fea-RP Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto
Universidade de São Paulo

Texto para Discussão

Série Economia

TD-E 06 / 2009

The Choice of Major: Effects on Welfare and an Evaluation of the No-Switching Majors Rule

Prof. Dr. Daniel Domingues dos Santos

Av. Bandeirantes, 3900 - Monte Alegre - CEP: 14040-900 - Ribeirão Preto-SP
Fone (16) 3602-4331/Fax (16) 3602-3884 - e-mail: cebelima@usp.br site:www.fearp.usp.br

Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto
Universidade de São Paulo

Universidade de São Paulo

Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto

Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto
Avenida dos Bandeirantes,3900
14049-900 Ribeirão Preto – SP

A série TEXTO PARA DISCUSSÃO tem como objetivo divulgar: i) resultados de trabalhos em desenvolvimento na FEA-RP/USP; ii) trabalhos de pesquisadores de outras instituições considerados de relevância dadas as linhas de pesquisa da instituição. A série foi subdividida em função das principais áreas de atuação da FEA-RP/USP: Economia, Administração e Contabilidade. Veja o site da CPq na Home Page da FEA-RP: www.fearp.usp.br. Informações: e-mail: cpq@fearp.usp.br

The choice of major: effects on wages and an evaluation of the no-switching majors rule

Daniel Domingues dos Santos

Advisor: James J. Heckman, The Henry Schultz Distinguished Service Professor in Economics and the College

ABSTRACT

Abstract

I estimate a dynamic discrete choice model of the decision of going to college in a specific major. Dynamics in the model result from a correlated bayesian learning structure about individual comparative advantages in the labor market, which allows me to decompose the income gains associated to college education into three components: (i) human capital accumulation, (ii) access to specialized segments of the labor market, and (iii) better exploration of comparative advantages due to the use of information acquired in college. The estimation suggests scientific majors are especially benefitted from (ii) whereas in non-scientific occupations the effect (i) dominates.

I then use the estimated structural parameters to simulate the impact of restrictions to changes of major during college, a policy commonly found outside North America. The results suggest that if American students were not allowed to switch majors, they would have welfare losses equivalent to their first year of labor income after college.

THE UNIVERSITY OF CHICAGO

THE CHOICE OF MAJOR: EFFECTS ON WAGES AND AN EVALUATION OF THE NO-SWITCHING MAJORS RULE

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE SOCIAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF ECONOMICS

BY
DANIEL DOMINGUES DOS SANTOS

CHICAGO, ILLINOIS
AUGUST 2008

ABSTRACT

I estimate a dynamic discrete choice model of the decision of going to college in a specific major. Dynamics in the model result from a correlated bayesian learning structure about individual comparative advantages in the labor market, which allows me to decompose the income gains associated to college education into three components: (i) human capital accumulation, (ii) access to specialized segments of the labor market, and (iii) better exploration of comparative advantages due to the use of information acquired in college. The estimation suggests scientific majors are especially benefitted from (ii) whereas in non-scientific occupations the effect (i) dominates.

I then use the estimated structural parameters to simulate the impact of restrictions to changes of major during college, a policy commonly found outside North America. The results suggest that if American students were not allowed to switch majors, they would have welfare losses equivalent to their first year of labor income after college.

TABLE OF CONTENTS

ABSTRACT iii

LIST OF TABLES v

LIST OF FIGURES vi

ACKNOWLEDGMENTS ix

1 THE CHOICE OF MAJOR: EFFECTS ON WAGES AND AN EVALUATION OF THE NO-SWITCHING MAJORS RULE 1
- 1.1 Introduction 1
- 1.2 Model 5
- 1.3 Empirical analysis 19
  - 1.3.1 The data 19
  - 1.3.2 The variables used in the econometric exercise 25
- 1.4 Empirical strategy and results 30
  - 1.4.1 Empirical strategy 31
  - 1.4.2 Results 43
- 1.5 What have we learnt from this research? - final remarks 57

A MAIN GRAPHS 64

B MAIN GRAPHS (CONT.) 83

C LIST OF MAJORS 91

D POSTERIOR DISTRIBUTIONS AND MCMC SEQUENCES 96

E SOME ROBUSTNESS CHECKS 114

F IDENTIFICATION 116

REFERENCES 125

LIST OF TABLES

1.1 Demographic profile of college graduates, college dropouts, and high-school graduates . . . 21
1.2 Enrollment in college, by declared major . . . 22
1.3 Transition probabilities, average across years . . . 23
1.4 Transitions conditional on the number of years of college education 23
1.5 Correlations between test scores . . . 24
1.6 Average parameters of the signal equations . . . 46
1.7 Average parameters of the wage equations . . . 49
1.8 Average parameters of the human capital accumulation equations . 51
1.9 Average parameters of the cost equations . . . 51
1.10 Average policy effects . . . 53
1.11 Comparison between the parameters obtained from the sample of white males and from the enlarged sample: signal and cost equations 61
1.12 Comparison between the parameters obtained from the sample of white males and from the enlarged sample: wage and human capital accumulation equations . . . 62
1.13 Comparison between the parameters obtained from the sample of white males and from the enlarged sample: Lambda and Omega . . 63
1.14 Comparison between the policy effects obtained from the sample of white males and from the enlarged sample . . . 63

C.1 List of majors - sciences . . . 92
C.2 List of majors - non-sciences . . . 93
C.3 Frequencies of majors by year - sciences . . . 94
C.4 Frequencies of majors by year - non-sciences . . . 95

E.1 Robustness checks of major-specific parameters . . . 114
E.2 Robustness checks of other parameters . . . 115

LIST OF FIGURES

A.1 Mechanics of the model - the role of G . . . 65
A.2 Mechanics of the model - the role of other elements of the information set . . . 66
A.3 Average standardized test scores, by field . . . 67
A.4 MCMC distribution of the intercept of the signal equations . . . 68
A.5 MCMC distribution of the ξ's in the signal equations . . . 69
A.6 MCMC distribution of the σ's in the signal equations . . . 70
A.7 MCMC distribution of the diagonal elements of Λ_0 . . . 71
A.8 MCMC distribution of the off-diagonal elements of Λ_0 . . . 72
A.9 MCMC distribution of the diagonal elements of Ω_0 . . . 73
A.10 MCMC distribution of the off-diagonal elements of Ω_0 . . . 74
A.11 MCMC distribution of the average (perceived) specific ability among college graduates . . . 75
A.12 MCMC distribution of the intercept of the wage equations . . . 76
A.13 MCMC distribution of the coefficient on labor market experience, in the wage equations . . . 77
A.14 MCMC distribution of the variance (ρ), in the wage equations . . . 78
A.15 Wage decomposition among sciences graduates . . . 79
A.16 Wage decomposition among non-sciences graduates . . . 80
A.17 MCMC distribution of the rate of accumulation of the unskilled talent, θ_{j0} . . . 81
A.18 MCMC distribution of the rate of accumulation of the scientific talent, $\theta_{j,sci}$ . . . 82

B.1 MCMC accumulation rate of the non-scientific talent, $\theta_{j,n-sc}$ . . . 84
B.2 MCMC distribution of the intercept of the cost equations . . . 85
B.3 MCMC distribution of the coefficient on parent's education, in the cost equations . . . 86
B.4 MCMC distribution of the option value of college education by major, in period 0 . . . 87
B.5 MCMC distribution of the average welfare loss associated to the no-switching majors rule . . . 88
B.6 MCMC distribution of the enrollment rate under the no-switching majors rule . . . 89
B.7 MCMC distribution of the graduation rate under the no-switching majors rule . . . 90

D.1 MCMC - signal, sciences . . . 101
D.2 MCMC - signal, non-sciences . . . 102
D.3 MCMC - diagonal elements of Λ_0 . . . 103

D.4 MCMC - off-diagonal elements of Λ_0 104
D.5 MCMC - elements of Ω_0 105
D.6 MCMC - elements of Ω_0 (cont.) 106
D.7 MCMC - intercepts of the wage equations 107
D.8 MCMC - coefficients on experience in the wage equations 108
D.9 MCMC - variances of the wage equations 109
D.10 MCMC - human capital accumulation rates 110
D.11 MCMC - human capital accumulation rates (cont.) 111
D.12 MCMC - cost equation, sciences 112
D.13 MCMC - cost equation, non-sciences 113

To my parents.

ACKNOWLEDGMENTS

I am especially thankful to James Heckman for his indispensable suggestions during the course of this research, and to CAPES (Brazilian Ministry of Education) and the University of Chicago for the financial support received during my PhD program. I am also indebted to Rodrigo Bueno, Carlos Corseuil, Barbara Cunha, Flavio Cunha, Marco Fiaccadori, Kenneth Judd, Hedibert Lopes, Gabriel Madeira, Heleno Pioner, Mari Sakudo, Sergio Urzua and George Yates and seminar participants at the University of Chicago and IPEA/RJ for comments and help received in the various parts of the project. Contact information: ddsantos@uchicago.edu

CHAPTER 1

THE CHOICE OF MAJOR: EFFECTS ON WAGES AND AN EVALUATION OF THE NO-SWITCHING MAJORS RULE

1.1 Introduction

Education policymakers seem to be very interested in making college curricula more flexible[1]. On one hand, there is the belief that flexibility may help a heterogeneous mass of students to adapt to a set of occupations that has also become increasingly heterogeneous as the technology change and the division of labor evolve. On the other hand, the relatively flexible North American university system was apparently successful in expanding college education.

The original focus of this research is to evaluate the economic importance of a specific aspect of the flexibility observed in the United States and Canada and not frequently found elsewhere, namely the freedom students have to change majors during the period in college, without losing the credits already earned so far. This question is interesting for two reasons: in terms of policy, it involves a change in the mechanism universities operate that could potentially bring positive economic results without implying in extra expenses to the society. On the other hand, an econometric evaluation of this policy is methodologically challenging, because whichever rule is chosen, it affects everybody at the same time, so that the direct comparison between people affected and not affected by this policy is not available to the analyst.

1. See the conclusions of the Bologna Convention, and the Brazilian project Nova Escola, among others.

The theoretical framework used to investigate this problem is one in which individuals complete high school with only partial information about their comparative advantage to different careers, and college education discloses new information about this matching that could be useful for them to make better professional choices. Major switching can then be justified by the fact that the information set evolves over time, and restrictions to this movement can be evaluated in terms of the ability to use this new information to improve on the individuals' previous choices. The model shows total gains of one period of college experience can be divided into an increase in the individual's productivity that could be used in high-school occupations even if the individual did not get the degree, and an option value of eventually graduating and being able to supply labor to a specialized segment of the market. The no-switching restriction represents in this case a change in this option value, which can be aggregated over individuals to have a global measure of its impact. The empirical strategy in this case consists on the estimation of a structural dynamic discrete choice model where agents always choose the alternative that maximizes their expected present value of wealth, with the use of the estimated structural parameters to simulate a counterfactual world where the individual maximization problems are constrained by a no-switching majors clause.

The underlying model is rich enough to allow for correlated learning and human capital accumulation, i.e., classes in a given major not only increase and bring new information about the talent required to perform tasks associated to that major, but also change the level and the information about other types of talent. It also allows me to quantify three channels through which college experiences could potentially affect ex-post wages, which constitutes the second goal of this paper.

In the first channel, which I call human capital effect, college courses increase the level of the various types of individual abilities, making them more productive. In the second, denoted as the credential effect, graduation in a given major let individuals get access to a privileged segment of the labor market, where personal traits are priced in a specific way. Finally, there is the informational effect, where the better knowledge of one's true talents allows her to make better choices and to explore more efficiently hers comparative advantages.

Although the specific policy evaluation just mentioned has not been conducted (to my knowledge) in the literature, there are a couple of papers closely related to this problem that should be briefly mentioned. Two of the three forces that interact in the model to determine the choice of going to college have been the objects of two vast literatures. In the first branch it is emphasized the importance of the increase in the agent's productivity associated with human capital accumulation, which in equilibria should equalize the foregone earnings during the period in college. The main references in this agenda are the early theoretical articles by Ben Porath (1967), Gary Becker (1962, 64), which met the empirical evidence found by Mincer (1958,74), exhaustively replicated by others until nowadays. In the second branch, inaugurated by Willis and Rosen (1979), it is proposed an alternative model where college and high-school occupations use productive traits in different ways, so that individuals self-select themselves into these two sectors in order to get the highest reward for their bundle of attributes, and where the main force driving the decision of going to college is the individual's comparative advantages. This line of investigation was followed by Heckman and Cameron (1998,2001), who incorporated dynamics and forward looking behavior to the agent's decision, and Cunha, Heckman and Navarro (2005), who proposed a way to incorporate and separate the unobserved components of initial heterogeneity and uncertainty in the individual dynamic problem. In this branch, the empirical strategy builds heavily upon the work of Roy (1951).

Partial information about specific talents with Bayesian learning coming from signals revealed though signals received after the first decision are a common aspect in the studies about choice of career by Jovanovic (1979), Miller (1984), and applied to the choice of major by Arcidiacono (2004). In the Multiarmed Bandit model adopted by Jovanovic and Miller, abilities are allowed to be multidimensional, but uncorrelated, which implies that the signals are informative only about the payoffs associated to the decision currently made. This simplification is useful to circumvent the curse of dimensionality (the impossibility of computationally solving problems with too many periods and choices per period), through the use of Dynamic Allocation Indexes, but is frequently though as unrealistic, and I do not use in the present paper. In the case of Arcidiacono, human capital is unidimensional (which in one sense can be seen as the other extreme, where talents are perfectly correlated), which also facilitates the characterization of the solutions, but implies the different occupations form a ladder so that most able individuals would always choose the good careers, and the others would go to the bad careers (in his case one only has to know about her position on this ladder). In my exercise, multidimensional abilities allow that individuals may be suited for some careers but not for others, with no predetermined hierarchy of majors.

The paper is structured in three parts following this introduction. In the next section I introduce and discuss the theoretical model used to analyze the questions posed above. In section 3 I describe the dataset and the empirical strategy used to perform the estimations and simulations required to answer these questions. In section 4 I discuss the limitations of this research and the extensions that should be carried out in order to improve the analysis.

1.2 Model

The model below describes a population of risk neutral individuals who want to maximize the expected present value of their income flows. The economy consists of $J+1$ types of occupations, each of which combines the observable productive characteristics X_i of workers, with a specific unobservable talent U_{ijt} (that evolves over time) to produce output. The specific ability is not observed neither by the individuals and firms, nor by the econometrician, but the agents do observe some signals about it and form their best guesses U^s_{ijt}, which are used to make their choices at time t. If one starts working, I assume he/ she stays in that occupation for his/ her entire working life (T periods), and no further decision takes place. Before that, agents have to decide whether to go to college, and in which major, and they know that college experience will not only increase their vector of abilities, U_i, but also bring some new information about it, so that their guesses about U_i may become more accurate (and hence allowing them to make better decisions). Only individuals who graduated in major j have the option to supply labor to that segment of the market, and the sector $j=0$ comprises all of the occupations that do not require a college degree, being therefore called the "high-school" or "unskilled" sector.

Environment

Formally, life begins at the end of high-school, and agents are endowed with a vector of (true) abilities and a vector of signals about them, related to one another through the equation:

$$\begin{aligned} U_{i0} &= U_{i0}^{s} + v_{i0} \\ & v_{i0} \perp U_{i0}^{s} \\ v_{i0} &\sim N\left(0, \Lambda_0\right) \\ U_{i0}^{s} &\sim N\left(0, \Omega_0\right) \end{aligned} \tag{1.1}$$

where v_{i0} is the portion of the abilities unknown to the agents at time 0. The vectors U_{i0}, U_{i0}^{s}, and v_{i0} have coordinates of the type $U_{ij0}, U_{ij0}^{s}, v_{ij0}$, corresponding to the respective components of ability in different sectors.

The choice set in period 0 is formed by $J+1$ possibilities, denoted $\left\{d_{ij0}\right\}_{j=0}^{J}$, with the restriction that agents can take only one action per period, which means that $d_{ijt}=1$ iff action j is taken in period t (and 0 otherwise), and:

$$\sum_{j=0}^{J} d_{ijt} = 1; \forall t$$

As mentioned before, d_{i0t} refers to the choice of going to the labor market in sector 0, which implies that the individual will no longer have further decisions to make, i.e.:

$$d_{i0t} d_{ijt+1} = 0; \forall j > 0$$

If one decides to go to college in major j, he/ she faces a new decision node in the next period, with a set of $J+1$ actions similar to the previous period. Each of these choices $j>0$ is costly, both because students have to pay tuition and fees to the university and due to the effort and other eventual personal sacrifices spent in that period. I call C_{ijt} the total cost of these choices, and it depends both on observable characteristics of agents, and on non-persistent shocks that arrive at each period and observed only by the agents:

$$\begin{aligned} C_{ijt} &= \beta_j^c X_i^c + \epsilon_{ijt} \\ X_i, U_{it}^s \perp \epsilon_{ijt} &\sim N(0,S), i.i.d. \end{aligned} \tag{1.2}$$

The initial information set upon which individuals make decisions can then be written as:

$$I_{i0} = \{X_i, U_{i0}^s, \epsilon_{i0}\}$$

College experience

If an individual decides to take one period of courses at college level, two things happen. First, these classes potentially improve the individual productivity by the acquisition of new skills, therefore affecting the level of true abilities. In particular, I assume that one period of classes in major j increase human capital in a deterministic fashion described by the equation:

$$\begin{aligned} U_{it+1} &= \theta_j U_{it} \\ \theta_j &= \begin{bmatrix} \theta_{j0} & & 0 \\ & \ddots & \\ 0 & & \theta_{jJ} \end{bmatrix} \end{aligned} \tag{1.3}$$

Second, college students observe new signals about their true abilities that are incorporated to their information sets. The new information is associated to the individual characteristics through:

$$G_{ijt} = \beta_j^g X_i^g + \varphi_j U_{ijt} + \sigma_j \varepsilon_{it} \tag{1.4}$$
$$X_i^g, U_{it} \perp \varepsilon_{it}$$
$$\varepsilon_{it} \sim N(0,1) \tag{1.5}$$

In this expression, ε_{it} is an i.i.d. non-persistent random shock that affects the signal in that period (e.g. if G_{ijt} is the GPA obtained in major j, at time t, ε_{it} could be sickness, family problems, etc., that affected the student's performance, but which is unrelated to abilities or other characteristics).

After revealed, the new signal is incorporated to the information set, that becomes:

$$I_{it+1}^{(j)} = \{I_{it}, G_{ijt}, \epsilon_{it}\}$$

The superscript (j) is used only to stress the fact that different decisions would in principle disclose different pieces of information. Although the set above in general contains everything the agents know until period t, we will see that only $\{X_i, E(U_{it}|I_t), \epsilon_{it}\}$ are relevant for the decisionmakers to solve their maximization problems (and in particular, previous realizations of ϵ_{it-s} are unimportant, as these terms are non-persistent and do not affect the future payoffs of any choice). In period 0, we directly have that $E(U_{i0}|I_0) = U_{i0}^s$, and the equations (1)-(3) deliver an analytical expression for the law of motion of $E(U_{it}|I_t)$ over time (which for simplicity is denoted just U_{it}^s):

$$U^s_{it+1} = \Lambda_{it+1}\left[\left(\Lambda_{it}\boldsymbol{\theta}_j\right)^{-1}U^s_{it} + \frac{\varphi_j}{\sigma_j^2}\left(G_{ijt} - \beta^g_j X^g_i\right)D_j\right] \quad (1.6)$$

$$\Lambda^{-1}_{it+1} = \left(\theta_j\Lambda_{it}\theta_j\right)^{-1} + \left(\frac{\varphi_j}{\sigma_j}\right)^2 D_j D'_j \quad (1.7)$$

where $D_j = [0, ...0, , 1, 0, ...0]$ is an indicator vector with the coordinate j being its only nonzero entry. The expressions above are a version of the Normal-Bayesian learning described in De Groot(1970) and Zellner(1971), and applied to economic problems by Jovanovic(1979), Miller(1984) and Arcidiacono(2004), among others, slightly generalized to allow for variations over time in the hidden variable U_i. As expected, the inverse of the variance matrix of the unknown component of abilities (also called the precision matrix, in the Bayesian jargon) increases proportionally to the precision in the previous period, and inversely to the rate of human capital accumulation. Moreover, the more the signal is related to the true abilities (summarized by $|\varphi|$), the faster is the revelation of information, and the noisier this signal is, the slower is informational disclosure. The first equation also says that the current information individuals have about abilities is described as a weighted average between what they knew in the previous period and the new signal. The distribution of U_{it} conditional on I_{it} can then be defined as:

$$U_{it}|I_{it} \sim N\left(U^s_{it}, \Lambda_{it}\right)$$

It is interesting to notice two things in the form of learning proposed here. First, the process of learning is correlated, and the standard methods used to solve the (uncorrelated) multi-armed bandit problem proposed by Gittins (1979) and Gittins and Jones (1979) will not apply in this case. When taking classes in major j, the student knows that all coordinates of her ability vector will change, and also that they will gather new information about this whole vector, causing the payoffs associated to all available choices to change as well. Second, because the hidden vector of abilities is also growing over time, a general result that the variance of the unknown portion of abilities decreases over time no longer holds in this case. Instead, what we can say is that this variance either decreases over time or increases at a slower rate than the variance of the known part of talents. To see this, notice that the motion of each element of Λ_{it} and Ω_{it} (denoted λ_{kmt+1} and ω_{mkt+1}, respectively) can be written as:

$$
\begin{aligned}
\lambda_{kmt+1} &= \theta_{jk}\theta_{jm}\lambda_{kmt} - \theta_{jk}\theta_{jm}\lambda_{kmt}\frac{\rho_{kjt}\rho_{mjt}}{\rho_{mkt}}\left(\frac{\xi_j^2\lambda_{jjt}}{1+\xi_j^2\lambda_{jjt}}\right) \\
\omega_{mkt+1} &= \theta_{jk}\theta_{jm}\omega_{mkt} + \theta_{jk}\theta_{jm}\lambda_{kmt}\frac{\rho_{kjt}\rho_{mjt}}{\rho_{mkt}}\left(\frac{\xi_j^2\lambda_{jjt}}{1+\xi_j^2\lambda_{jjt}}\right) \\
\xi_j &= \frac{\theta_{jj}\varphi_j}{\sigma_j}; \rho_{kjt} = \frac{\lambda_{kjt}}{\sqrt{\lambda_{kkt}\lambda_{jjt}}}
\end{aligned}
$$

such that, in general there is a transference of $\theta_{jk}\theta_{jm}\frac{\rho_{kjt}\rho_{mjt}}{\rho_{mkt}}\left(\frac{\xi_j\lambda_{jjt}}{1+\xi_j\lambda_{jjt}}\right)$ from Λ_{it} to Ω_{it}, and for the diagonal elements ($m = k$) that represent the variances of v_{it} and U_{it}^s, respectively, this term necessarily lies in the interval $\left(0, \theta_{jk}\theta_{jm}\lambda_{kmt}\right)$. The case $\left\{\theta_{jk} = 1, \forall j, k : \lambda_{km0} = 0, \forall j \neq k\right\}$ would recover the standard multi-armed bandit model (with $\theta_{jk}\theta_{jm}\frac{\rho_{kjt}\rho_{mjt}}{\rho_{mkt}}$ being 1 in the diagonal and 0 elsewhere).

