choices model, we firstly describe four possible choices after compulsory education and their related utility functions. The optimal decision in each stage is derived by maximising individuals' life-time expected utilities after selecting the appropriate choice in each period conditional on the information set at that time and personal discount factor and the utility choices function could be solved by conditional logit model after regulating the disturbance term. The conditional logit model can tackle individuals' heterogeneity (such as discount rate, information set, non-pecuniary utility) very well, but not uncertainty. As analyzed in chapter 2 that there are three major uncertainties involved in each individual's education choices, it is vital to find out the influence of uncertainty on education choices. Then we analyzed the uncertainty and risks related with each choice to help determine the optimal education choice by decision trees as well as utility function conditional on the information set at time t. In order to solve the maximized utility function considering the uncertainty factor, we have to borrow from financial theory. We find investing on education is quite similar to buy a call option that one have the right to secure a level of qualification or not depending on the corresponding wages. Motivated by this idea, we treat the education investment as buying a series of European call option and use the B-S option model to estimate the maximal cost that individuals could invest on education.

The empirical evidence on the B-S model and multinomial logit model is provided in chapter 2. We first evaluate the ex-ante wages for each education choice conditional on their personal characteristics. The predicted wages are quite close to the realized wages in sizes and dimensions. Based on the estimated ex-ante wages, we estimate the wages multinomial logit model, utility multinomial logit model, wages B-S option model and utility B-S option model separately in order to find out the significance of non-pecuniary utility and uncertainty on individuals career move. The estimation results are quite close to the actual results, especially when we consider the effort costs and uncertainty. However, even if we consider these factors, the estimated results are still lower than the actual results. The results also suggest that volatility is not the main reason to stop individuals from being educated for technical qualifications but is the main factor to discourage individuals from attending academic qualification. Individuals who intend to acquire technical qualifications will put more weights on the non-pecuniary utility or the effort costs.

In chapter 4, we model the determinants of getting a graduate level of job by pecking order theory. Based on this theory and three assumptions, we work out a threshold level q_t^{ol}, at time t for the minimum quality q_{it} of individual i to whom an offer is made of a level $l > 0$ job. Through a survey data conducted by a project in June 2003, there are around 20 percent graduates are believed they are overeducated in the China's labour market, with corresponding percentages for college graduate, bachelor, master and doctor of 12.9 percent, 21 percent, 36 percent and 42 percent respectively. The results suggest overeducation in China is more frequent among higher degree than among lower degree. The statistics for gender are quite interesting, namely male graduates are more likely to be overeducated than female graduates if equalizing other conditions.

Then our interests turn to test whether or not there is an equal rate of return on additional years of education, in terms of raising the probability of securing a job of a given level throughout the range from college diploma to PhD. Disaggregating the schooling variable reveals that it is the coefficients on additional years at the masters and PhD levels, particularly for males, which have the most significant effect on boosting an individual's position in the pecking order.

Among all the affecting factors in both ordered probit model and ordered logit model, Party membership, family background, English skills and registration status play a significant role on securing a graduate level of job. Among which, the role of family background is quite interesting, which was represented by father's career rank and father's qualification. Parents belonging to the lowest social rank and with qualifications beyond the master's level all have a significant negative impact on their children's position in the pecking order. In terms of subject, only the graduates from computer & electronic program and language program are easily to find a matched job.

The expected return to personal characteristics, subject of study, type of employment, working location and job levels is analyzed in chapter 5. The return is not equally distributed among education and graduates with master degree receives the highest rewards. University rank, English skills and gender in line with the findings of the determinants on getting a graduate level of job, all have a significant impact on earnings. Whereas, Party membership and class of degree do not have a significant positive effect on graduation wages in China. Relative to working in other sectors, working in the government, state-owned companies and education institutions tends to imply a lower wage level and with starting salaries in joint ventures significantly higher. In the context of subjects, the economic return and the probability to get a graduate job is quite different. Literature, physics, construction, art and medicine all have opposite signs of coefficients in the ordered probit model and OLS logwages regression.

In addition, two stylized facts on the wages to surplus education also hold in China.

After analyzing the determinants of individuals' wages, one can work out the optimal education choices considering the risk of overeducation. Due to the sample limitation, we cannot work out individuals' exact education choices, but a relationship between average wages growth rate and the critical probability that individuals will be not be overeducated by setting the net present value equals to zero. According to our estimation, investing in master education is the most profitable investment and individuals will not be overeducated as long as the wages growth rate is larger than 2.3 per cent.

Due to the externality of education, certain percentages of individuals' overeducation do not imply the whole economy was overeducated. The results in chapter 6 show human capital investment play an important role in China's economic growth by using both NL model and augmented Solow method, especially for upper secondary education. Due to the significant effect of human capital investment on economic growth, we develop the Lucas (1988) endogenous model of optimal social human capital investment in order to estimate the social optimal human capital investment in China. Based on the evidence from *Education at a Glance 2006* that those who have a high GDP growth rate also have a high speed increase on human capital investment, I consider the physical capital growth rate, when I model the human capital increase rate. The estimation results from this new social optimal human capital investment model show that China's human capital investment is insufficient indicating the whole economy has not been overeducated. In other words, the economy can absorb more educated individuals.

In the thesis, we develop individuals' education choices model under uncertainty with strict restrictions, such as individuals are risk neutral, individuals' wages satisfy a Geometric Brownian Motion (GBM) with drift. However, individuals come from poor family background may attach a high discount rate to future uncertain return and their wages may not follow a drifted GBM process even if we control all the personal characteristics. Future empirical uncertainty studies may try to relax these assumptions. In addition, though B-S education choices method can take into account the wages stochastic process, it cannot evaluate the return to uncertainty. An accomplished ex-ante wages equation considering uncertainty factor need to be developed, which can estimate the return to systematic risks and specific risks.

Though our Chinese graduate survey data that we analyzed in chapter 4 and 5 record individuals' comprehensive information, it is only a cross-section data, which restricts our attention on the career mobility. Future researches can use panel data to discuss the overeducation rate overtime and the expected return when experience was considered. When we evaluate the social optimal human capital investment, we only consider the effect of physical capital, human capital and population, which explains around 40 to 50 percent economic growth. Other factors, such as openness, political stability, institutions, etc should be considered in determining the socially optimal level of human capital investment.

Bibliography

Acemoglu, D. & Pischke, J. S. (2001). Changes in the Wage Structure, Family Income and Children's Education, *European Economic Review, Papers and Proceedings*, 45, 890-904.

Aghion, P. & Howitt, P. (1998). *Endogenous growth theory.* MIT Press, Cambridge.

Aitchison, J. & Brown, J. (1963). *The Lognormal Distribution*, Cambridge University Press.

Altonji, J. (1993). The Demand for and Return to Education When Education Outcomes Are Uncertain. *Journal of Labor Economics, 11(1)*, 48-83.

Battu, H., Belfield, C. R. & Sloane, P.J. (1999). Overeducation among graduates: a cohort view. *Education Economics, 7(1)*, 21-38.

Becker, G. (1960). Underinvestment in College Education? *The American Economic Review, 50(2)*, 346-354.

— (1962). Investment in Human Capital: A Theoretical Analysis" *The Journal of Political Economy*, 70(5), 9-49.

— (1993). *Human Capital: a Theoretical and Empirical Analysis with Special Reference to Education*, University of Chicago Press, USA.

Belfield, C. R. (2000). *Economic principles for education: theory and evidence*, Edward Elgar publishing limited.

Ben-Porath, Y. (1967). The production of Human Capital and the Life Cycle of Earnings. *Journal of Political Economy, 75*, 352-365.

Benhabib, J. & Spiegel, M. M. (1994). The role of human capital in economic development: evidence from aggregate cross-country data. *Journal of Monetary Economics, 34(2)*, 143-173.

Black, F. & Scholes, M. (1973). The pricing of options and corporate liabilities, *Journal of Political Economy*, 81 (3), 637-654.