Labor market

In this world information is symmetric, although incomplete. In every sector, firms and workers agree about a wage that captures the perceived productivity in that segment, which should then be a function of observable characteristics and the known part of the specific sectoral ability, i.e., $w_{ijt} = f_j\left(X_{it}, U^s_{ijt}, \eta_{it}\right)$, where η_{it} is just an i.i.d. shock (observed by the agents but not by the econometrician) that affects the workers' productivity at a given point in time. Because the focus of this research is on learning during college, I assume no further information about talents is disclosed after the last period in college, denoted by τ_i. Therefore, $U^s_{it} = U^s_{i\tau_i}, \forall t \geq \tau_i$. Finally, the observed part of personal attributes, X_i has only one element that varies over time, in a deterministic way, which is the experience in the labor market (defined as $t - \tau_i$). For simplicity, I assume the wage equation is linear, i.e.:

$$w_{ijt} = \beta^w_j X^w_i + A_j\,(t - \tau_i) + U^s_{ij\tau_i} + \rho_j \eta_{it} \tag{1.8}$$

$$X^w_i, U^s_{ij\tau_i} \perp \eta_{it} \sim N\,(0, 1)$$

$$X^w_i \subseteq X_i \tag{1.9}$$

We are now in a position to investigate how tertiary education affects wages in this economy. First, a college degree determines that the individual characteristics $\left(X^w_i, t - \tau_i\right)$ will no longer be priced in the same way as high-school occupations, since the hedonic prices of this characteristics are now indexed by j. Second, people who did not finish college will go to the labor market earlier in life, and will not only start making money before the college graduates, but also will have more experience, since their τ_i will be lower. Third, college education will affect the specific ability required in sector j in two ways, through information and human capital accumulation. Because the wage equation is linear as well as the law of motion of U_{it}^s, it is possible to exactly separate these effects (and to quantify them, as shown in the next section). In particular,

$$w_{ijt} = \beta_j^w X_i^w + A_j (t - \tau_i) + \rho_j \eta_{it} + \left(\prod_{s=0}^{\tau_i} \theta_{k_s j} \right) U_{ij0}^s + H \left(U_{i0}^s, G_i^{\tau_i}, X_i, D_i^{\tau_i} \right) \tag{1.10}$$

where $H \left(U_{i0}^s, G_i^{\tau_i}, X_i, D_i^{\tau_i} \right)$ is also a linear function of the initial signal, U_{i0}^s, the decision history, $D_i^{\tau_i} = \left[d_{ij_1 1}, ..., d_{ij_\tau \tau} \right]$, and the history of college signals, $G_i^{\tau_i} = \left[G_{ij_1 1}, ..., G_{ij_\tau \tau} \right]$. Given this structure, the average income gains (in terms of first wages) of getting a college degree can therefore be exactly decomposed into three effects:

$$\begin{aligned} w_{ij\tau} - w_{i0\tau} &= \left(\beta_j^w - \beta_0^w \right) X_i^w + U_{ij0}^s - U_{i00}^s \\ &+ \left(\prod_{s=0}^{\tau_i} \theta_{k_s j} - 1 \right) U_{ij0}^s \\ &+ H \left(U_{i0}^s, G_i^{\tau_i}, X_i, D_i^{\tau_i} \right) \end{aligned}$$

where $k_s = k \Leftrightarrow d_{ik s} = 1$.

In the first line of the right hand side of the equation above, we see what I call the *credential effect*, or the difference in wages that would appear without any modification in the specific abilities, and due only to the fact that labor in sector j rewards individual characteristics in a different way than sector 0 does. In the second line we see the *human capital effect,* resulting from the fact that abilities of type j were raised by a $\left(\prod_{s=0}^{\tau_i} \theta_{k_s j} - 1 \right)$ factor in the years spent in college. Finally, $H \left(U_{i0}^s, G_i^{\tau_i}, X_i, D_i^{\tau_i} \right)$ captures the new information about abilities that was revealed by the college signals, and which is also priced in the labor market. Similar exercises could be imagined for dropouts, that would end up competing by the sector 0 jobs (so that the credential effect would be zero), but would still display eventual gains from going to college in the form of an increase in human capital and more accurate information about talents.

The decision problem

As mentioned before, agents are assumed to be risk neutral and to maximize the expected present value of their income flows in the T years after the moment they finish high-school. Moreover, whenever an individual starts working, I assume he/ she stays in the labor market for the rest of his/ her active period of life, giving up the chance of eventually going to college and changing careers (this is somehow coherent with the assumption that the labor market does not bring relevant information about one's comparative advantages).

If an individual has already gone to the labor market, after τ_i periods in college, his/her expected value of income is:

$$
\begin{aligned}
W_{it}\left(j, I_{i\tau_i}\right) &= \sum_{s=t}^{T} \delta^{(s-t)} \left[\left(s-\tau_i\right) A_j + \beta_j^w X_i^w + U_{ij\tau_i}^s\right] \\
&= \left[\frac{1-\delta^{T-t+1}}{1-\delta}\right] \left(\left(t-\tau_i\right) A_j + \beta_j^w X_i^w + U_{ij\tau_i}^s\right) \\
&\quad + \delta \left[\frac{1-\delta^{T-t+1}}{(1-\delta)^2} - \frac{(T-t+1)\,\delta^{T-t}}{1-\delta}\right] A_j
\end{aligned}
$$

On the other hand, if the person is still enrolled in college and considers to choose major j in that period, he/she pays the cost C_{ijt} (which includes direct and psychic costs), and keeps it open the possibility of eventually graduating in some major with the correspondent payoff, i.e.

$$V_{ijt}\left(I_{it}\right)=-C_{ijt}+\delta E\left[\max_{d_k}\sum_{k=0}^{J}d_{ikt+1}V_{ikt+1}\left(I_{it+1}\right)|I_{it}\right] \tag{1.11}$$

We can then state the maximization problem as:

$$\begin{aligned} & \max_{\{d_{ijt}\}_{j=0}^{J}}\sum_{j=0}^{J}d_{ijt}V_{ijt}\left(I_{it}\right) && (1.12)\\ s.t. \quad : \quad & I_{it+1}=I_{t+1}\left(I_{it},G_{ijt},\epsilon_{it},d_{ijt}=1\right) \\ : \quad & \sum_{j=0}^{J}d_{ijt}=1 \\ : \quad & d_{i0t}d_{ijt+1}=0;\forall j>0 \end{aligned}$$

The first restriction shows the law of motion of the information set, which may vary with the current choice of major. The second forbids agents to take more than one choice per period. The third constraint says that one cannot leave college today and come back in the next period. Finally, I impose that graduation in college takes three periods (because in most 4-year programs students do not have to declare a major at the end of the first year), so that at the beginning of the third period $V_{ijt}\left(I_{it}\right)=-C_{ijt}+\delta E\left[W_{it+1}\left(j,I_{it+1_i}\right)|I_{it}\right]$ for every $j>0$. Notice that , by construction, $V_{i0t}\left(I_{it}\right)=W_{it}\left(0,I_{it}\right)$ in every period.

The exercise is a standard version of dynamic discrete choice models, but contains some elements not so often found in the literature. Besides the fact that learning is correlated and is about a hidden variable that also varies over time, the solution to this question clearly incurs in a dynamic selectivity problem that any empirical analysis must be aware of. The difficulty is that even though U_{i0}^{s} is randomly distributed across the population, the subsample that is observed taking classes in a given major probably had particularly good draws of this vector (a least in the coordinate correspondent to that major), which indicated he/she could be a good professional in that career. As the results of college experience are revealed, those with a really good perceived ability in the previous period are more willing to tolerate bad realizations of the new signals than those who were at the margin between taking that major or doing something else, thus biasing even more the distribution of U_{it}^{s} in that class of students. Finally, in the same cohort we have people with different previous decision histories, and therefore with information sets that evolved according to different laws of motion. This difficulty certainly plays a role in the choice of the estimation procedure suited to perform the empirical analysis in the next section.

The no-switching majors rule

The model just introduced rationalizes the decision of going to college in a specific career, providing a reasonable mechanism which justifies why agents may optimally want to review their previous decisions, based on the motion of the state variables over time. The no-switching majors rule is a restriction on the choice set of the individuals which limits the options after the first period to a binary decision between staying put in the same major or dropping out of college. In most countries, students have some sort of limitation in their abilities to change majors during college, or to use credits previously earned in other majors if they decide to change their fields of study. Evaluating this type of policy is not trivial, because the treatment in principle affects the whole population, and in this sense it is similar to the problem proposed by Marschak (1953), who wanted to estimate the impact of an increase in a commodity tax on its demand (likewise, the increase in taxes affects everybody and reduces the number of choices available to the agents). Like Marschak, our method of evaluation involves the estimation of structural parameters that determine choices in an unrestricted environment, and then use these parameters to simulate how the individual behavior and welfare would change after the restriction is imposed, except that in our case the behavior is described by a dynamic problem and the choice set by a discrete number of actions.

In practice, the first step is to solve the dynamic discrete choice problem both in the unrestricted and in the restricted cases, for every member of the population, and then find ways to aggregate the solutions to these problems in synthetic measures of the policy effect. To illustrate how the policy may affect individuals, let us assume that there are only two types of specialized occupations, say scientific and non-scientific, and the unskilled positions, and let us also reduce the number of decision periods to two. In both the restricted and unrestricted environments, the period 0 decision set of all individuals has three elements, $j = 0, 1, 2$, and choice 0 in this period implies that individuals have no further decision to make. Those who choose either 1 or 2 start college in period 0, and have to decide whether to complete it in period 1 or to dropout. The difference now is that in the restricted case those who want to stay in college must keep the same major as the one previously declared, whereas in the unrestricted case they are free to move to the other major.

The analogue of equation (1.11) in this case is:

$$V_{ijt}^{R}\left(I_{it}\right) = -C_{ijt} + \delta E\left[\max_{d_{ijt+1}}\begin{pmatrix} d_{ijt+1}V_{ijt+1}^{R}\left(I_{it+1}\right) + \\ \left(1-d_{ijt+1}\right)W_{it+1}\left(0, I_{it+1}\right)\end{pmatrix}|I_{it}\right] \quad (1.13)$$

The gains of going to college in a given major, as measured by the difference between the value function associated to a specific major and the value of going to the labor market can be written as:

$$V_{ijt}\left(I_{it}\right) - W_{it}\left(0, I_{it}\right) =$$

$$\begin{aligned}
&\left[-C_{ijt} - w_{i0t}\left(I_{it}\right)\right] + \\
&\delta\left[W_{it+1}\left(0, I_{it+1}^{(j)}\right) - W_{it+1}\left(0, I_{it}\right)\right] + \\
&\delta E\left[\max\left\{0; \max_{d_k} \sum_{k=0}^{J} d_{ikt+1}\left(V_{ikt+1}\left(I_{it+1}^{(j)}\right) - W_{it+1}\left(0, I_{it+1}^{(j)}\right)\right)\right\} | I_{it}\right]
\end{aligned}$$

$$V_{ijt}^{R}\left(I_{it}\right) - W_{it}\left(0, I_{it}\right) =$$

$$\begin{aligned}
&\left[-C_{ijt} - w_{i0t}\left(I_{it}\right)\right] + \\
&\delta\left[W_{it+1}\left(0, I_{it+1}^{(j)}\right) - W_{it+1}\left(0, I_{it}\right)\right] + \\
&\delta E\left[\max\left\{0; \left(V_{ijt+1}\left(I_{it+1}^{(j)}\right) - W_{it+1}\left(0, I_{it+1}^{(j)}\right)\right)\right\} | I_{it}\right]
\end{aligned}$$

In this form, it is easy to identify how the policy affects the individual welfare. In both expressions, the first two lines of the right hand side are identical, and only the last is different. In the first line, we see the total cost of going to college, which contains the costs associated with the respective major plus the foregone labor income in that period. In the second line, we have the (discounted) gain in the continuation value of income in the unskilled sector, which is likely to be positive, as experience in major j also affects abilities of type 0 and can make workers more productive in that sector.

The third line resembles the familiar expression found in finance for option values. Indeed, it captures the fact that only individuals currently enrolled in college will have the option of eventually staying in college from the next period on, and it is in this term that the no-switching rule affects the payoffs. This is equivalent to say that the policy effect could be measured by the difference in the average option value of college in these two circumstances.

The Graph A1 shows the solution to the agent's problem in these two environments in period 1 for an hypothetical individual who have chosen to take classes in major 2 at period 0, and observed shocks $\epsilon_{i1} = (\overline{\epsilon}_{i11}, \overline{\epsilon}_{i12})$. In part (a), the optimal choice conditional on the realization of the signal G_{i21} would be the same in the two situations if $G_{i21} \in \left(-\infty, G\_\right)$ or $G_{i21} \in (G_{+}, \infty)$, but if it takes values in $\left(G\_, G_{+}\right)$, then the optimal choice in the unrestricted world is no longer feasible if the no-switching majors rule is imposed. The ex-post loss caused by the constraint can be measured by the distance between the curves describing the restricted and unrestricted values in period 1 for every realization of G. Now, if we move back to period 0 we see that the ex-ante reduction in the payoffs associated to choice 2 are a weighted average of the discounted values of the period 1 valuation, for every realization $G_{i21} \in \left(G\_, G_{+}\right)$, with the weights given by the distribution of G conditional on I_{it}, $F_{G|I_{i1}}$. In part (b), we see that heterogeneity in observables, X_i, unobservables, U_{i1}^{s}, and shocks, ϵ_{i1} all cause the intercepts of the lines V_{ij1} to shift, with no modification on their slopes. In particular, it may be the case that for some combinations of $\left(X_i, U_{i1}^{s}, \epsilon_{i1}\right)$, not all of the choices are relevant to the agent's problem, since they would not be optimal for any realization of G_{i21}.

Finally, and because in period 0 ϵ_{i1} is still unknown to the agents and can assume values in the whole space $\Re^{J}$, we have that the payoffs associated to all choices $d_{ij0} : J > 0$ are affected by the imposition of the no-switching majors rule, since changes of major could always be desirable with positive probability.

The only group that is really unaffected by this constraint is the subpopulation of individuals that would not go to college even in the unrestricted case. The reason is that while the payoffs of start working right after high-school do not change with this rule, all of the payoffs associated with the different majors diminish, since the continuation value of these options are now maxima subject to an extra restriction which is binding for some realizations of ϵ_{i1}. Therefore, if not going to college is the optimal choice in the unrestricted case, so it is in the restricted one, being the expected present value of income equals to $W_{i0}(0, I_{i0})$ in both cases. This fact is interesting (and generalizes to more realistic problems with many majors and periods), because it allows us to separate the universe of agents in a fraction that is potentially affected by the policy, and its complement which is unaffected, suggesting the possibility of measuring the effects of the "treatment on the treated " and the "average treatment effect", similar to the ones defined in the experimental literature.

1.3 Empirical analysis

1.3.1 The data

The dataset used to estimate the model above and simulate the no-switching majors rule is the National Longitudinal Study of the High-School Class of 1972 (NLS-72, for short), which follows a randomly sampled cohort of 22654 American high-school seniors[2] for 14 years, in a series of 7 interviews with retrospective

2. To be precise, the sample consists of an stratified sample of schools in which the country was divided into 600 areas, and two schools were sampled at random in each region. Then, for each school a random sample of 18 students was selected, and each individual was surveyed and completed a battery of tests that included questions in mathematics, reading, mosaic comparisons, vocabulary, picture numbering and letter grouping, besides information on the students' academic performance.

questions to fill in the gaps in order to have a complete history of educational and professional choices of the participants. The dataset is also complemented by post high-school transcript files of those who took higher education courses afterwards, a battery of cognitive tests applied to the students before the first wave of interviews (which is particularly good, since the results are not distorted by significant heterogeneity in age and level of schooling), and information about the school, military service, and other data they could collect from other sources. The questionnaires are especially rich in information about majors taken, dropouts, and characteristics of the jobs for those who worked in this period, as well as demographic characteristics. The first five waves of this panel contain similar questionnaires, but critical methodological changes took place in the last survey. This fact, together with the sharp increase of attrition verified between the last two waves of the panel[3], implied in the exclusion of this last wave of the analysis.

In Table 1.1 we see the demographic profile of college graduates, college dropouts and high-school graduates. According to the table, women are less likely to go to college than men, and among college students, the white are less likely to graduate than the non-white. When we divide the sample by the maximum level of schooling of the individual's parents, a clear pattern emerges, suggesting that the children of more educated parents have a much higher probability of going to college and complete it than the children of low educated people. Finally, we also see that agents who finish high-school at earlier ages are more likely to go to college, whereas older high-school seniors are more likely to start working right after graduation in this cycle.

3. The data is carefully collected, with a loss of less than 20% of the individuals between 1972 and 1979. In the last wave, however, only 65% of the original sample was present, which is still low for this type of data, but may bring extra selectivity problems to the analysis.

Table 1.1: Demographic profile of college graduates, college dropouts, and high-school graduates

	Total	College Graduates	College Dropouts	High-School Graduates	Not Identified
Gender					
women	50.5	25.4	30.0	31.7	12.9
men	49.5	28.0	30.8	27.5	13.7
Race					
white	22.4	18.1	34.6	31.1	16.2
non-white	77.6	29.4	29.2	29.1	12.4
Parent's education					
< high-school	21.2	11.4	25.5	46.2	16.9
high-school	36.0	18.7	30.6	35.8	14.9
some college	22.0	31.6	36.1	20.6	11.7
4-year college	11.1	50.7	31.2	10.9	7.2
graduate school	9.7	54.7	28.7	7.9	8.7

Source: NLS-72

In Table 1.2, we see the college enrollment by groups of majors[4]. In every year of the panel, scientific courses seem to attract about a half of the students, but as the time passes, enrollment in non-scientific courses increases systematically. The proportion of degrees issued in scientific majors is also slightly higher than the non-scientific ones, suggesting the students of these majors are more likely to finish their programs.

Although students may change majors at any point in time, we can just add up the observed transitions in order to have a synthetic measure of the probability that a student enrolled in a given major moves to a different position in the next

4. the exact composition of these groups are presented in Table A1, in the appendix

Table 1.2: Enrollment in college, by declared major

Year	Natural Sciences	Social Sciences	Sciences Total	Business	Education	Non-Sciences Total
1972	29.5	24.4	53.9	19.3	26.8	46.1
1973	27.4	25.3	52.7	21.3	26.0	47.3
1974	29.1	26.4	55.5	21.8	22.7	44.5
1975	28.5	27.1	55.6	22.7	21.8	44.4
1976	31.0	24.9	55.9	24.1	20.0	44.1
1977	28.7	23.4	52.1	25.4	22.5	47.9
1978	28.7	21.6	50.3	28.0	21.6	49.7
1979	29.7	20.6	50.3	29.6	20.2	49.7
Major of Graduation	25.3	28.9	54.2	19.2	26.6	45.8

Source: NLS-72

period. In the third table, we see that, overall, Natural Sciences students are more likely to stay put in their current choices, whereas social sciences and business students are more willing to dropout of college. The fraction of the students that stay in college but change majors between periods ranges from 7.3% (business) to 14% (education), and except for those coming from Education, Business seems to be the main destination of the movers. Table 1.4 complements these information by showing that dropout rates decrease consistently over the first 5 years of college (which is supposed to last 4 years), and then increase among the delayed students. Switching is particularly intense between the second and third years of college, and the rates are more volatile in non-scientific courses.

As mentioned above, the dataset contains an interesting battery of tests that quantifies different abilities of the students. Even though these talents cannot be seen as the empirical counterparts of the hidden abilities modeled in the previous

Table 1.3: Transition probabilities, average across years

From/ To	Dropout	Natural Sciences	Social Sciences	Business	Education
Nat. Sciences	30.2	60.3	3.3	3.1	3.1
Soc. Sciences	31.8	2.6	56.3	4.1	5.3
Business	36.4	2.0	3.1	56.3	2.2
Education	30.9	5.4	5.6	3.0	55.1

Table 1.4: Transitions conditional on the number of years of college education

Grade Transition	From Sciences, to:			From Non-sciences, to:		
	Sciences	Non-sciences	dropout	Sciences	Non-sciences	dropout
1st-2nd	72.1	10.9	16.9	10.8	69.1	20.0
2nd-3rd	68.6	13.4	18.0	17.4	60.9	21.7
3rd-4th	74.8	10.4	14.8	10.8	72.7	16.5
4th-5th	76.2	10.4	13.3	7.7	77.6	14.8
5th-6th	77.0	9.1	13.9	7.9	79.0	13.1
6th-7th	73.4	9.9	16.7	7.0	77.0	16.1
7th-8th	73.3	5.1	21.6	7.2	71.3	21.5

Source: NLS-72

section, it may bring us some illustration about how important the features of the model can be. The tests are divided in questions and exercises about Mathematics, Vocabulary, Mosaic Comparison, and Figure Number, and Table 1.5 below shows the estimated correlation between these parts. The first interesting observation is that correlation is in general positive, but relatively low, except in the case of Math-Vocabulary, where it gets close to 50%, suggesting students may have different strengths and weaknesses that could not be captured by a scalar index of human capital (as in Arcidiacono, 2004). Second, we see that the subsample of individuals who have switched majors during college displays systematically higher correlations than those who did not switch, which is in the spirit of the learning process assumed before, where correlations between talents may play an important role in predicting switching.

Table 1.5: Correlations between test scores

Total	Math	Mosaic Comparison	Figure Numbering	Vocabulary
Math	1			
Mosaic Comparison	0.40	1.00		
Figure Numbering	0.42	0.38	1.00	
Vocabulary	0.59	0.28	0.31	1.00
Non-Switchers				
Math	1.00			
Mosaic Comparison	0.34	1.00		
Figure Numbering	0.31	0.27	1.00	
Vocabulary	0.49	0.18	0.20	1.00
Switchers				
Math	1.00			
Mosaic Comparison	0.39	1.00		
Figure Numbering	0.41	0.38	1.00	
Vocabulary	0.57	0.28	0.30	1.00

Another key aspect of the theoretical model that enriches the analysis is the presence of a dynamic selection bias, such that as new decisions are made, the survivors are likely to be the individuals who either had a very high perceived ability at time 0, or who have received good signals afterwards. The Graph A3 aims to illustrate this process, by constructing the average of the test scores in the population of students who stay in college up to a given period. All of the scores were standardized such that the mean was brought to zero and the variance to one to facilitate comparisons. First, notice that the average test scores in all batteries were higher among those who entered in college than those who did not (since the weighted average of both should be zero, this is seen by the fact that the averages in period one are all positive). As time passes, the students with the lowest test scores have higher probability to dropout of college, so that the average among the stayers increases with the years. Second, different majors select people with different specific abilities. While the Natural Sciences students display the highest average in Math, it is the Social Sciences students who have the best vocabulary skills.

1.3.2 The variables used in the econometric exercise

The goal of this part is to describe how the variables used in the estimation of the structural model introduced in Section 1 were constructed. The data needed in this procedure contains the observable characteristics of the agent determined outside of the model, X_i, the signals that arrive after the first period to those who decided to go to college, G_{ijt}, the wages earned after the last period of education, w_{ijt}, and the decision history of the sample members, d_{ijt}. Notice that the observables are divided according to their role in the model, so that variables $X_i^w \subseteq X_i$ denote attributes that are assumed to influence wages, $X_i^c \subseteq X_i$ help to determine the costs of going to college, and $X_i^g \subseteq X_i$ affect the signals.

Observables:

All of the observable traits are constructed directly from answers to specific questions of the questionnaires. The exogenous determinants of wages are sex (dummy variable that is 1 iff the individual is male), race (dummy variable indicating whites), and experience, defined as the number of years after the last period of education.

The cost components include also sex and race, but it also contains the highest educational level of the parents (the reason is that more educated parents may help the students to learn, have books and other materials needed in college already at home, etc). Parental education was constructed as follows: the surveys ask, for each parent, to which of 5 categories (less than high-school, high-school, some college, college, graduate level) the parent belongs. I then took either the maximum of this variable between the two parents, or the education of the non missing parent, when information about the other was missing. Finally, I assumed that a linear index ranging from 1 to 5 could summarize the information contained in these 5 discrete, qualitative levels.

In the case of grades, I included only sex and race as potential explanatory arguments.

Since the same questions were asked repeatedly in the successive waves of the panel, I tried to solve eventual contradictions by using the most frequent information when at least $N-1$ waves agreed about the respective answer.

Wages:

In the survey, participants are asked about their weekly labor income with the typical reference date being at the beginning of fall. Since the empirical counterpart of the model takes in general one year as the correspondent to one period, I multiplied the reported income by 365/7 in order to have an approximation of the yearly earnings. Zero values for this variable were counted as missing information, and the annualized income was divided by 10,000, because the empirical estimation heavily relies on numerical approximations that may become more accurate if the magnitudes of the different variables involved in the calculations were similar.

Decisions:

Every person was given a value for three variables, $d_{ijt} : t = 1, 2, 3$. I first used all of the information I found in the surveys to separate the individuals who had never gone to college. Because the questionnaires are quite long, it is plausible that some college students decide to say they are not in college just to avoid answering the educational part of the survey. For this reason, I created a category of undefined college status to label all of the people who showed contradictory answers to these questions. Moreover, some people also reported to be in college, but refused to tell the major, and these were also included in the undefined block. Finally, I accepted as valid college enrollment in major j all individuals who reported to be taking classes either in a 4-year college program, or in a 2-year academic (mostly community) college program, or in a 2-year program whose credits could be used in a 4-year college program. The reason to include 2-year students is that, since they can potentially use their credits in a 4-year program (and since I am not modeling explicitly the choice of school), they still keep the option of transferring and eventually getting a 4-year college degree. Finally, the key variable used to classify students into majors was the field of study (FOS) asked in every survey, and for every year. These fields were first grouped into 79 broader categories, detailed described in Table C1 in the appendix, with the respective original FOS codes that compose each group. The 79 "majors" were than aggregated into 4 areas: Natural Sciences (and other courses intensive in Math and/ or Biology), Social Sciences (together with humanities and arts), Business (and communication), and

Education (every course oriented to form high-school teachers). Due to computational restrictions, I further grouped the first two of these into a larger definition of scientific majors, and the last two were labeled as Non-Sciences.

There is an important detail worth to be mentioned here. College programs typically last four years, whereas in general students declare majors only three times. Instead of extending the model to have a fourth decisional period (where the only choices would be stay put in college or dropout), I assumed that for those who went to college, the two first years corresponded in fact to one period, which generates an asymmetry whose consequences were not yet investigated, but should certainly be in future steps of this research. Crucial to this decision was the enormous computational burden that an extra period would bring, with no clear compensation in terms of results.

Signals:

The main source of information available in the data about information that arrived during college and that could influence the students' decisions is the transcript file. There, we can find the grades obtained by the students during college, for every single discipline, which is also labeled according to its main subject, as defined by the Classification of Instructional Programs (CIP), together with the exact period the course was taken, and the name and FICE[5] code of the school, among other relevant items. However, there are two difficulties to be circumvented in order to create an empirical counterpart that satisfies the model proposed in this article. First, the choice of school is not explicitly modeled, and it is well known that there is selectivity in the matching between students and schools, such that better schools usually get better students and vice-versa. In this sense, a very

5. Federal Interagency Committee of Education

good grade in a major may be a good signal that an individual is suited to that career if it comes from a good school, but a poor signal if it comes from a school that does not have a good reputation. Second, among schools of the same level the grading policy may be heterogeneous. If a school has its marks very concentrated around one point (say, the maximum), getting a grade close to that point is not informative about your talent, whereas if the grading distribution is more balanced, the opposite happens.

The solution I propose is to compare individual grades obtained in courses classified as typical of the declared major with grades obtained in the remaining courses. If in a given period the grades of the first are much better than the grades of the last, this means that the student indeed may have a clear comparative advantage in that major, and since both grades are obtained in the same school, this procedure minimizes the influence of specific grading policies on the signal.

To be precise, I first constructed a subsample of college graduates, and assign to each observation a major, according to the broader 79-category definition. Then, I merged these data with the transcript file, making sure I kept only transcripts from the college the individual got his/ her first degree (each person may have up to 7 transcripts in the dataset). The third step is to construct a list of course subjects each individual studied, being cautious to avoid repetition (e.g., if one took calculus I and calculus II, both courses will appear in the transcript with the same CIP code). Then, within each major I took the frequency that a CIP code appears in the individual transcripts. If it is greater than 25%, I classify this CIP as typical of that major, and if not, as a distribution. Next, and going back to the whole subsample of college students, I computed, for each period, the partial GPA relative to the typical courses and the partial GPA of the non-typical courses, and the difference between these two averages would be my first approximation to the Signal students received in that period. The problem with this index is that it only provides an indicator of relative ability, that could be important to decide whether to stay in the same major or to switch, but does not contain much information about the absolute level of that ability, which could be important in the decision of staying in college or dropping out (and as we saw in the last graph, individuals with lower general abilities are more likely to leave college). For this reason, I defined the average between the Math and Vocabulary scores in the battery test as a measure of general ability, and the variable I used as the signal ended up with the form:

$$G_{ijt} = \left(\frac{Math + Vocab}{2}\right)\left(1 + \frac{GPA_{jt} - GPA_{-jt}}{GPA_t}\right)$$

where GPA_{jt} and GPA_{-jt} stand for the average grade obtained in courses typical and non-typical to major j, respectively, and GPA_t is the total GPA obtained in period t. Finally, some individuals reported to be attending more than one college institution in the same period. In these cases, I first discarded information from 2-year colleges when the person was enrolled in at least one 4-year institution. If more than one school was a 4-year college, I computed the signal in both of them, and then took the average weighted by the number of courses taken in each institution.

1.4 Empirical strategy and results

In this section I propose a method to estimate the theoretical model described in section 2, and present the respective results obtained from the NLS-72 data. The estimation is carried out by Markov Chain Montecarlo (MCMC), which has three main advantages over other methods: (i) it does not use minimization, which is computationally burdensome; (ii) it allows us to use eventual prior information to complement the distribution functions that form the likelihood of the model; and (iii) estimation and simulation can be performed together in an integrated algorithm.

1.4.1 Empirical strategy

MCMC - overview

As every Bayesian method, MCMC treats a $K-$ dimensional vector of parameters, Υ, as if each of its components were a random variable itself. The standard algorithm consists of a sequence of draws of each element (or block of elements), Υ_k from its posterior distribution, conditional on the other parameters, $\Upsilon_{k^-} = (\Upsilon_1, ..., \Upsilon_{k-1})$ and $\Upsilon_{k^+} = (\Upsilon_{k+1}, ..., \Upsilon_K)$ and the data, Z. In our case, we also use data augmentation to facilitate the procedure, i.e., we not only draw the parameters that appear in the model, but also some of the latent variables (as if they were individual-specific parameters). A typical example of this class of algorithms would be:

(i) set the seed $\Upsilon^{(0)}$

(ii) start iteration 1 by drawing $\Upsilon_1^{(1)}$ from the conditional posterior distribution $f\left(\Upsilon_1^{(1)}|\Upsilon_{1^+}^{(0)}, Z\right)$, then sample $\Upsilon_2^{(1)}$ from $f\left(\Upsilon_2^{(1)}|\Upsilon_{2^-}^{(1)}, \Upsilon_{2^+}^{(0)}, Z\right)$, and at every k, sample $\Upsilon_k^{(1)}$ from $f\left(\Upsilon_k^{(1)}|\Upsilon_{k^-}^{(1)}, \Upsilon_{k^+}^{(0)}, Z\right)$

(iii) at iteration m, repeat the step above, using the most recent sampled values available for the elements of Υ to sample $\Upsilon_k^{(m)}$ from $f\left(\Upsilon_k^{(m)}|\Upsilon_{k^-}^{(m)}, \Upsilon_{k^+}^{(m-1)}, Z\right)$

The statistical theory behind this system says that if the conditional posteriors form proper densities, this sequence of draws converges to the ergodic distribution of the chain, which is precisely the full posterior distribution of (Υ, Z) [6]. However, because the seed used to start the chain does not need to belong to the "heart" of the support of posterior (the subset of its support that contains the most frequent realizations of Υ), the first iterations (called the "warm up phase") of the chain not necessarily belong to its ergodic distribution, and must be discarded in the statistical analysis.

The posterior distribution of the parameters conditional on the data is obtained by applying the Bayes theorem to the joint distribution of Z and Υ :

$$f(\Upsilon|Z) = \frac{f(\Upsilon, Z)}{f(Z)} = \frac{f(Z|\Upsilon)\, f(\Upsilon)}{f(Z)} \propto f(Z|\Upsilon)\, f(\Upsilon)$$

In words, the conditional distribution of Υ on Z is proportional to the product of the likelihood function, $f(Z|\Upsilon)$, and the unconditional (also called prior) distribution of the parameters, $f(\Upsilon)$. Moreover, we can also rewrite it as:

$$\begin{aligned} f(\Upsilon|Z) &= f(\Upsilon_k|\Upsilon_{k-}, \Upsilon_{k+}, Z)\, f(\Upsilon_{k-}, \Upsilon_{k+}|Z) \\ f(\Upsilon) &= f(\Upsilon_k|\Upsilon_{k-}, \Upsilon_{k+})\, f(\Upsilon_{k-}, \Upsilon_{k+}) \\ &\Rightarrow f(\Upsilon_k|\Upsilon_{k-}, \Upsilon_{k+}, Z) \propto f(Z|\Upsilon)\, f(\Upsilon_k|\Upsilon_{k-}, \Upsilon_{k+}) \end{aligned}$$

Now, the econometrician can include in the function $f(\Upsilon_k|\Upsilon_{k-}, \Upsilon_{k+})$ all of the information that may be relevant and not already present in the likelihood, $f(\Upsilon|Z)$. Finally, the right hand side of the expression above can always be written as the product of a (minimal) term that includes Υ_k by a term that does not include

6. For a comprehensive textbook about MCMC, see Lopes and Gammermann (2007).

it, i.e.,

$$f\left(\Upsilon_k|\Upsilon_{k-},\Upsilon_{k+},Z\right)=K\left(\Upsilon_k,\Upsilon_{k-},\Upsilon_{k+},Z\right)\overline{f}\left(\Upsilon_{k-},\Upsilon_{k+},Z\right)$$

If the expression for $K\left(\Upsilon_k,\Upsilon_{k-},\Upsilon_{k+},Z\right)$ coincides with the kernel of a known distribution in Υ_k for given $\left(\Upsilon_{k-},\Upsilon_{k+},Z\right)$, then we can complete this distribution and have directly the conditional posterior $f\left(\Upsilon_k|\Upsilon_{k-},\Upsilon_{k+},Z\right)$, from which we can sample Υ_k. This is called the *Gibbs sampler procedure*, and should be used whenever it is possible, since it provides the fastest convergence of the chain.

If $K\left(\Upsilon_k,\Upsilon_{k-},\Upsilon_{k+},Z\right)$ does not form the kernel of a known density, however, then some other sampling strategy must be adopted, the most popular being the Metropolis -Hastings (MH) algorithm. The MH technique is essentially an importance sampling procedure, where, at iteration m, the (block of) parameters $\widetilde{\Upsilon}_k$ are sampled from some known distribution $f^p\left(\Upsilon_k|.\right)$ proposed by the analyst, and then a criterion consistent with the ergodic properties of the chain is used to decide whether to accept the new draw or to keep the previous draw in that iteration, i.e., whether $\Upsilon_k^{(m)}=\widetilde{\Upsilon}_k$ or $\Upsilon_k^{(m)}=\Upsilon_k^{(m-1)}$. In the general case, it is shown that the proposed density function may depend not only on the remaining parameters and data, but also on the previous draw of the parameter k, which means that

$$f^p\left(\Upsilon_k^{(m)}|.\right)=f^p\left(\Upsilon_k^{(m)}|\Upsilon_{k^-}^{(m)},\Upsilon_{k^+}^{(m-1)},\Upsilon_k^{(m-1)}Z\right)$$

, and the criterion should be

$$\text{let } \alpha = \frac{f\left(\widetilde{\Upsilon}_k|\Upsilon_{k-}^{(m)},\Upsilon_{k+}^{m-1},Z\right)}{f^p\left(\widetilde{\Upsilon}_k|\Upsilon_{k-}^{(m)},\Upsilon_{k+}^{(m-1)},\Upsilon_k^{(m-1)}Z\right)} / \frac{f\left(\Upsilon_k^{(m-1)}|\Upsilon_{k-}^{(m)},\Upsilon_{k+}^{m-1},Z\right)}{f^p\left(\Upsilon_k^{(m-1)}|\Upsilon_{k-}^{(m)},\Upsilon_{k+}^{(m-1)},\widetilde{\Upsilon}_k Z\right)}$$

$$\Upsilon_k^{(m)} = \begin{cases} \widetilde{\Upsilon}_k \text{ if } \alpha \geq U \\ \Upsilon_k^{(m-1)} \text{ if } \alpha < U \end{cases}$$

where U is a random number sampled from a uniform distribution. It is clear that if $\alpha \geq 1$, then $\Upsilon_k^{(m)} = \widetilde{\Upsilon}_k$ regardless of the realization of U, since this is the upperbound of the uniform's support. It is also easy to see that if

$$f^p\left(\widetilde{\Upsilon}_k|\Upsilon_{k-}^{(m)},\Upsilon_{k+}^{(m-1)},\Upsilon_k^{(m-1)}Z\right) = f\left(\widetilde{\Upsilon}_k|\Upsilon_{k-}^{(m)},\Upsilon_{k+}^{m-1},Z\right)$$

then $\alpha = 1$ and all draws are accepted, which makes the Gibbs sampler a particular case of the MH.

Even though every distribution f^p with a support at least as large as the support of $f(\Upsilon_k|.,.,.)$ is a valid choice for the proposal, some of them are more convenient, in order to speed up convergence. Ideally, we should pick f^p as similar as possible as $f(\Upsilon_k|.,.,.)$, but similarity is not easily defined. There are two ways of electing f^p that are very common in the literature. The first, called random-walk, consists on a symmetric distribution (if it is possible to find one) centered on the previous draw of the respective parameter, i.e.:

$$f^p\left(\widetilde{\Upsilon}_k|\Upsilon_{k-}^{(m)},\Upsilon_{k+}^{(m-1)},\Upsilon_k^{(m-1)}Z\right) = h\left(\widetilde{\Upsilon}_k - \Upsilon_k^{(m-1)},\Upsilon_{k-}^{(m)},\Upsilon_{k+}^{(m-1)},Z\right)$$

, for some function h. The two advantages of this function are that (i) α simplifies to

$$\alpha = f\left(\widetilde{\Upsilon}_k|\Upsilon_{k^-}^{(m)},\Upsilon_{k^+}^{m-1},Z\right)/f\left(\Upsilon_k^{(m-1)}|\Upsilon_{k^-}^{(m)},\Upsilon_{k^+}^{m-1},Z\right)$$

, and (ii) h usually has some hyperparameters that characterize it, and which can be calibrated in order to keep the rate of acceptance of new draws satisfactory. The leading example of this class is a normal distribution with mean $\Upsilon_k^{(m-1)}$, and variance V. If V is low, the new draw will probably be very close to the previous draw $\left(V \rightarrow 0 \Rightarrow \widetilde{\Upsilon}_k \rightarrow \Upsilon_k^{(m-1)}\right)$, and α will approach 1, increasing the acceptance rate of new draws. The drawback is that if S is too low, the sequence of draws $\left(\Upsilon_k^{(1)},\Upsilon_k^{(2)}...\right)$ will take more iterations to cover the whole parameter space, and if the initial value is far from the interval that concentrates most of the realizations of this variable, it will take longer to find a threshold M after which the realizations of $\left\{\Upsilon_k^{(m)}\right\}_{m>M}$ can be a reasonable approximation to the ergodic distribution of Υ_k. Geweke (1989) suggests that an acceptance rate around 45% for scalar variables and 21%-25% for blocks of parameters (depending on the dimension of this block) maximize the speed of convergence of the chain if the proposed distribution is the normal - random walk.

The second popular choice of f^p relies directly on the functional form of the posterior. If $f\left(\Upsilon_k|\Upsilon_{k^-}^{(m)},\Upsilon_{k^+}^{m-1},Z\right)$ contains a piece that can be identified as the kernel of a known distribution, then we should importance sampling from this distribution, i.e.: if

$$f\left(\Upsilon_k|\Upsilon_{k^-}^{(m)},\Upsilon_{k^+}^{m-1},Z\right) = K\left(\Upsilon_k|\Upsilon_{k^-}^{(m)},\Upsilon_{k^+}^{m-1},Z\right)h\left(\Upsilon_k,\Upsilon_{k^-}^{(m)},\Upsilon_{k^+}^{m-1},Z\right)$$

, and K is the kernel of the density $\widetilde{f}$, then we should make $f^p = \widetilde{f}$. The idea is that since K comes directly from the true posterior, it could be more similar to

$f\left(\Upsilon_k|\Upsilon_{k^-}^{(m)},\Upsilon_{k^+}^{m-1},Z\right)$ than other competing alternatives. Again, by doing this α simplifies to

$$\alpha = h\left(\widetilde{\Upsilon}_k,\Upsilon_{k^-}^{(m)},\Upsilon_{k^+}^{m-1},Z\right)/h\left(\Upsilon_k^{(m-1)},\Upsilon_{k^-}^{(m)},\Upsilon_{k^+}^{m-1},Z\right)$$

In my estimation, two parameters are fixed exogenously, as if I had the prior information that they would assume a specific value with probability 1. These are the discount factor δ, set to 0.9, and the variance of the unobservable component of costs, $S = 10$. They were fixed because it helps to get a smoother evolution of the chain towards convergence, and because I still do not have a formal proof these parameters are identified from the data. I provide a small robustness check in the appendix, by replicating the main exercise of this section with different values of δ and S, and the main conclusions are not affected, although a more careful analysis of the role of these parameters should be taken in the future.

The blocks of quantities that form the chain are then the parameters

$$\Omega_0,\Lambda_0,\left\{\rho_j,A_j,\beta_j^w\right\}_{j=0}^J,\left\{\theta_j,\sigma_j,\xi_j,\beta_j^g,\beta_j^c\right\}_{j=1}^J$$

, and the latent variables $U_{i0}^s, v_{i0}, \left\{\widehat{V}_{ijt}\right\}_{j=1}^J ;_{t=1}^{\mathcal{T}}$. Here some reparameterizations are worth to notice. First, I sample $\xi_j = \frac{\varphi_j\theta_{jj}}{\sigma_j}$, instead of the original parameter φ_j. The reason is that this combination of parameters appear very often in the chain, causing the successive draws of $\left(\theta_{jj},\varphi_j,\sigma_j\right)$ to be very correlated, and therefore slowing down the process of convergence. It is easy to see that there is a 1-1 correspondence between $\left(\theta_{jj},\varphi_j,\sigma_j\right)$ and $\left(\theta_{jj},\xi_j,\sigma_j\right)$. Second, I augment the data by the $(J-1)$ latent variables $\widehat{V}_{ijt} = V_{ijt}\left(I_{it}\right) - V_{i0t}\left(I_{it}\right) : j > 0$, instead of the J variables $V_{ijt}\left(I_{it}\right)$. In this case, we know from the random utility models that a choice between J discrete actions can be fully described by $J-1$ unobservables, and the choice of $\left\{\widehat{V}_{ijt}\right\}_{j>0}$ also revealed to be more stable than the alternative $\left\{V_{ijt}\right\}_{j>0}$ (that in principle could be used, after discarding the redundant V_{i0t}).