Burdett, K. & Mortensen, D. T. (1998). Wage Differentials, Employer Size and Unemployment. *International Economic Review, 39(2)*, 257-273.

Card, D. (1993). Using geographic variation in college proximity to estimate the return to schooling. NBER working paper, No. 4483.

— (1995). Earnings, schooling and ability revisited. *Research in Labor Economics*, *14*, 23-48.

— (1999). Causal Effect of Education on Earnings. *Handbook of Labor Economics* 3A, Chapter 30.

Chen, S. (2003). Does Wage Volatility Increase for College Attendees? Department of Economics working paper, SUNY at Albany.

Chevalier, A & Walker, I. (1999). Further Results on the Returns to Education in the UK. Keele University working paper.

Chevalier, A. (2003). Measuring over-education. *Economica*, *70*, 509-531.

Cramer. J. (2001). *An Introduction to the Logit Model for Economists* (Second edition), Timberlake Consultants, London.

Davidson, R. & Mackinnon, J. (1993). *Estimation and Inference in Econometrics*, Oxford, Oxford University Press.

— (2003). *Econometric Theory and Methods*, Oxford, Oxford University Press.

De la Fuente, A. & Domenech, R. (2000). Human capital in growth regressions: how much difference does data quality make? Campus de la Universidad Autonoma de Barcelona.

Dearden, L (1998). Ability, Families, Education and Earnings in Britain. IFS working paper 98/14

— (1999). Qualifications and earnings in Britian: how reliable are conventional OLS estimates of the returns to education?" IFS Working Papers, W99/07.

Dearden, L., McIntosh, S., Myck, M. &Vignoles, A. (2002). The Returns to Academic and Vocational Qualifications in Britain. *The Bulletin of Economic Research*, *54(3)*, 249-274.

Dixit, R.K. & Pindyck, R.S. (1994). *Investment under Uncertainty*. Princeton University Press.

Dolton, P. & Silles, M. (2001). Over-education in the graduate labour market: some evidence from alumni data, Centre for the Economics of Education, London School of Economics and Political Science.

Dolton, P. & Vignoles, A. (2000). Incidence and effects of overeducation in the UK graduate labour market. *Economics of Education Review, 19(2)*, 179-198.

Dolton, P. & Vignoles, A. (2002). The return on post-compulsory school mathematics study. *Economica*, *69*, 113-41.

Dominitz, J. & Manski, C. F. (1996). Eliciting student expectations of the returns to schooling. *Journal of Human Resources*. 31(1):1-26.

Duncan, D. C., Turner, M. D. & Pape, A. D. (2003). Minimum wages and school enrollment of teenagers: a look at the 1990's. *Economics of Education Review, 22(1)*, 11-21

Duncan, G. J. & Hoffman, S.D. (1981). The incidence and wage effects of overeducation. *Economics of education Review, 1(1)*, 75-86.

Ellwood, D. & Thomas, J. K. (2000). Who is getting a college education: family background and the growing gaps in enrolment. *Securing the Future*. Sheldon Danziger and Jane Waldfogel, eds. New York: Russell Sage Foundation.

Engelbrecht, Hans-J. (2003). Human capital and Economic Growth. *Economic Record*, *79*, S40-S51.

Fisher, I. (1912). *The Nature of Capital and Income*, New York: Macmillan.

Freeman, R. B. (1971). *The market for college-trained manpower*. Havard University Press.

— (1986). Demand for Education, *Handbook of Labor Economics, 6(1)*, edited by O Ashenfelter and R Layard.

Frenette, M. (2004). The overqualified Canadian graduate: the role of the academic program in the incidence, persistence and economic returns to overqualification. *Economics of Education Review*, *23*, 29-45.

Fuller, W., Manski, C. & Wise, D. (1982). New Evidence on the Economic Determinants of Postsecondary Schooling Choices. *The journal of Human Resources, 17(4)*, 477-98.

Greene, W. (2002). *Econometric Analysis* (Fifth edition), New York: New York University Press.