Let Υ_i denote the subset of Υ composed by the latent variables, and Υ^* the remaining parameters. Then the individual contribution to the full posterior could be written as:

$$F_i = f_i\left(Z|\Upsilon\right) f_i\left(\Upsilon_i|, \Upsilon^*\right) =$$

$$\left[\prod_{t>\tau_i} f_w\left(w_{it}|U^s_{i0}, G_{i1}, G_{i2}, G_{i3}, X_i, \Upsilon^*, d_{ij0}=1, d_{ik1}=1, d_{il2}=1\right)\right] *$$
$$\overline{f}_G\left(G_{i2}|U^s_{i0}, v_{i0}, X_i, \Upsilon^*, d_{ij0}=1, d_{ik1}=1, d_{il2}=1\right)^{1-d_{i02}}$$
$$\left[\begin{array}{c} 1\left(\widehat{V}_{il2} > \max\left\{0, \max_{m\neq l}\widehat{V}_{im2}\right\}\right) * \\ f_{\widehat{V}}\left(\widehat{V}_{i2}|U^s_{i0}, G_{i1}, G_{i2}, X_i, \Upsilon^*, d_{ij0}=1, d_{ik1}=1\right) * \\ \overline{f}_G\left(G_{i2}|U^s_{i0}, v_{i0}, X_i, \Upsilon^*, d_{ij0}=1, d_{ik1}=1\right) \end{array}\right]^{1-d_{i10}}$$
$$\left[\begin{array}{c} 1\left(\widehat{V}_{ik1} > \max\left\{0, \max_{l\neq k}\widehat{V}_{il1}\right\}\right) * \\ f_{\widehat{V}}\left(\widehat{V}_{i1}|U^s_{i0}, G_{i1}, X_i, \Upsilon^*, d_{ij0}=1\right) * \\ f_G\left(G_{i1}|U^s_{i0}, v_{i0}, X_i, \Upsilon^*, d_{ij0}=1\right) \end{array}\right]^{1-d_{i00}}$$
$$1\left(\widehat{V}_{ij0} > \max\left\{0, \max_{k\neq j}\widehat{V}_{ik0}\right\}\right) f_{\widehat{V}}\left(\widehat{V}_{i0}|U^s_{i0}, X_i, \Upsilon^*\right) *$$
$$f_{U^s}\left(U^s_{i0}|\Upsilon^*\right) f_v\left(v_{i0}|\Upsilon^*\right)$$

where the pieces of this distribution are:

$$
\begin{aligned}
f_{U^s}\left(U_{i0}^s|.\right) &= N\left(0;\Omega_0\right)\\
f_v\left(v_{i0}|.\right) &= N\left(0;\Lambda_0\right)\\
f_G\left(G_{it}|.\right) &= N\left(\beta_j^g X_i^g+\left(\prod_{s=0}^{t-1}\theta_{k_s j}\right)\xi_j\sigma_j\left(U_{ij0}^s+v_{ij0}\right);\sigma_j\right)\\
f_{\widehat{V}}\left(\widehat{V}_{it}|.\right) &= N\left(-\beta^c X_i^c+\delta E\left(V_{it+1}|d_{it},I_{it}\right)-V_{i0t};SxI_{J-1}\right)\\
f_w\left(w_{it}|.\right) &= N\left[\sum_{j=0}^{J}d_{ij3}\left(A_{ij}\left(t-\tau_i\right)+\beta_j^w X_i^w+U_{ij3}^s;\rho_j\right)\right]
\end{aligned}
$$

where $k_s = k \Leftrightarrow d_{iks} = 1$. Furthermore, the indicator functions of $d_{ijt} = 1$, $1\left(\widehat{V}_{ijt} > \max\left\{0, \max_{k\neq j}\widehat{V}_{ikt}\right\}\right)$ impose truncations to the normal functions $f_{\widehat{V}}\left(.|.\right)$. The full posterior is then given by the expression:

$$
f\left(\Upsilon|Z\right) = f_\Upsilon\left(\Upsilon^*\right)\prod_{i=1}^{N}F_i
$$

where the prior distribution is assumed to be in general non-informative, except in the case of θ_j, which gets a uniform prior ranging from 0 to $\overline{\theta}$, with $\overline{\theta}$ being a sufficiently high real number (in practice it just intends to impose the restriction that θ is non-negative. The uniform with $\overline{\theta}\rightarrow\infty$ places a condition indistinguishable to an indicator $1\left(\theta_j > 0\right)$ in the sampling procedure). The precise expressions for the conditional posteriors of each block of parameters is left to the appendix.

After all, I was able to use the Gibbs sampler for the blocks: Ω_0, $\left\{\rho_j\right\}_{j=0}^{J}$, and the latent variables $\left(v_{i0}, \left\{\widehat{V}_{it}\right\}_{t=1}^{\tau_i}\right)$. A pure random-walk MH was used for the blocks $\left\{\beta_j^w, \beta_j^g, \theta_j, \Lambda_0{}^7\right\}$ and the latent vector U_{i0}^s, and a pure procedure taking part of the posterior as the proposed distribution was used for the blocks $\left\{\sigma_j\right\}_{j=1}^{J}$. Finally, a mixture proposal with positive probabilities for both the random walk and the "part-of-the-posterior" method was chosen to sample the blocks $\left\{\beta_j^c, \xi_j, A_j\right\}$[8].

The "Emax" problem and the convergence of the chain

In the whole literature of dynamic discrete choice models, the most difficult part is to deal with the object $E\left(V_{it+1}|d_{it}, I_{it}\right) = E\max_k\left(V_{ikt+1}|d_{it}, I_{it}\right)$, which in general does not have an analytical form (in my case only $E\left(V_{i3}|d_{i3}, I_{i3}\right)$ does), and involves a multidimensional numerical integral with $J-1$ coordinates. If the problem has finite horizon, as it is my case, we should also integrate in the temporal direction, which multiplies the dimension by the number of periods between the current and the terminal ones. In the case of the exercise presented in this paper, with only 3 choices per period and 3 decision periods, the period 0 valuations involve an 8-dimensional integral, that has to be computed twice for each individual, and for all of the N observations, at least once per update of a parameter block during a single iteration m. This is obviously computationally burdensome and every work in this area has to be careful in order to choose the integration method that is at the same time somehow precise and feasible. Solutions found elsewhere include the spline approximation of $E\max$ evaluated on a grid of possible realizations of the unobserved shocks, as proposed by Keane and Wolpin (1994), but which also required the unobservables to be non-persistent (unlike U_{i0}^s, v_{i0}); the choice of particular functions that deliver an analytical solution or a solution close to be analytical (as in Rust 1987), transformations of the maximization problem in order to obtain variables that are sufficient statistics to the $E\max$ function of to the solution of the individual maximization problem (as in Hotz and Miller,

8. In this case I first draw a uniform quantity and, if it is above a predetermined threshold, I use the random walk and otherwise the alternative method. I thank Hedibert Lopes for this suggestion.

1993), the use of previous computations of $E\left(V_{it+1}^{(m-s)}|d_{it},I_{it}^{(m-s)}\right)$ to approximate $E\left(V_{it+1}^{(m)}|d_{it},I_{it}^{(m)}\right)$ (as in Imai(2007) and Norets(2006)), among others. Unfortunately I could not find, among these solutions, one that could be directly applied to my case, and I conservatively decided to use Gauss-Hermite quadratures with 7 nodes in each dimension to approximate the $E\max$ function.

It turns out that a trick can also be used in the particular case treated here. The multidimensional Gaussian quadrature integration usually requires that, for a given integrand of the form $h\left(z_1,...,z_K;X\right)$, the integral

$$\int_K h\left(z_1,...,z_K;X\right)dz_1,...,dz_K$$

be calculated from a weighted average of the function $h\left(.\right)$ evaluated on a grid of points $\left\{\overline{z}_q\right\}_{q=1}^{Q}$ that solves the $Q^{th}-$order system of Gauss-Hermite polynomials. Therefore, in principle QxK evaluations are required to approximate this integral, but in our particular case, the integrand has the form: $h\left(\varepsilon,\epsilon;I_{it}\right)=\max\left\{\overline{V}_{i0t+1}\left(\varepsilon_{it+1}\right),\overline{V}_{i1t+1}\left(\varepsilon_{it+1}\right)+\epsilon_{i1t+1},...,\overline{V}_{iJt+1}\left(\varepsilon_{it+1}\right)+\epsilon_{iJt+1}\right\}$. We can then first condition on ε (which is one-dimensional) and compute the integral over ϵ, and then integrate the ε component out. But notice that, if on a given node $\left(\overline{z}_{q1},...,\overline{z}_{qJ}\right)$, we find out that

$$\begin{aligned}h\left(\varepsilon_{it+1},\overline{z}_{q1},...,\overline{z}_{qJ};I_{it}\right) &= \max\left\{\begin{array}{c}\overline{V}_{i0t+1}\left(\varepsilon_{it+1}\right),\overline{V}_{i1t+1}\left(\varepsilon_{it+1}\right)+\overline{z}_{q1},\\ ...,\overline{V}_{iJt+1}\left(\varepsilon_{it+1}\right)+\overline{z}_{qJ}\end{array}\right\}\\ &= \overline{V}_{ijt+1}\left(\varepsilon_{it+1}\right)+\overline{z}_{qj}\end{aligned}$$

i.e., on that combination of nodes choice j was the maximum, then it is also true that whenever we choose another $\overline{z}_{qj}' > \overline{z}_{qj}$, keeping the other coordinates constant, the maximum will still be $\overline{V}_{ijt+1}(\varepsilon_{it+1})+\overline{z}'_{qj}$, and whenever we reduce one or more of the coordinates $k \neq j$, keeping the others no bigger than $(\overline{z}_{q1}, ..., \overline{z}_{qJ})$, j will also be the maximum. In the integral procedure, we can take advantage of this property by doing the following strategy: First, sort $\overline{V}_{ijt+1}(\varepsilon_{it+1})$ from the highest to the lowest. Call the highest $l = 1$ (it does not matter its original label j). Then, find, for every node $\overline{z}_q^{l=1}$ correspondent to the respective shock associated to decision l, the thresholds that would make it no longer the maximum. By doing this, we find all combinations of the nodes $\overline{z}_{q'}^{-l}$ associated to the remaining alternatives that would be dominated by $\overline{V}_{ilt+1}(\varepsilon_{it+1}) + \overline{z}_q^l$, for every q. We can than just sum the weights associated to the combinations of $\left(\overline{z}_q^l, \overline{z}_{q'}^{-l}\right)$, and multiply it by $\overline{V}_{ilt+1}(\varepsilon_{it+1})+\overline{z}_q^l$, in order to have its contribution to the integral. After finishing with $l = 1$, go to $l = 2$ (the second highest), and find the thresholds, being careful not to include the comparisons with $l = 1$ that already proved to be dominated by the first choice, and keep going until the lowest alternative. This simple idea alone speeded up the running time of the computational code in $J - 1$ times.

The second remarkable difficulty in the estimation of the proposed model is to find proposal densities that can be efficiently used in the MH steps of the sampling procedure. The main problem found in this particular case was that one of the latent variables that had to be sampled to augment the data could not be Gibbs sampled. When only parameters that are common to the whole population have to be sampled with MH algorithms, the main consequence in terms of convergence is that it takes more iterations than the Gibbs sampler, and the literature provides a number of tricks that can be used to circumvent this problem, such as the use of random walk importance sampling with calibration of the variance to achieve the desired acceptance rate of the MCMC blocks. It turns out that it is much more difficult to monitor the evolution and the acceptance rate of individual-specific "parameters", such as the latent variables used to augment the data, and even though we can control the overall number of observations that have their latent variables updated in a single iteration, it is possible that it is the same individual that has been always updated, while others get stuck for many iterations in the same realization of the latent variable. Not surprisingly, the most volatile part of the chain was the distribution of U_{i0}^{s}, which in the case of the white males exposed in the last section, seemed to have converged only after 30000 iterations, while most of the parameters did not show substantial changes in their distributions since iteration 15000. Solving computational problems is a major step to allow us to divide majors into a larger number of categories, thus better capturing the impacts of the no-switching majors rule.

Simulations and other exercises

In this research, we not only desire to estimate the structural parameters of the theoretical model proposed in Section 1, but also to perform some decompositions (of wages, to see the relative importance of signal information, human capital accumulation and labor market segmentation related to college experiences; and of value functions, to see the magnitudes of the global cost component, the gain associated to the stream of wages in the unskilled sector with and without college, and the option values) and simulations (to evaluate the impact of the no-switching majors rule on social welfare and on the college enrollment and graduation). These exercises can be fully integrated to the MCMC procedure, as it will become clear now.

First, notice that the referred decompositions take an endogenous variable, $y =$

$y(\Upsilon, Z)$, and decompose it additively, as $y = y_1(\Upsilon, Z) + y_2(\Upsilon, Z) + ... + y_K(\Upsilon, Z)$. Since y is a function of random variables, it is a random variable itself, with underlined distribution fully derived from the joint distribution of (Υ, Z), and in the same spirit, the distributions of $y_1, ..., y_K$ are also fully described by the same joint distribution. We know that the sequence of draws of $\Upsilon^{(m)}$ form a sample coming from this joint distribution, and therefore, if we keep computing (and collecting) $y_1^m, ..., y_K^m$, for every iteration $m > M$ (after the warm up phase), we should end up with vector representing the distributions of $y_1, ..., y_K$. Similarly, the policy evaluation is carried out by comparing value functions of the type $V_{ij0}(\Upsilon, Z)$, with its counterparts in the world where the no-switching majors restriction is present, $V_{ij0}^R(\Upsilon, Z)$, and again we can keep computing $\left(V_{ij0}^{(m)}(\Upsilon, Z), V_{ij0}^{R(m)}(\Upsilon, Z)\right)$ until we have a valid representation of the distributions of these quantities (the same is true for the difference $W(\Upsilon, Z) = \max_j \{V_{ij0}(\Upsilon, Z)\} - \max_j \left\{V_{ij0}^R(\Upsilon, Z)\right\}$). Decisions are just the solutions of the individual maximization problems, and can also be represented as $d_{ijt} = d_{ijt}(\Upsilon, Z)$, so that the distributions of the policy effects on enrollment rates and graduation rates can be done likewise.

1.4.2 Results

Preliminary warnings

The results presented in this section use a selected subsample of white males coming from the NLS-72 dataset, with 1097 individuals that satisfied the following requirements: (i) if the agent has ever gone to college, it must be a single spell starting either in 1972 (the year after graduation in high-school), or 1973; (ii) if one chose not to go to college or dropped out of it before graduation, he/ she must also have reported not to have the intention to pursue further education at college level in the remaining waves of the panel; (iii) no missing/ contradictory information about the explanatory variables X_i; (iv) if the person has ever had a college experience, information about grades must be available, as well as the test scores in the battery test. 15% of the schools refused to send transcripts to the NLS-72 committee, and not every student took in, every period in college, classes (with valid grades, which excludes pass/fail) both in subjects related to the major and in subjects unrelated to it; (v) after the first period in the labor market, the individual must have had at most one year with missing information on wages. The reason of this selection is twofold: it reduces the sample size, which helps to make it feasible the time required to run the code, and it makes the sample as similar as possible to the population described in the model. Of course, the cost is that the results are not necessarily valid for the whole population, as this subsample may suffer from additional selectivity issues.

Second, even though the primary information about majors is organized in the NLS-72 data according to the FOS codes in more than a hundred majors, we are not able to deal with this many choices per period, so that we have to group majors into broader categories. Since it is plausible that most of the changes of major occur between fields relatively similar to one another, and which would be in the same broader category, any evaluation of restrictions to major switching would miss the movements that happen within categories, so that the impact of this restriction on welfare and other economic variables is likely to be underestimated.

Third, I managed to have what I thought was the best approximation of a signal that could influence the agent's decision from the transcript files. In reality, though, it may be the case that the most important information is in fact the relative performance of the individual when compared to his peers. Unfortunately, such measures of relative performance are not available in the data, and both the GPA and the signal proposed before revealed to be very little correlated with observed wages (which is the main brick of the value functions, and hence, decisions). Not surprisingly, the signal equations did not show a significant importance of specific abilities in the determination of the signal, and the consequence was that the true importance of learning and the informational structure on the exercises is probably underestimated. Moreover, because the laws of motion impose the information set is increased only by the signal and the cost shocks at every period, the cost shocks acquire a disproportional importance in determining the agents' choices, as the signal used in the exercise is not very informative. The use of datasets with detailed information about the individual performance in college would be an important advance regarding this research.

Fourth, the model contains a very rigid structure that may not conform to the true relations between the dependent and explanatory variables in reality. In particular, risk neutrality simplifies a lot the calculations, by allowing us not to model the credit markets explicitly. Even accepting risk-neutrality, nothing in reality or in economic theory imposes that wages, costs and signals must be linearly related to the covariates. The choice of a subsample of white males to construct the main exercise presented in this section helps to minimize this problem because it homogenizes the sample, but again, some caution should be taken when looking at its results.

In spite of all these limitations, the estimation and simulation will show two things: first, that the method works and could be used for classes of problems similar to this. Second, these weaknesses point out where additional effort should be spent in order to improve on this research.

The learning process and the informational structure

The standard MCMC graphs with the draws made from the ergodic distribution of the chain can be seen in the appendix E.

In the model, it is assumed the key ingredient of the learning process is the arrival of new signals, G_{ijt}, associated to enrollment in college in major j, time t. Because this subsample is homogeneous in the observable characteristics, the only element of β_j^g is the intercept, which together with $(\xi_j, \sigma_j, \theta)$ fully characterize the dependence of G_{ijt} on the state variables $(X_i^g, U_{it}, \varepsilon_{it})$

I will leave the analysis of the human capital accumulation rate, θ_j, to the next subsection. The average obtained from the ergodic distribution of the chain, and the respective 95% interval containing the most frequent realizations for the remaining parameters were:

Table 1.6: Average parameters of the signal equations

	β^g	$\xi\ (x10^{-4})$	σ
Sciences	-0.079	-0.03	3.00
	$(-0.19 : 0.04)$	$(-0.49 : 0.40)$	$(2.88 : 3.14)$
Non − sciences	-0.055	0.02	3.18
	$(-0.18 : 0.08)$	$(-0.20 : 0.28)$	$(2.99 : 3.37)$

In this table, and in the Graphs A4-A6, we see that both the intercept and the combination of parameters $\xi_j = \frac{\theta_{jj}\varphi_j}{\sigma_j}$ have symmetric distributions with averages close to zero and relatively large dispersion, suggesting that the signal is not influenced by abilities. The variance of this equation, on the other hand, is accurately measured and with similar magnitude in both majors. The ratio $\frac{\varphi_j}{\sigma_j}$ is therefore small, implying that $\Lambda_{it+1} \approx \theta_j\ \Lambda_{it}\theta_j$, and if the human capital accumulation rate is positive ($\theta_j >> 1$), we should observe the variance of the unknown part of abilities increasing over time (and at approximately the same rate of the variance of the known component).

Turning now to the initial conditions of the decision problem, the averages of the realizations of Λ_0 and Ω_0 were:

$$\overline{\Lambda}_0 = \begin{bmatrix} 1.06 & \underset{(-2.25\,:\,2.99)}{0.13} & \underset{(-1.60\,:\,1.55)}{-0.14} \\ 0.13 & \underset{(20.47\,:\,75.32)}{27.46} & \underset{(-20.83\,:\,21.95)}{-0.55} \\ -0.14 & -0.55 & \underset{(20.52\,:\,94.78)}{25.93} \end{bmatrix} 9$$

$$\overline{\Omega}_0 = 10^{-4} * \begin{bmatrix} \underset{(4.56\,:\,5.66)}{5.19} & \underset{(-0.24\,:\,0.17)}{0.004} & \underset{(-0.28\,:\,0.20)}{-0.0002} \\ 0.004 & \underset{(1.37\,:\,1.65)}{1.50} & \underset{(-0.19\,:\,0.10)}{-0.0002} \\ -0.0002 & -0.0002 & \underset{(1.94\,:\,2.33)}{2.12} \end{bmatrix}$$

It is clear from these matrices that the component of true abilities unknown to the agents at the end of high school is the dominant term of the total variance of true abilities, U_{i0}. The difference lies in the order of 10^4, which suggests either that the agents know very little about their abilities at that point in time, or that the structure of the model does not capture what agents really know at the first period of their decision problems. In the Graphs A7-A10, we see the distributions of the elements of these matrices. The first two pictures contain the distribution of the diagonal elements of these objects, except for the unidentified element λ_{00}. It is interesting to notice that while the average constitutes a good approximation of a typical realization of ω_{jj}, the posterior distributions of λ_{jj} are very skewed to the right, with fat tails and the mode floating around 1/4 of the average (which is still much higher than the respective values of ω_{jj}). Furthermore, while the distributions of $\lambda_{11}, \lambda_{22}$ almost coincide, the variances of the known part of initial abilities display very different symmetrical distributions, being ω_{11} the one with lowest mean and dispersion, and ω_{00} the one with the highest. Turning now to the off-diagonal elements of these matrices, we see that in both cases the magnitude is much lower than their diagonal counterparts. This result is important, because it suggests that a diagonal matrix would approximate reasonably well the shape of them, simplifying the learning process to the one assumed in the (uncorrelated) multiarmed-bandit model, for which a much simpler estimation method is available.

The dynamic bias and the distribution of U_{ij0}^s

Together with the matrices Λ_0 and Ω_0, the initial state I_{i0} also contains the unobserved (from the point of view of the econometrician) signals about specific abilities, U_{i0}^s that move over time. At each decision node, agents tend to self-select themselves into majors that could eventually allow them to supply labor to the segment of the market where they believe to have comparative advantage. Therefore, while the ex-ante variable U_{i0}^s was normally distributed across agents with mean zero, the ex-post distribution of this variable among the subsample of college graduates in a specific major may have a very different distribution. On each iteration of the chain, I computed the average of this distribution to see how important dynamic bias is, and the results are shown in the Graph A11

These graphs are very informative about the importance of self-selection in the model. While the distribution of U_{i10}^s, U_{i20}^s (signal related to scientific and non-scientific abilities, respectively) has mean zero in the whole population, its average among college graduates in sciences and non-sciences is approximately US\$420.00, and US\$ 2,400.00, respectively. To have an idea of the importance of this effect, notice that the first yearly income among science graduates is around US$ 8,645, and among non-sciences, US$ 9,528. This means that 5% of the sciences income comes from individual advantages in this field, and impressive 25% of the starting earnings of non-science graduates come from comparative advantages in this area.

Labor earnings

The wage equations proposed in this article are linear, and with specific coefficients associated to each different segment of the labor market. The table below shows the average realization of the parameters that characterize this equation from the ergodic MCMC (units are US$ 10,000 of yearly earnings), which is complemented by the Graphs A12-A14:

Table 1.7: Average parameters of the wage equations

Career	β^w	A	ρ
Unskilled	0.64 (0.63 : 0.66)	0.14 (0.138 : 0.146)	0.33 (0.3212 : 0.3352)
Sciences	0.81 (0.79 : 0.83)	0.22 (0.215 : 0.222)	0.07 (0.066 : 0.075)
Non − sciences	0.73 (0.68 : 0.78)	0.21 (0.207 : 0.223)	0.18 (0.1669 : 0.1914)

First, notice that, as expected, the wages in occupations that do not require a college degree display lower average level (intercept), and slower evolution over time (A). If this were not true, it would be necessary a much greater comparative advantage in one of the specialized fields to justify delaying the beginning of the professional career in order to get a college diploma, and on the top of that, paying costs associated to higher education. Furthermore, these parameters are slightly higher in sciences than in non-sciences occupations, which suggests that the main attractive of this career is that observable attributes are better rewarded there, in contrast to the important role played by the perceived specific ability in non-science positions, as mentioned in the last part. Regarding to the last parameter, ρ, which measures the variance of the (non-persistent) shocks unrelated to personal traits and that affect wages in each period, we observe that the dispersion is much lower in sciences than in other careers. If we compare the remaining dispersion after controlling for X_i, U_{it}^s with the total dispersion of wages in each segment of the market (3.40, 0.18 and 1.10, for unskilled, science and non-science, respectively), we see the same ordering appears in the raw data, and also that the unexplained variation in the model lies between 9.7% and 38.9% of the total (unconditional) variance observed in wages.

The decomposition of the different effects of college education on wages confirms our suspicion that credential effects are relatively more important to explain the college premium in scientific occupations than in non-scientific occupations. As the Graphs A15-A16 indicate, 87% of the labor income increase in the scientific segment is explained by the fact that personal characteristics are rewarded in a particularly favorable way, while this effect accounts for no more than 37% in the non-sciences occupations.

Indeed, what really matters to non-science workers is the human capital accumulated in college, which composes 63% of the gain. This explanation becomes clear once we analyze how classes in different majors impact specific abilities. The next table and Graphs A17,A18 and B1 show the average realizations of θ_{jk}, i.e., the impact of experience in major j on abilities of type k (to recover θ_{jk} from this table, one has to divide these numbers by 100 and add 1) :

In fact, it is clear that one year of classes in non-sciences generates an impact on this type of human capital (83%) much bigger than one year of sciences on scientific abilities (almost a half of it), which justifies the tremendous importance

Table 1.8: Average parameters of the human capital accumulation equations

major	ΔU_{i0} (%)	ΔU_{i1} (%)	ΔU_{i2} (%)
Sciences	211.89	41.65	34.89
	(200.28 : 223.30)	(39.68 : 44.25)	(22.87 : 42.33)
Non − sciences	97.80	32.21	83.06
	(84.54 : 113.67)	(8.51 : 55.80)	(80.7 : 85.6)

of this effect on wages compared to sciences. The numbers also show a very important impact of college classes on abilities used in unskilled jobs, which should cause unskilled workers with some college experience to earn much higher wages than those who went to the labor market right after graduation in high-school. Alternatively, it possibly suggests that using college diploma acquisition to divide people across labor market segments may be misguiding, since in fact people with some college may be competing for specialized occupations with college graduates, but unfortunately the data do not provide details about this.

Costs

In our estimations, the cost equations were the only ones that included covariates, namely, the maximum education achieved by one of the parents of the agent. On the other hand, the variance of the random term of costs, S, is one of the parameters fixed exogenously, which means that the scale of these parameters not necessarily correspond to what would appear if we included S in the estimation. The average values of these parameters is shown below:

Table 1.9: Average parameters of the cost equations

	Intercept	*Parent's education*
Sciences	3.45	−0.50
	(3.15 : 3.75)	(−0.59 : −0.40)
Non − sciences	2.96	−0.27
	(2.60 : 3.32)	(−0.37 : -0.15)

Assuming $S = 10$ is the correct value for this parameter, these numbers say that total cost of going to college for children of parents with less than high-school is US$ 30,000.00. After that, each addition to parental education reduces costs by US$ 2,700.00 if the agent takes classes in non-scientific majors, and twice as much if he/ she chooses Sciences. This means that more educated parents generate a different type of comparative advantage in Sciences to their children, not related to specific talents, since costs will reduce up to US$ 10,000 among children of people who went to graduate school. In the context of this exercise this is not irrelevant, especially when we take into account that the learning mechanism is not very significant and choices are likely to be very influenced by the composition of costs. The distributions of these parameters are presented in the Graphs B2-B3.

In fact, if we consider that the option value of going to college is, on average, around US$ 122,000 in both science and non-science majors, and this cost is basically the price to keep it open the possibility of eventually graduating in college, we have that this price may be 25% of the expected payoff for children of parents with less than high-school, and only 10% if the individual has at least one parent with post-college education and chose sciences in the first period. The full distribution of the option values (in period 0) of going to college is displayed in the Graph B4 , for both sciences and non-sciences, and it shows a similar format for the two distributions, with sciences in slight advantage.

Policy effects

Turning now to the counterfactual simulations about college decisions under the no-switching majors rule, we elected four main indicators to analyze. The first two intend to capture the welfare loss associated to the extra constraint imposed on the individual's maximization problem, both for the average citizen and for the potential students for who this restriction is binding (which may be seen as an analogue of the Treatment on the Treated effect found in the literature about experiments). These measures involve variations in the expected present value of income evaluated at the optimal choice of the agents. The last two focus on the variations in the optimal choice itself, and aim to compute variations in the proportion of high-school graduates that choose to go to college, and the proportion of college students who effectively graduate in some major.

The table below shows the average impact of the restriction on the welfare of the members of our sample.

Table 1.10: Average policy effects

	Average	5%(−)	5%(+)
Impact on the total population (US$)	5,149.80	4,897.41	5,414.68
Impact on the treated (US$)	9,337.73	8,880.09	9,818.02

and the Graph B5 depicts the distribution of the "Treatment on the Treated":

The average loss on individuals who seriously consider to go to college after high school are of the same magnitude as the average initial earnings of a college graduate, which can be considered a significant impact (even though it is about only 5% of the total value of going to college). In other words, if these numbers could be generalized to other situations, it would suggest that a country where the no-switching majors rule is currently adopted could improve the welfare of talented people in an amount equivalent to one year of labor income, just by removing this constraint, which in principle does not involve changes in the governmental educational budget. Of course, this exercise is just a first approximation to these impacts, but it may be just the lowerbound of the true impacts, if we consider that most changes of major occur within the broad categories of sciences and non-sciences, and are not captured by this simulation. Even if one does not believe in these numbers, it provides a strong argument in favor of a serious research around this topic.

The results look even more impressive when we investigate the changes in the proportion of high-school graduates who decide to go to college. As the Graph B6 emphasizes, the number of college students decays from 492 (44.85%) to an average of 388.8 (35.44%), after the prohibition of switching majors. It is true that some caution should be used here, since we already mentioned that individual decisions may be disproportionately affected by circumstantial (cost) shocks, since the signals obtained from the data were not very informative. Good signals would reinforce the importance of the persistent unobserved abilities in the model, which could in principle make the solutions to the individual problems less susceptible to changes in the environment.

Decisions are also reflected in the graduation rates, which likewise seem to be very affected by the restriction to major switching. In the sample, 396 students got a degree after four years in college (36.1%), while, on average, only 165.9 people finished college in the simulations (15.12%).

Enlarged sample

The subsample of white males was a workhorse for this exercise. On one hand, the pieces of the model had to be made as simple as possible in order to allow the realization of the empirical part, but nothing in economic theory suggests that wages, costs and signals have to be linear in the covariates. A set of homogeneous observations on this dimension helps to minimize the dependence of the conclusions on the arbitrary structure of the model, but raises a natural question about the generality of the respective results, and this subsection aims to fill in this gap by presenting some results obtained from a sample that also includes non-whites and women.

Costs and signal The numbers shown on Table 1.11 indicate that the signal is once again not explained by neither the observed differences across agents, nor by the sampled unobservables, and the only significant difference from the previous exercise was an increase in the estimated variance of the equation, σ_j.

Regarding to costs, we see that the average level for white males is similar to the numbers found in the previous section (as captured by the sum of the intercept and the dummy variable of sex), but classes in sciences are much costlier to women and non-white, who have to pay, on average, 20,7% and 14% (respectively) more than white males for one period of classes in this field. This is in sharp contrast with classes in non-sciences, which do not display any significant cost differences between genders or races. The cost equations also show that the importance of parents' education increases in sciences and decreases in non-sciences when we include women and non-white in the estimation, which may reinforce the effect of differences in costs across majors on the racial composition of the scientific labor force, as the average education among non-whites' parents is significantly lower than whites' (especially considering this data comes from 1972 high-school seniors).

Wages and human capital accumulation Table 1.12 compares the estimations of the wage equations in the two subsamples mentioned above. The first thing that immediately calls our attention is the remarkable decrease in the premia associated to experience in the labor market in all majors, when women and non-white are included in the estimation. In particular, the differentials between unskilled and specialized careers almost disappear, suggesting that college education no longer leads to a faster income growth over the lifetime.

Analyzing the race and gender differentials, the first interesting fact is that non-whites appear to have higher wages in scientific careers than whites, which together with the estimations of the cost equations indicate that the main incentive whites have to pursue a degree in this field is that it is less costly to them to do so, whereas the incentive for non-whites is the higher payoffs associated to this choice. Still about racial differences, we see that this variable is not significant to explain wage differences in non-sciences, and that whites have a great advantage over non-whites in unskilled occupations.

In terms of sex, women get occupations that on average pay lower salaries in all fields. The magnitudes are similar in the two specialized careers, and equivalent to approximately 25% of the initial earnings in these jobs. However, the gender differences are lower in college occupations than in high-school jobs.

When we verify how the rate of human capital accumulation changes after enlarging the sample, the main fact is that the increase in college abilities becomes bigger, while the increase in the high-school ability becomes smaller. In the unskilled sector, this means that having some college experience does not pay off as much for the average individual as it does for the average white male, which would potentially represent a greater incentive to complete college among non-whites and women.

Other parameters and policy effects Table 1.13 shows that no significant changes appear in the variance matrices of the distributions of U_{i0}^s and v_{i0}. In both cases, these objects appear to be reasonably approximated by diagonal matrices, and the variances of the known components of ability are on average slightly smaller in the enlarged sample, in contrast with the variance of the unknown part, which increases even more in the more heterogeneous set of observations.

Finally, Table 1.14 contains the results of the simulations of the effect of the no-switching majors rule on the enlarged sample. Interestingly, this table induces to different conclusions about the consequences of changing the sample composition, depending on the indicator analyzed. On one hand, the welfare loss associated to this rule seems to be smaller in the enlarged sample than in the group of white males, even though the magnitudes of the treatment effects in both cases is similar. However, if we focus on the enrollment and dropout rates, we see that the no-switching restriction has a significantly higher impact among the enlarged sample than in the previous exercise. These two facts suggest that non-whites and women are closer to the indifference thresholds that determine the choices of going to college and major than the white males, and are therefore more sensitive to perturbations in the environment. If true, this may suggest that policies aiming to increase the number of non-whites and women in college would be more effective if the environment was more flexible than otherwise.

1.5 What have we learnt from this research? - final remarks

This research has two types of contributions. First, it tries to quantify the relative importance of three different channels through which college education may affect labor earnings, and to evaluate the impact of a policy commonly found in many countries of the world, namely, the prohibition that college students switch majors during their courses. Second, it develops a method to answer these questions, which involves estimating a structural dynamic model of choice of major where individuals learn about their abilities during college, and self-select themselves into careers where they think they could have comparative advantage based on the available information. I divide my conclusions into two aspects: what I learned about how to approach this problem, and what I learned about the answers to the questions that motivated this research.

In terms of the questions proposed at the beginning of this article, we found that both human capital accumulation and labor market segmentation are important to explain why college graduates earn more than high-school graduates, but the relative importance of these channels may be very different for different types of occupations. In our exercise, occupations that require a diploma in a scientific major reward particularly well the individual productive attributes, being this the main incentive to choose these careers. On the other hand, majors associated to non-scientific specialized careers have a high content of human capital accumulation, and this is what especially attracts people to take classes in this area.

Regarding the policy evaluation, we found that forbidding people to switch majors may cause a welfare loss equivalent to one year of labor earnings in a specialized job. If we take into account that this constraint could in principle be removed without major costs to the educational policymakers, it seems unjustifiable that some countries keep this policy. Of course, richer models could provide more accurate answers, and the supporters of this restriction could argue there are other costs to change the system, but the magnitude of this impact certainly asks for more effort to be spent on this question.

The limitations of the empirical strategy were also useful to point out where we should focus our energy in order to improve our results. First, the variable constructed to capture the new information that arrives during college revealed to be non-informative about one's specific talents, which may be an indication that this type of information is not decisive to explain the agent's behavior, or (which I believe) that individual grades are not suited to be the empirical counterpart of the signal modeled in Section 2. Indeed, the attempt of using individual grades (or transformations of it) as the signal suffered from problems of heterogeneous grading policies across educational institutions and from the lack of measures of relative performance, such as the rank in the class or the distance between the grade obtained and the average of the student's peers. The ideal dataset should also include specific information about honors in specific classes, since the same CIP code may be used to denote classes with very different degrees of difficulty, attenuating the relation between individual specific talents and observed grades. Moreover, it is still necessary a deeper discussion concerning whether the learning structure adopted in the theoretical model is realistic or not. On one hand, if more than one signal arrives per period, the analyst should be able to accommodate these variables in the model, in order to better describe the individual's decision behavior. On the other hand, the whole dynamics of the model is fully determined by a very particular combination of the distributions of the unobservables, which should be relaxed.

The implementation of the empirical procedure can also be improved, which requires the experimentation of new numerical methods to solve the excessive computational burden associated to the multidimensional integration that appears in the solution of the individual maximization problems, and the experimentation of other sampling techniques to speed up the convergence of the MCMC.

Overall, I can say the method was successful in providing an integrated procedure that combines estimation with simulation, and allows us to evaluate a class of policies where the changes occur in a discrete set of choices and where the whole population is affected. However, at this stage the answers obtained should still be seen with reserve, due both to the rigid structure of the underlying model and the limitations of the data.

	White Males		Enlarged sample	
	Average	90% interval	Average	90% interval
Signal equation				
Sciences				
Intercept	-0.08	-0.19 : 0.04	-0.14	-0.42 : 0.13
Sex	-	-	0.03	-0.15 : 0.21
Race	-	-	0.04	-0.25 : 0.31
Ability (x10^4)	-0.03	-0.49 : 0.40	0.00	-0.11 : 0.10
Variance	3.00	2.88 : 3.14	3.17	3.05 : 3.29
Non-sciences				
Intercept	-0.06	-0.18 : 0.08	-0.32	-0.60 : -0.04
Sex	-	-	0.02	-0.17 : 0.20
Race	-	-	0.29	-0.00 : 0.58
Ability (x10^4)	0.02	-0.20 : 0.28	-0.01	-0.11 : 0.07
Variance	3.18	2.99 : 3.37	3.42	3.30 : 3.54
Cost equation				
Sciences				
Intercept	3.45	3.15 : 3.75	3.96	3.59 : 4.28
Sex	-	-	-0.68	-0.87 : -0.48
Race	-	-	-0.46	-0.70 : -0.21
Parent's Education	-0.50	-0.59 : -0.40	-0.62	-0.69 : -0.55
Non-sciences				
Intercept	2.96	2.60 : 3.32	2.29	1.97 : 2.61
Sex	-	-	0.05	-0.16 : 0.26
Race	-	-	-0.23	-0.55 : 0.09
Parent's Education	-0.27	-0.37 : -0.15	-0.24	-0.31 : -0.18

Table 1.11: Comparison between the parameters obtained from the sample of white males and from the enlarged sample: signal and cost equations

Table 1.12: Comparison between the parameters obtained from the sample of white males and from the enlarged sample: wage and human capital accumulation equations

	White Males		Enlarged sample	
	Average	90% interval	Average	90% interval
Wage equation				
Sciences				
Intercept	0.81	0.79 : 0.83	0.911	0.879 : 0.951
Sex	-	-	0.276	0.253 : 0.295
Race	-	-	-0.069	-0.111 : -0.034
Experience	0.22	0.215 : 0.222	0.044	0.042 : 0.045
Variance	0.07	0.066 : 0.075	0.071	0.069 : 0.073
Non-sciences				
Intercept	0.73	0.68 : 0.78	0.863	0.818 : 0.905
Sex	-	-	0.270	0.239 : 0.299
Race	-	-	-0.037	-0.084 : 0.013
Experience	0.21	0.207 : 0.223	0.048	0.046 : 0.051
Variance	0.18	0.167 : 0.191	0.121	0.117 : 0.126
High-school				
Intercept	0.64	0.63 : 0.66	0.556	0.536 : 0.576
Sex	-	-	0.367	0.352 : 0.382
Race	-	-	0.052	0.033 : 0.072
Experience	0.14	0.138 : 0.146	0.037	0.035 : 0.038
Variance	0.33	0.321 : 0.335	0.436	0.430 : 0.442
Law of motion (theta)				
Sciences				
HK to sciences	1.42	1.40 : 1.44	1.540	1.525 : 1.555
HK to non-sciences	1.35	1.23 : 1.42	1.527	1.423 : 1.625
HK to high-school	3.21	3.20 : 3.22	2.822	2.686 : 2.960
Non-sciences				
HK to sciences	1.32	1.09 : 1.56	1.483	1.429 : 1.536
HK to non-sciences	1.83	1.81 : 1.86	1.834	1.821 : 1.847
HK to high-school	1.98	1.84 : 2.14	1.759	1.480 : 2.064

Table 1.13: Comparison between the parameters obtained from the sample of white males and from the enlarged sample: Lambda and Omega

	White Males		Enlarged sample	
	Average	90% interval	Average	90% interval
Lambda				
(1,2)	0.130	-2.25 : 2.99	0.033	-1.81 : 2.16
(1,3)	-0.140	-1.60 : 1.55	0.073	-0.27 : 1.59
(2,2)	27.460	20.47 : 75.32	53.859	16.85 : 177.27
(2,3)	-0.550	-20.83 : 21.95	-4.573	-52.52 : 28.03
(3,3)	25.930	20.52 : 94.78	35.996	12.23 : 93.44
Omega (x10^4)				
(1,1)	5.190	4.56 : 5.66	3.941	3.75 : 4.14
(1,2)	0.004	-0.24 : 0.17	0.002	-0.06 : 0.06
(1,3)	0.000	-0.28 : 0.20	-0.001	-0.07 : 0.07
(2,2)	1.500	1.37 : 1.65	0.901	0.85 : 0.95
(2,3)	0.000	-0.19 : 0.10	0.000	-0.03 : 0.03
(3,3)	2.120	1.94 : 2.33	0.959	0.91 : 1.01

Table 1.14: Comparison between the policy effects obtained from the sample of white males and from the enlarged sample

	White	Males	Enlarged	Sample
	Average	90% interval	Average	90% interval
Treated on the treated	0.93	0.89 : 0.98	0.82	0.79 : 0.84
Average treatment	0.51	0.49 : 0.54	0.47	0.46 : 0.49
Δ Enrollment rate (p.p.)	-9.40	$-11.21 : -7.57$	-14.37	$-15.51 : -13.21$
Δ Graduation rate (p.p.)	-15.78	$-11.21 : -7.57$	-18.45	$-19.52 : -17.40$