Griffin, P. & Edwards, A. C. (1993). Rates of Return to Education in Brazil: Do Labor Market Conditions Matter? *Economics of Education Review, 12(3)*, 245-55.

Groot, W. (1996). The incidence of, and returns to overeducation in the UK. *Applied Economics, 28*, 1345-1350.

Halevy, Y. (2002). The Structure of Intertemporal Preferences under Uncertain Lifetime, Department of Eonomics, University of British Columbia.

Hartog, J., Pfann, G. & Ridder, G. (1989). (Non-)Graduation and the Earnings Function. *European Economic Review, 33*, 1373-1395.

Hartog, J. & Vijverberg, W. (2002). Do wages really compensate for risk aversion and skewness affection. IZA Discussion Paper No. 426

Hartog, J. & Vijverberg, W. (2006). Schools, skills and risks. Conference paper for the economics of education, Dijon.

Heckman, J. (1976). A life-cycle model of earnings, learning, and consumption. *Journal of Political Economy, 84*, S11-44

— (1979). Sample Selection Bias as a Specification Error. *Econometrica, 47(1)*, 153-161.

Hey, J. D. (2003). *Intermediate Microeconomics: People are Different.* McGraw-Hill.

Hogan, V. & Walker, I. (2005). Education choices under uncertainty: implications for public policy. Conference paper for the risk in education meetings.

Hull, J. C. (2000). *Options, Futures, and Other Derivatives (5th Edition)*, Prentice Hall.

Johnes, G. (1993). *The Economics of Education.* The Macmillan Press Limited.

— (1998). Human capital versus sorting: new data and a new test, *Applied Economics Letters*, 5, 665-67.

— (2006). Education and economic growth, Lancaster university working paper.

Johnes, G. & R. McNabb (2004). Never give up on the good times: student attrition in the UK, *Oxford Bulletin of Economics and Statistics*, 66, 23-48.

Kane, T. J. (1994). College Attendance By Blacks Since 1970: The Role of College Cost, Family Background and the Returns to Education, *Journal of Political Economy*, 102(5), 878-911.

Keane, M. & Wolpin, K. I. (1994). The Solution and Estimation of Discrete Choice Dynamic Programming Models by Simulation and Interpolation: Monte Carlo Evidence. Research Department Staff Report 181, Federal Reserve Bank of Minneapolis.

— (1997). The Career Decisions of Young Men. *The Journal of Political Economy, 105(3)*, 473-522

Kodde, D. (1986). Uncertainty and the demand for education. *The Review of Economics and Statistics, 68(3)*, 460-67.

Kohn, M., Manski, C. & Mundel, D. (1976). An Empirical Investigation of Factors Influencing College Going Behaviour. *Annals of Economic and Social Measurement*, 5, 391-419.

Levhari, D. & Weiss, Y. (1974). The effect of risk on the investment in human capital. *The American Economic Review, 64(6)*, 950-63.

Lucas, R. (1988). On the mechanics of economic development. *Journal of Monetary Economics, 22*, 3-42

Mankiw, G., Romer, D., & Weil, D. (1992). A Contribution to the Empirics of Economic Growth. *Quarterly Journal of Economics, 106*, 407-437.

Manski, C. & Wise, D. (1983). *College Choice in America*, Harvard University Press.

Mayston, D. (2002). *Evaluating the Benefits of Career Guidance*, Centre for Performance Evaluation and Resource Management, York. and Centre for Guidance Studies.

Mcfadden, D. (1974). Conditional Logit Analysis of Qualitative Choice Behavior. In P. Zarembka, ed., *Frontiers of Econometrics*. New York: Academic Press.

Merton, R. C. (1973). Theory of rational option pricing, *Bell Journal of Economics and Management Science*, *4 (1)*, 141-183.

Miller, P. W. (1984). The Causes and Consequences of Interruptions to Full-Time Education. *Australian Economic Papers*, *23(42)*, 61-70.

Mincer, J. (1974). *Schooling, experience and earnings*. New York: Columbia University Press for the National Bureau of Economic Research.