APPENDIX A

MAIN GRAPHS

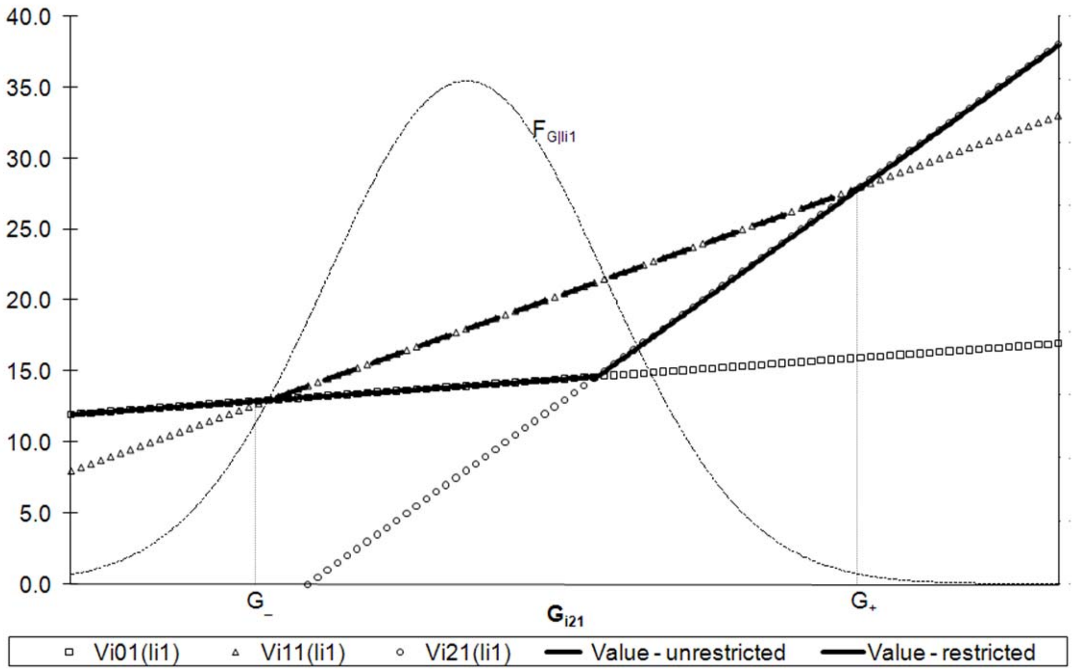

Figure A.1: Mechanics of the model - the role of G

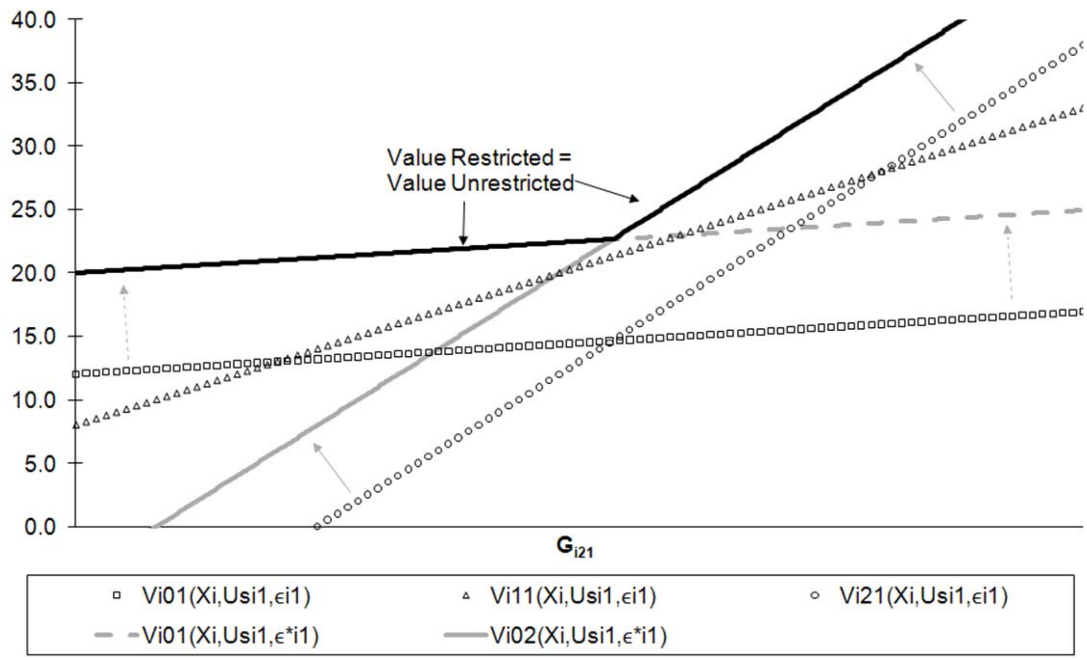

Figure A.2: Mechanics of the model - the role of other elements of the information set

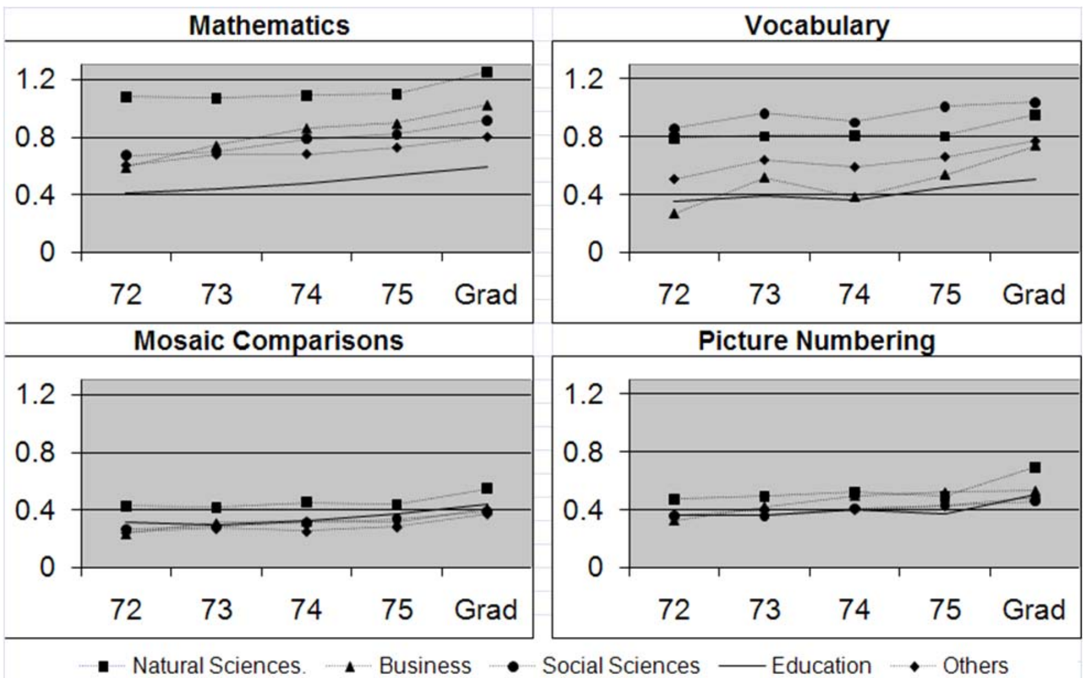

Figure A.3: Average standardized test scores, by field

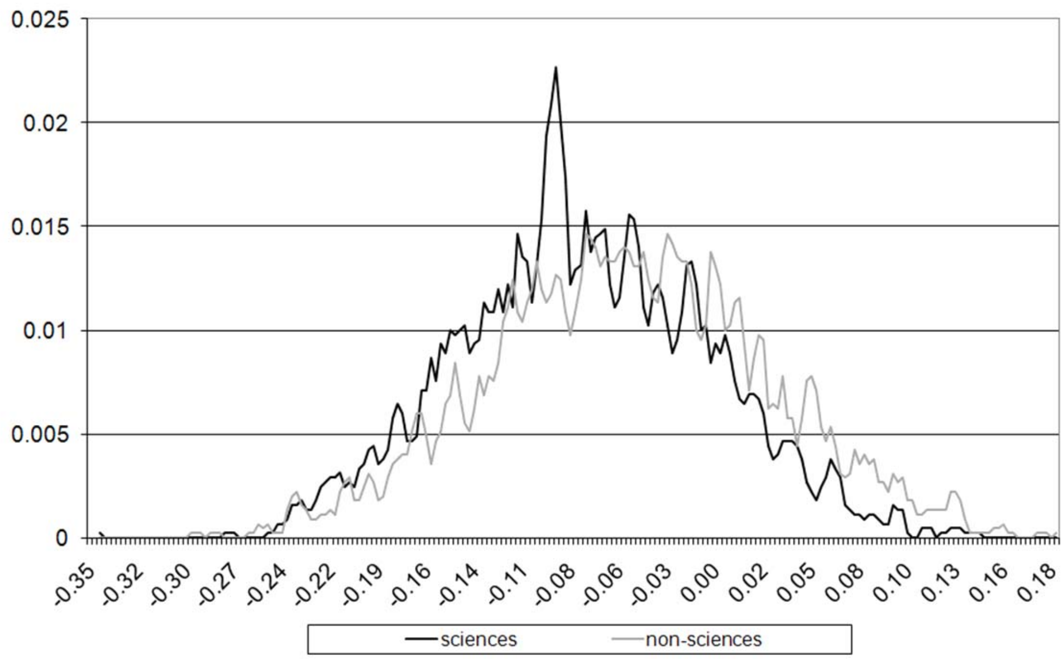

Figure A.4: MCMC distribution of the intercept of the signal equations

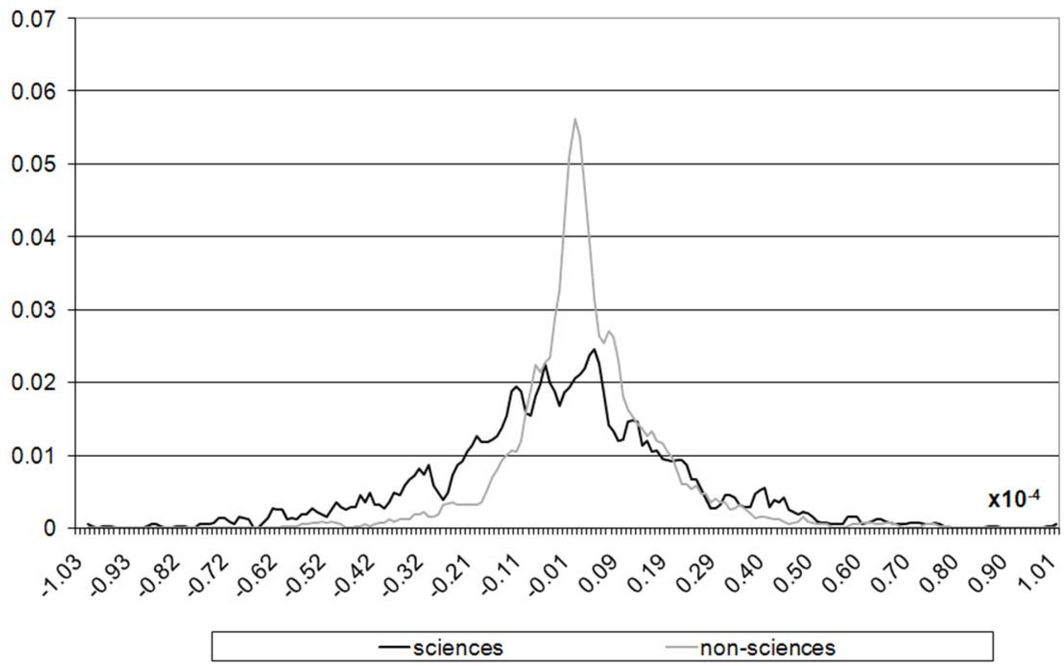

Figure A.5: MCMC distribution of the ξ's in the signal equations

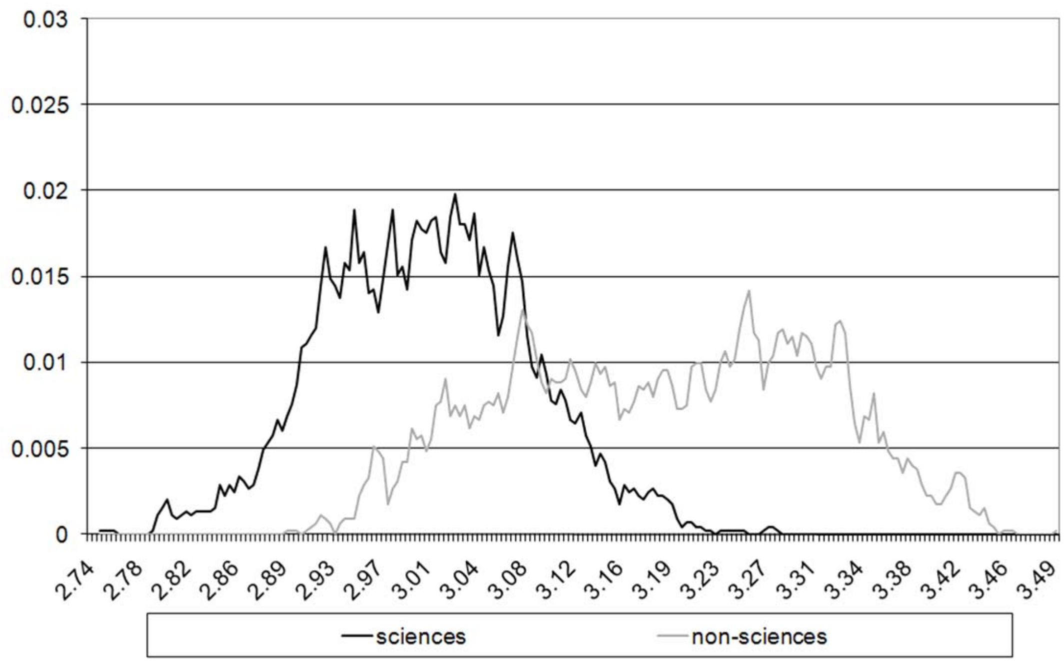

Figure A.6: MCMC distribution of the σ's in the signal equations

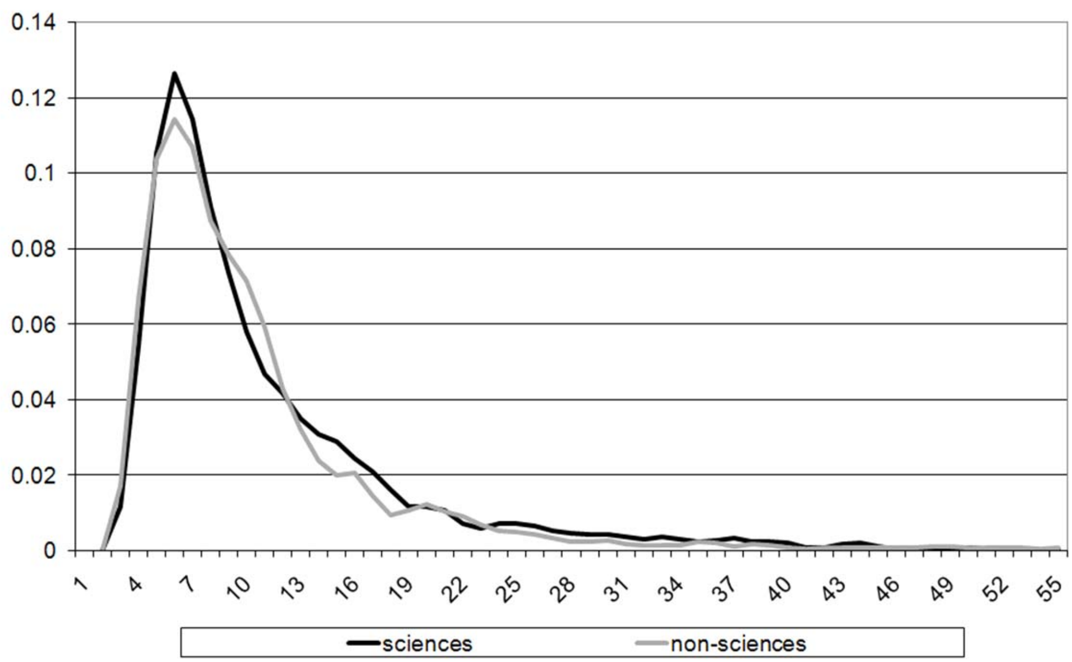

Figure A.7: MCMC distribution of the diagonal elements of Λ_0

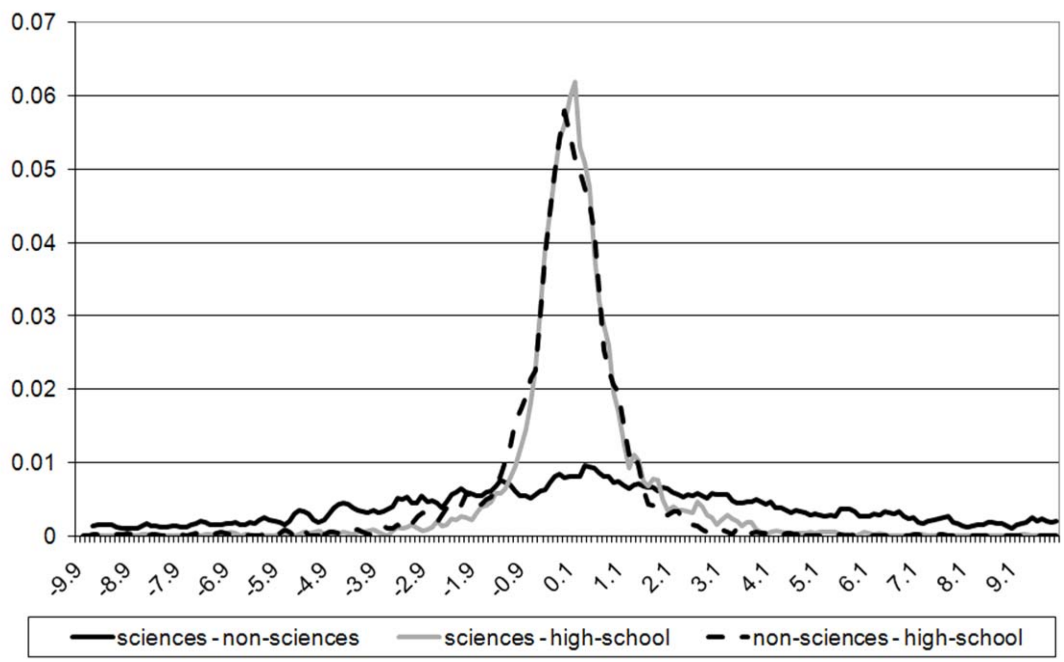

Figure A.8: MCMC distribution of the off-diagonal elements of Λ_0

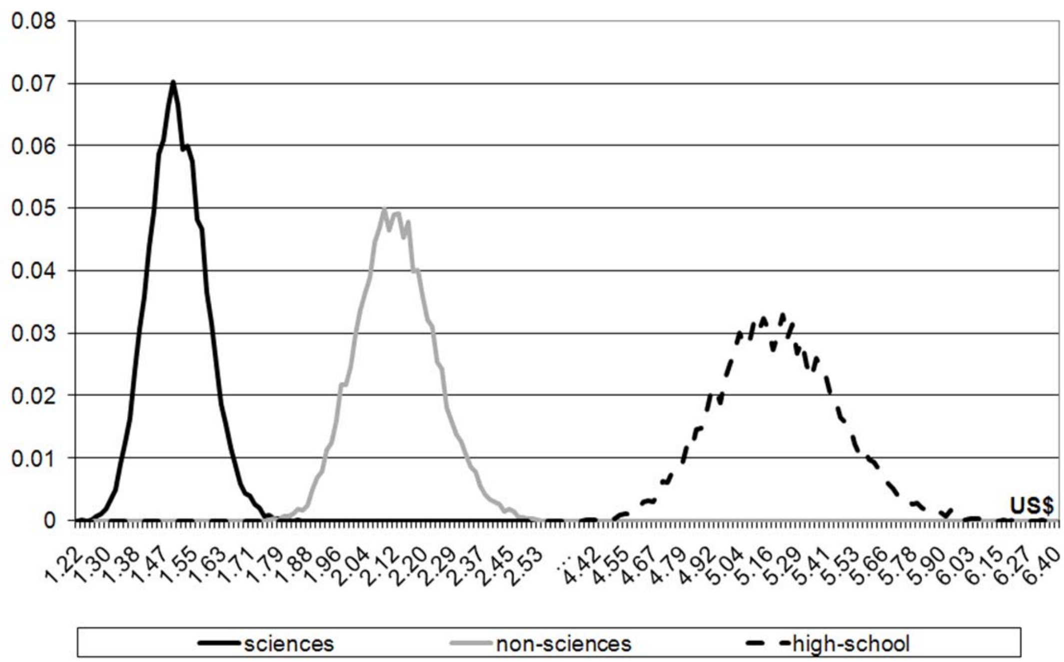

Figure A.9: MCMC distribution of the diagonal elements of Ω_0

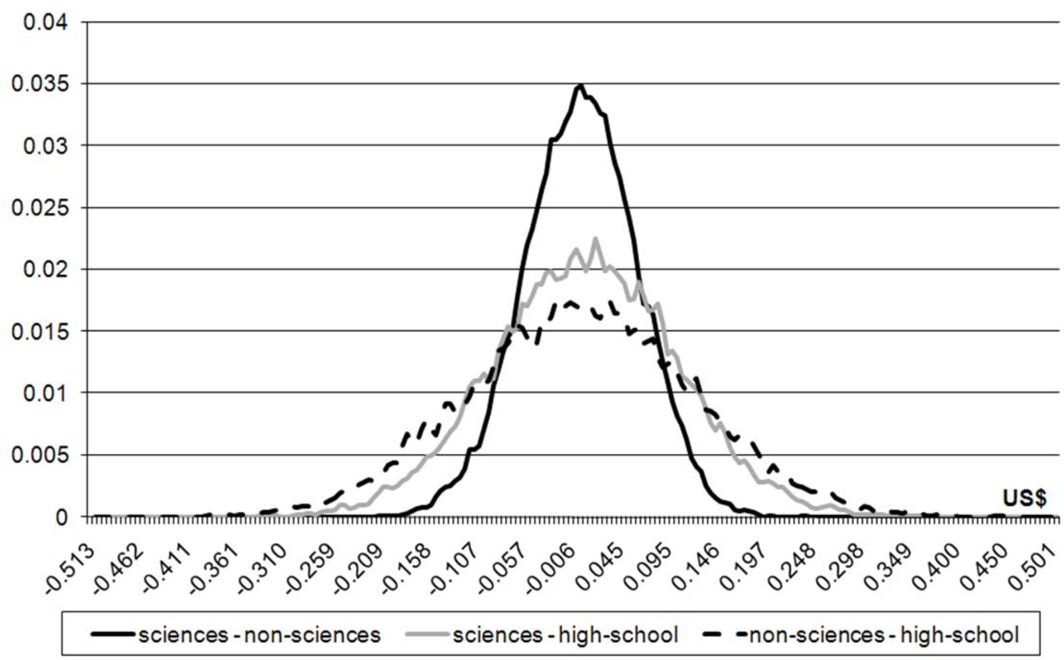

Figure A.10: MCMC distribution of the off-diagonal elements of Ω_0

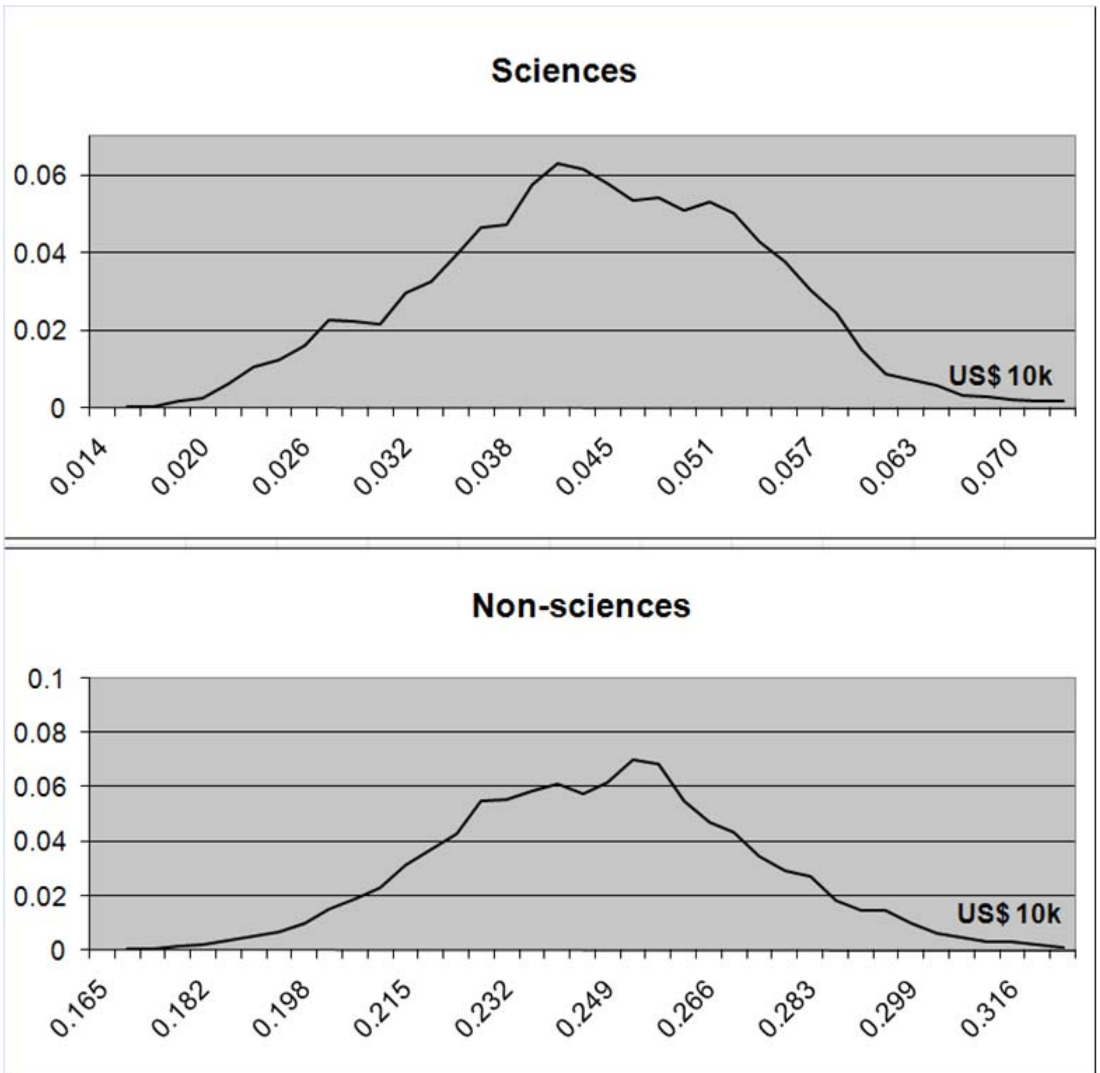

Figure A.11: MCMC distribution of the average (perceived) specific ability among college graduates

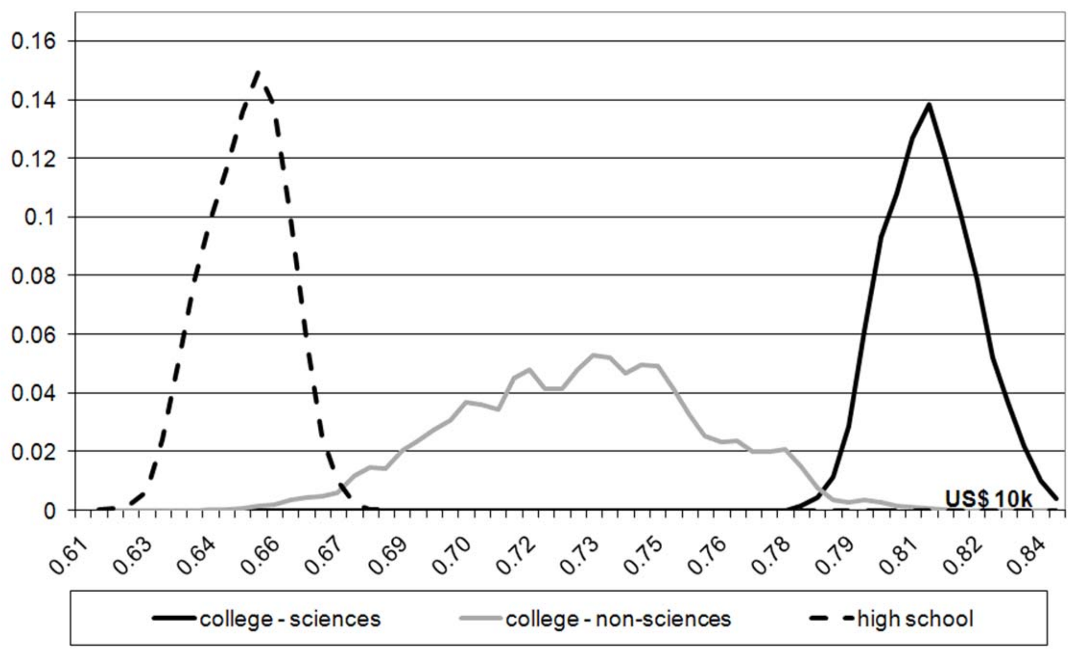

Figure A.12: MCMC distribution of the intercept of the wage equations

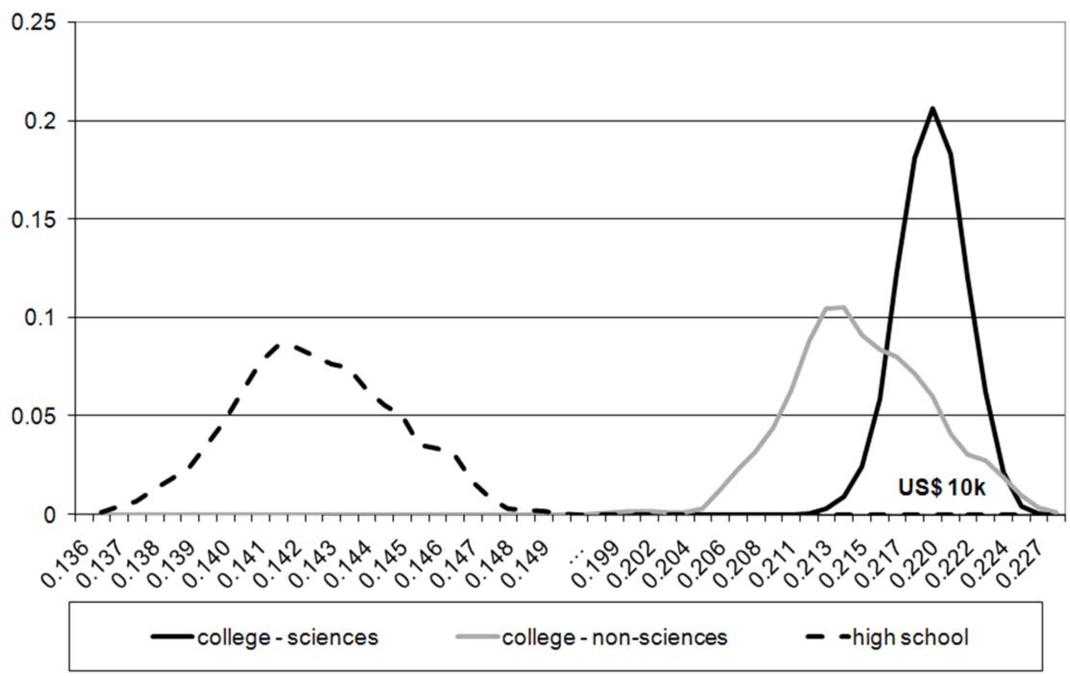

Figure A.13: MCMC distribution of the coefficient on labor market experience, in the wage equations

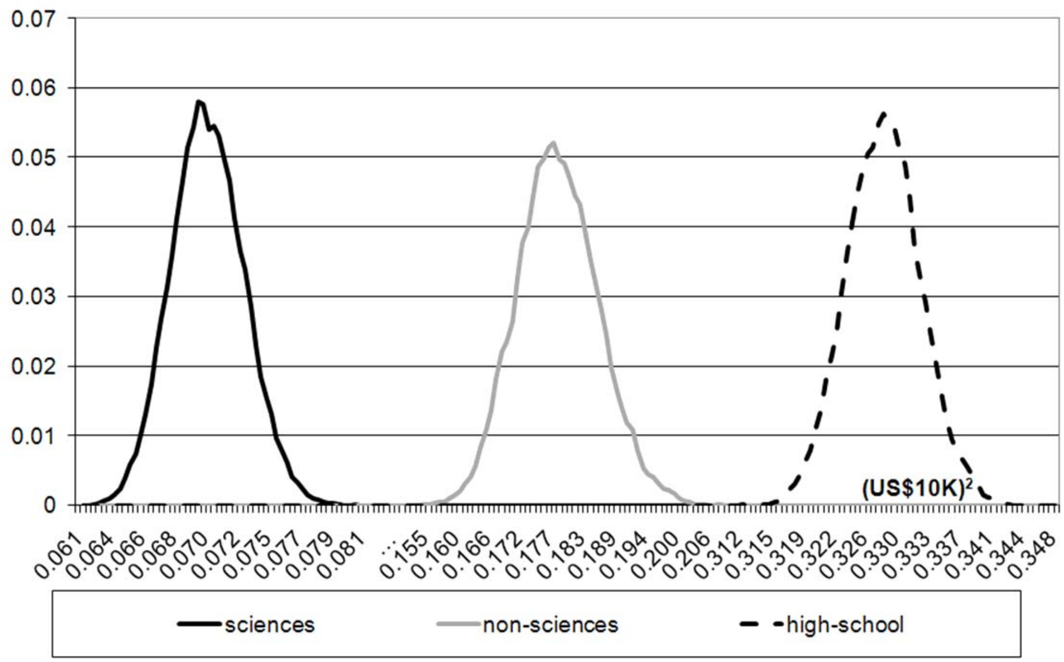

Figure A.14: MCMC distribution of the variance (ρ), in the wage equations

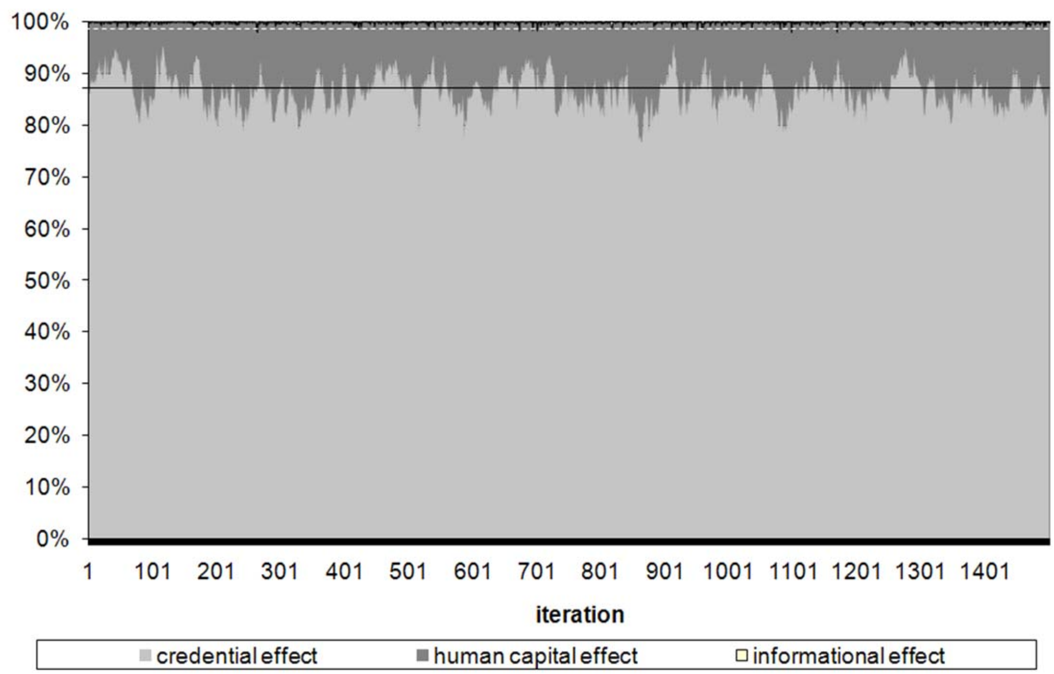

Figure A.15: Wage decomposition among sciences graduates

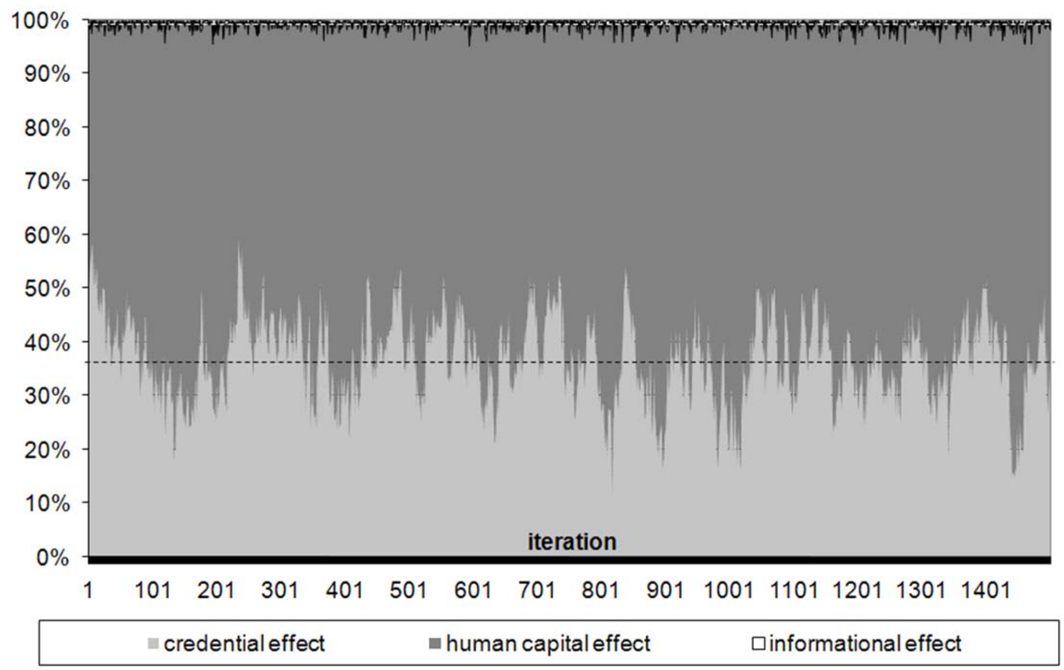

Figure A.16: Wage decomposition among non-sciences graduates

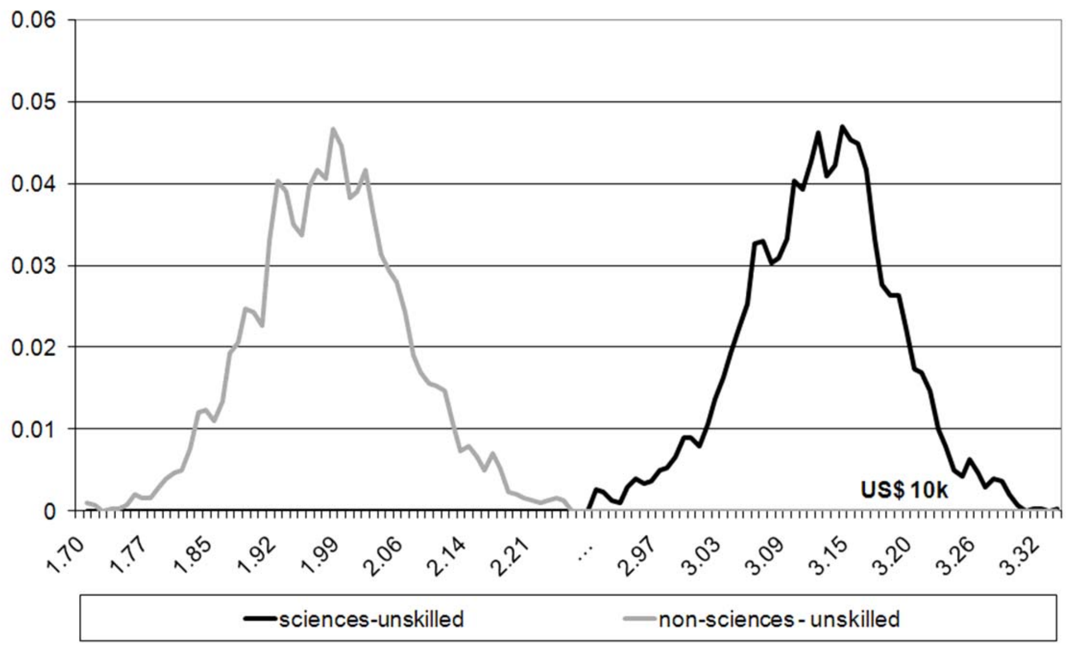

Figure A.17: MCMC distribution of the rate of accumulation of the unskilled talent, θ_{j0}

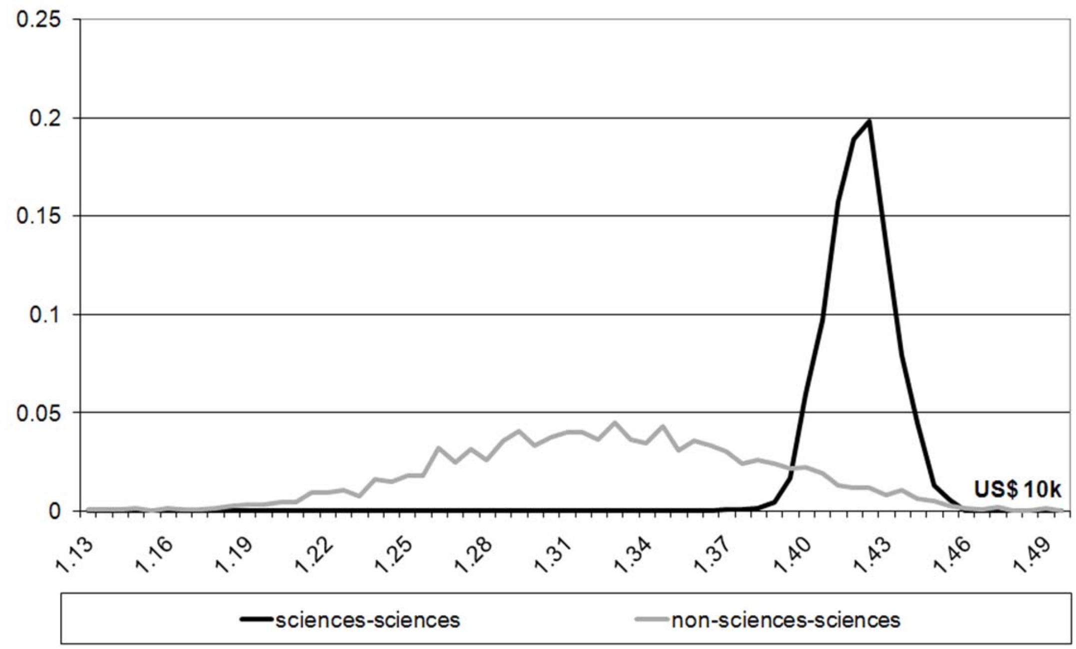

Figure A.18: MCMC distribution of the rate of accumulation of the scientific talent, $\theta_{j,sci}$

APPENDIX B

MAIN GRAPHS (CONT.)