Nelson, R. & Phelps, E. (1966). Investment in Humans, Technological Diffusion, and Economic Growth. *American Economic Review*, *61*, 69-75.

Oosterbeek, H. & Ophem, H. (2000). Schooling choices: Preference, discount rates and rates of return. *Empirical Economics*, *25(1)*, 15-34.

Pakes, A. (1986). Patents as Options: Some Estimates of the Value of Holding European Patent Stocks. *Econometrics*, *54(4)*, 755-784.

Palacios-Huerta, I. (2003). An Empirical Analysis of the Risk Properties of Human Capital Returns. *The American Economic Review*, *93(3)*, 948-964.

Pratt, J. (1964). Risk Aversion in the Large and in the Small, *Econometrica, 32*, 122-136.

Pritchett, L. (1996). Where has all the education gone? World Bank Policy Research Department working paper No.1581.

Rebelo, S. (1991). Long-Run Policy Analysis and Long-Run Growth. *The Journal of political Economy*, *99*, 500-521.

Romer, P. (1990). Endogenous Technological Change. *Journal of Political Economy*, *98*, S71-S102.

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

Rumberger, R. W. (1981). The rising incidence of overeducation in the U.S. labor market. *Economics of Education Review*, *1(3)*, 293-314.

Rumberger, R.W., & Thomas, S. L. (1993). The economic returns to college major, quality and performance: A multilevel analysis of recent graduates. *Economics of Education Review*, 12(1), 1-19.

Sattinger, M. (1993). Assignment models of the distribution of earnings, *Journal of economic literature*, *31(2)*, 831-880.

Sicherman, N. (1991). Overeducation" in the labour market. *Journal of Labour Eonomics*, *9(2)*, 101-122.

Sloane, P. J., Battu, H. & Seaman, P.T. (1999). Overeducation, undereducation and the British labour force. *Applied Economics*, *31(11)*, 1437-1453.

Snow, A. & Warren, R. S. (1990). Human capital investment and labor supply under uncertainty. *International Economic Review*, *31(1)*, 195-206.

Solow, R. (1956). A Contribution to the Theory of Economic Growth. *Quarterly Journal of Economics*, *70*, 65-94.

Spence, M. (1973). Job market signalling. *Quarterly Journal of Economics*, *87(3)*, 355-374.

Stiglitz, J. E. (1975). The theory of "screening" education and the distribution of income. *American Economic Review*, *LXV*, 283-300.

Thaler, R. H. (1999). Mental Accounting Matters. *Journal of Behavioural Decision Making*, *12*, 183-206.

Verdugo, R. & Verdugo, N. (1989). The impact of surplus schooling on earnings: some additional findings. *Journal of Human Resources*, *24 (4)*, 629 – 643.

Walker, I & Zhu, Y. (2003). Education, earnings and productivity: recent UK evidence, Office for National Statistics.

Webbink, D & Hartog, J (2004). Can students predict starting salaries? Yes! *Economics of Education Review*, *23*, 103-113.

Weisbrod, B. A. (1962). Education and Investment in Human Capital. *Journal of Political Economy, 70(2)*, 106-123.

Weiss, A. (1995). Human capital vs signalling explanations of wages. *Journal of Economic Perspectives*, *9 (4)*, 133 – 154.

William, G & Gordon, A. (1981). Perceived earnings functions and ex-ante rate of return to post compulsory education in England, *Higher Education, 10(2)*, 199-227

Williams, J. T. (1979). Uncertainty and the Accumulation of Human Capital Over the Life Cycle. *Journal of Business*, *52(4)*, 521-48.

Willis, R. J. & Rosen, S. (1979). Education and self-selection. *Journal of Political Economy (Supplement) 87(5)*, S7-S36.

Winston, G. C. (1999). Subsidies, hierarchy and peers: the awkward economics of higher education. *Journal of economic perspectives*, *13*, 13-37.

Wooldridge, J. (2002). *Econometric analysis of cross section and panel data*, The MIT Press, USA.