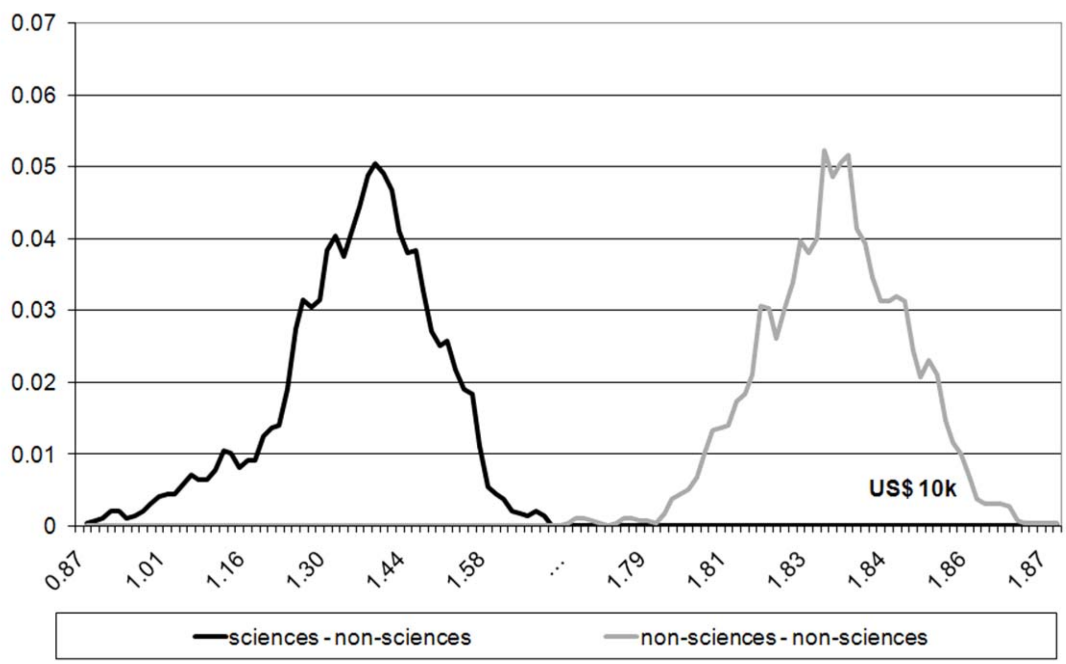

Figure B.1: MCMC accumulation rate of the non-scientific talent, $\theta_{j,n-sc}$

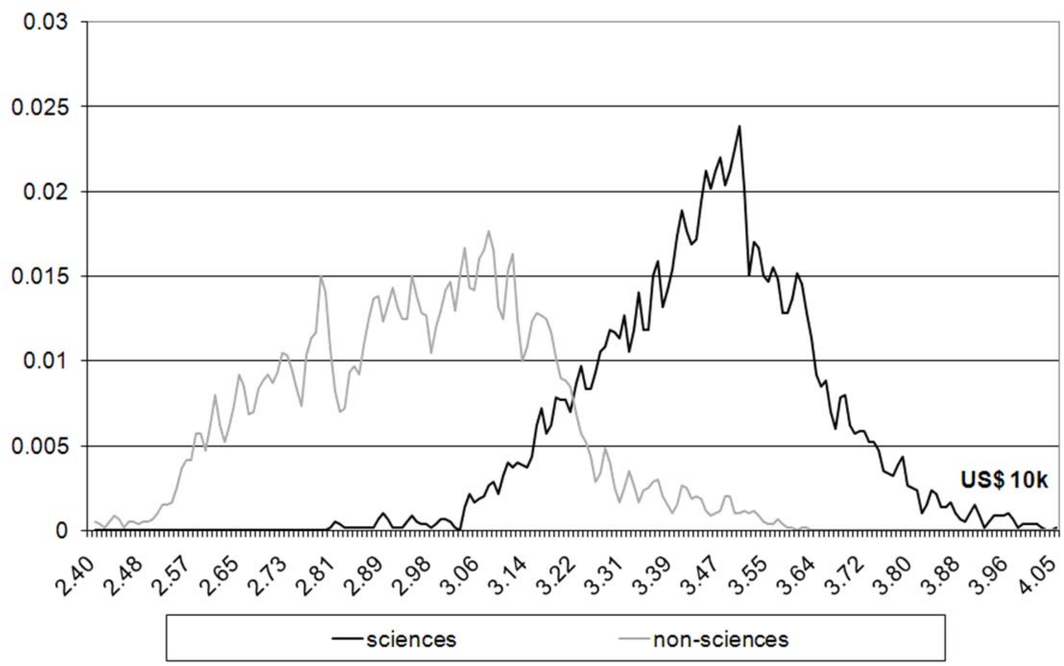

Figure B.2: MCMC distribution of the intercept of the cost equations

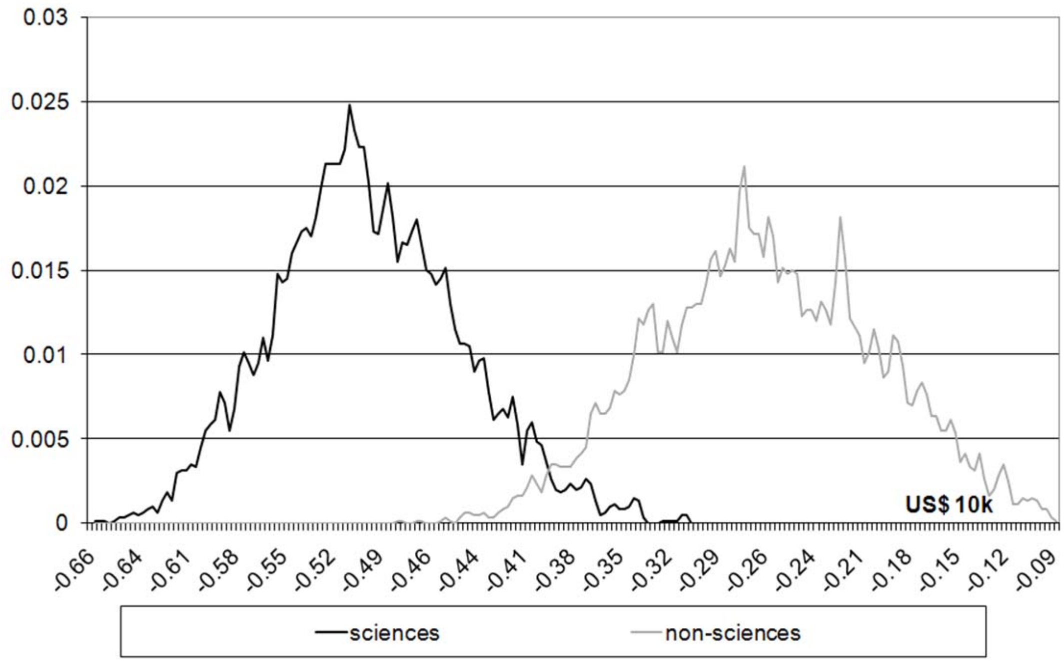

Figure B.3: MCMC distribution of the coefficient on parent's education, in the cost equations

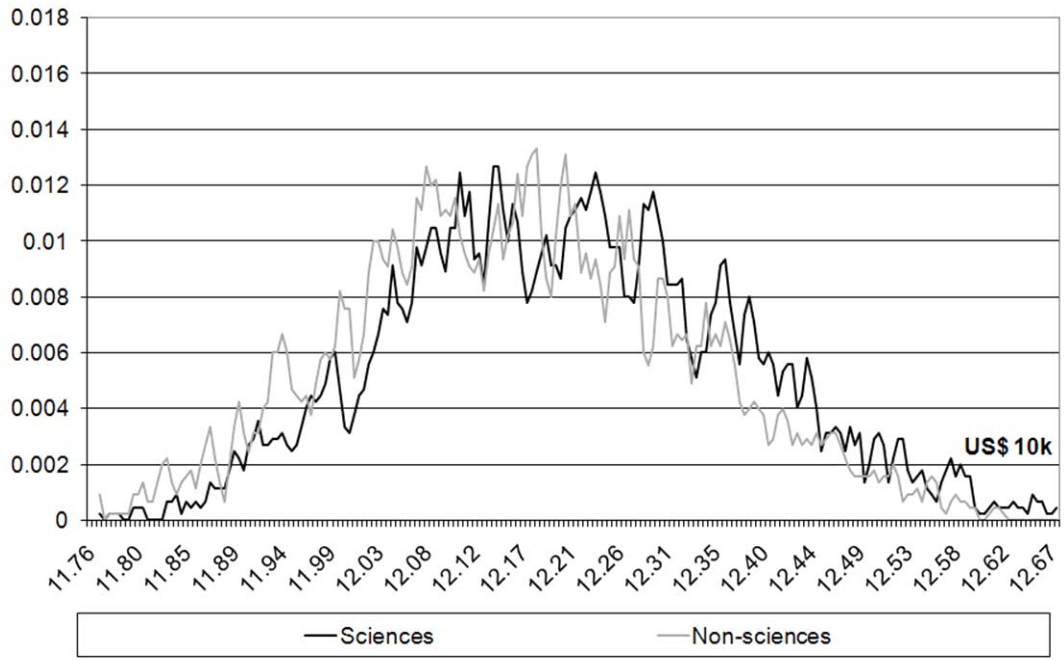

Figure B.4: MCMC distribution of the option value of college education by major, in period 0

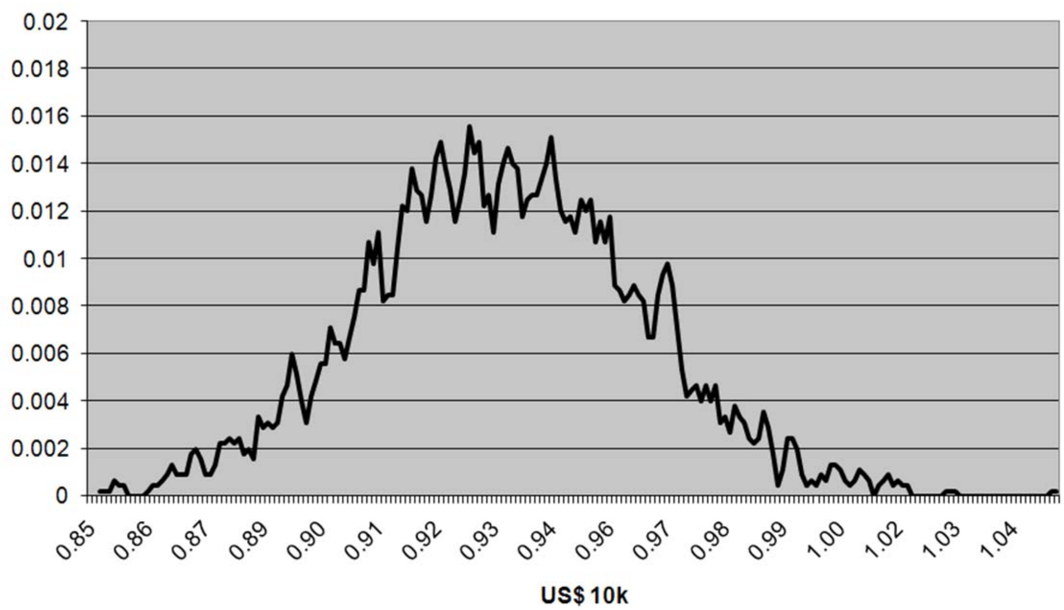

Figure B.5: MCMC distribution of the average welfare loss associated to the no-switching majors rule

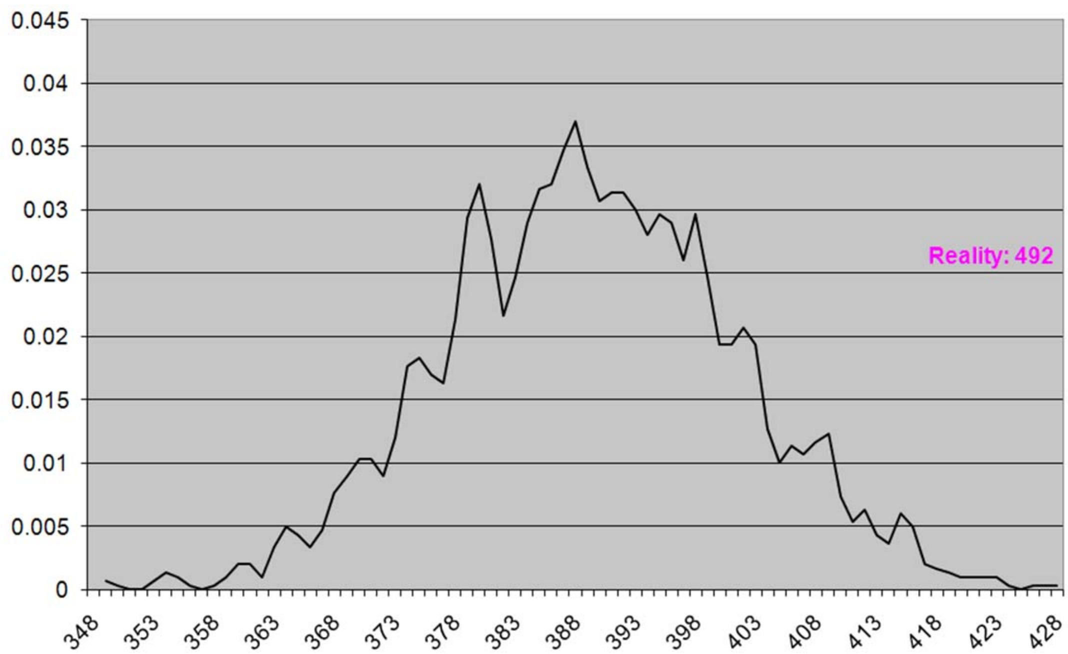

Figure B.6: MCMC distribution of the enrollment rate under the no-switching majors rule

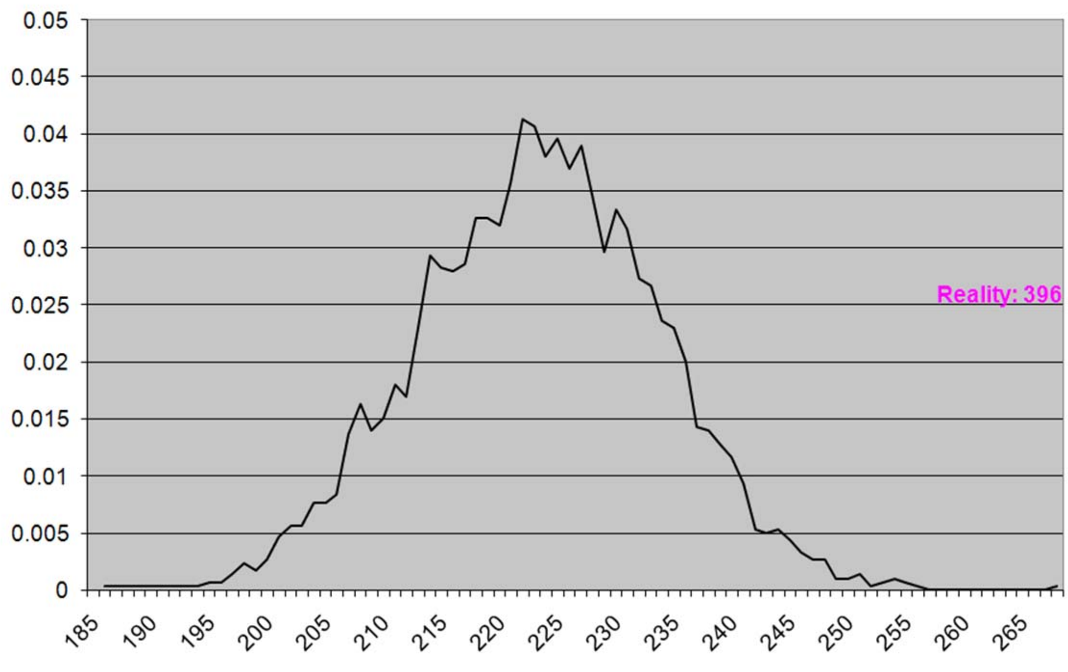

Figure B.7: MCMC distribution of the graduation rate under the no-switching majors rule

APPENDIX C

LIST OF MAJORS

Table C.1: List of majors - sciences

Natural Sciences		Social Sciences	
Field	FOS	Field	FOS
agriculture	101,110,113,199	architecture	201-206,160106
agronomy	102,103,108,109	regional studies	301-313,1111
husbandry, poultry	104-107	fine arts	1001,1003,1099
natural resources	114-198	painting	1002
biology	401,402,404,406, 412,418-498	music	1004-1006
zoology	407-410	dramatic arts	1007,1008
microbiology	403,411,414-417,499	commercial art	1009-1011,170700
computer science	701-705,799,140200	foreign languages	1101,1106-1115
engineering, general	901,903,905,919,925, 999,4904,160140	french	1102
chemical eng.	906,915,920,922	german	1103
civil eng.	904,908,992	spanish	1105
electrical eng.	909,160108	law	1401
mechanical eng.	907,910,913,917,921,923	english	1501,1505,1507
nursing	1203,70399	literature	1502-1504
public health	2004,2005,1201,1208, 1214,1215,1222,70401	rethoric	1506
optometry, dentistry,etc.	1209,1204-1207,1210, 1213,1216,1218, 1219,1221,1299	psychology	822,2001-2003, 2006-2010
pharmacy	1211	religion	1510
physical therapy	1212	philosophy	821,1509
laboratory tech's	1223-1225	law enforcement	2105,160605
nutrition	1306	soc. Sci.(general)	2201,2290
mathematics	1701-1703	anthropology	2202,2203,2211-2213
physics	1902,1904,1911-1913	economics	2204,111,517
chemistry	1905,1907-1909	history	2205
geology	911,912,914,916,918,924, 1913-1999	geography	2206
		political science	2207
		sociology	2208
		criminology	2209
		int. relations	2210
		urban studies	2214

Table C.2: List of majors - non-sciences

Business		Education	
Field	FOS	Field	FOS
business	112,501,513	education (general)	801,806;824-826,899,1601
commerce	502,140,100	elementary	802
finance	503-505,512	secondary	803-807
marketing	509;604,4903,179900	special	808-819
administration	506-508,510,511,827,828,1202,1307,2102,8270	kindergarten	823
communication	601	art	831
journalism	602	music	832
radio & tv	603	science & math	829,830,833,834;1508
media	605	specific subjects (health,commerce,...)	836-839
recreation	2103	physical education	835
		home economics	1301-1303
		family relations	1304,1305
		social work	2101,2104,2106,2199
		theology	2301-2304,2371
		interdisciplinary studies	4901
		natural sciences	1901,4902

Table C.3: Frequencies of majors by year - sciences

	1972	1973	1974	1975	1976	1977	1978	1979	graduation
agriculture	46	46	41	40	27	26	23	13	35
agronomy	13	14	18	33	26	25	14	13	17
husbandry, poultry	74	71	51	47	26	26	13	13	39
natural resources	67	50	34	28	13	17	14	10	21
biology	261	233	252	226	120	92	58	51	245
zoology	19	22	35	30	21	15	11	10	41
microbiology	22	24	33	37	32	34	25	13	29
computer science	101	81	65	73	50	54	68	86	31
engineering, general	99	88	58	35	23	47	33	32	54
chemical eng.	30	27	23	22	23	24	22	16	22
civil eng.	115	105	87	78	67	63	58	55	42
electrical eng.	123	114	94	86	64	63	54	64	45
mechanical eng.	151	133	124	128	101	82	82	81	68
nursing	66	75	311	267	193	169	148	158	142
public health	25	31	41	52	44	34	43	47	35
optometry, etc.	538	461	281	180	107	95	81	88	78
pharmacy	77	83	84	67	55	48	13	10	47
physical therapy	69	65	41	27	26	19	6	10	17
laboratory tech's	194	157	100	82	55	36	29	30	26
nutrition	11	12	14	25	19	17	12	10	17
mathematics	170	111	88	58	26	31	17	18	62
physics	47	41	35	32	16	8	11	11	26
chemistry	72	52	60	66	36	37	33	20	70
geology	49	47	47	55	42	33	35	30	34
architecture	98	87	76	70	52	38	23	22	39
regional studies	5	6	19	27	14	7	5	2	21
fine arts	28	30	25	30	12	16	11	14	40
painting	115	108	75	64	50	46	31	36	32
music	150	115	76	76	58	52	45	39	56
dramatic arts	64	59	52	47	31	31	24	21	35
commercial art	113	111	92	76	55	48	49	40	24
foreign languages	32	33	21	21	7	11	5	6	24
french	32	30	27	21	5	6	5	3	31
german	6	8	7	7	2	6	3	2	11
spanish	33	34	21	22	7	7	4	9	20
law	165	150	72	57	39	37	35	39	13
English	105	120	135	103	63	61	54	52	128
literature	13	15	21	31	22	11	12	9	24
rethoric	18	17	13	14	6	6	5	2	12
philosophy	20	23	24	20	18	22	14	10	27
religion	16	15	12	26	13	17	8	5	26
psychology	309	315	306	260	147	146	110	98	213
law enforcement	118	116	82	74	49	62	57	45	20
soc. Sci.(general)	49	49	55	47	14	18	14	10	45
anthropology	23	34	29	27	22	23	16	15	22
economics	59	68	87	88	45	44	31	25	91

Table C.4: Frequencies of majors by year - non-sciences

APPENDIX D

POSTERIOR DISTRIBUTIONS AND MCMC SEQUENCES

In this appendix I present the conditional posteriors of each block that constitutes the MCMC algorithm. Superscript G in the proposal distributions indicate that a block can be Gibbs sampled.

(i) Ω_0 :

$$\begin{aligned} posterior &\propto \left|\Omega_0^{-1}\right|^{N/2} \exp -\frac{1}{2} tr\left(\Omega_0^{-1} U_0^{s\prime} U_0^s\right) \\ proposal^G &= IWish\left(N+J+1, U_0^{s\prime} U_0^s\right) \end{aligned}$$

(ii) Λ_0^{-1} :

$$\begin{aligned} posterior &\propto \prod_{i=1}^{N}\left(f_v\left(v_{i0}|.\right)\prod_{j=1}^{J}\prod_{t=0}^{\tau} f_{\widehat{V}}\left(\widehat{V}_{it}|.\right)\prod_{t=\tau}^{T_{79}} f_w\left(w_{it}|.\right)\right) \\ proposal &= Wish\left(N+J+1, \frac{\Lambda_0^{(m-1)-1}}{N+J+1}\right) \end{aligned}$$

(iii) σ_k :

$$\begin{aligned} posterior &\propto \prod_{i=1}^{N}\left(\prod_{j=1}^{J}\prod_{t=0}^{\tau} f_G\left(G_{it}|.\right)\prod_{t=\tau}^{T_{79}} f_w\left(w_{it}|.\right)\right)^{d_{ik}} \\ proposal &\propto \prod_{i=1}^{N}\left(\prod_{j=1}^{J}\prod_{t=0}^{\tau} f_G\left(G_{it}|.\right)\prod_{t=\tau}^{T_{79}} f_w\left(w_{it}|.\right)^{d_{i03}, w_t>0}\right)^{d_{ik}} \end{aligned}$$

which implies $proposal = gamma\,(a, b)$

$$
\begin{aligned}
a &= 1 + 1/2 \sum_{i=1}^{N} \sum_{t=0}^{\tau_i} d_{ikt} \\
b &= \frac{1}{2} \sum_{i=1}^{N} \left[\begin{array}{l}
\sum_{t=0}^{\tau_i} d_{ikt} \left(G_{it} - \beta_k^{g\prime} X_i^g \right)^2 + \\
d_{i01} d_{ik0} \rho_0^{-2} \left(M_1^{G(0)} \left(G_{i1} - \beta_k^{g\prime} X_i^g \right) \right)^2 \sum_{t=1}^{T_{79}} 1\,(w_{i0t} > 0) + \\
d_{i02}\,(1 - d_{i01})\, \rho_0^{-2} \left(\sum_{t=2}^{T_{79}} 1\,(w_{i0t} > 0) \right) * \\
\left(\sum_{s=0}^{1} d_{iks} M_2^{G(0,s+1)} \left(G_{is+1} - \beta_k^{g\prime} X_i^g \right) \right)^2 + \\
(1 - d_{i02})\, \rho_0^{-2} \left(\sum_{t=3}^{T_{79}} 1\,(w_{ijt} > 0) \right) * \\
\left(\sum_{s=0}^{2} d_{iks} M_3^{G(0,s+1)} \left(G_{is+1} - \beta_k^{g\prime} X_i^g \right) \right)^2
\end{array} \right]
\end{aligned}
$$

(iv) ξ_k :

$$
\begin{aligned}
posterior &\propto \prod_{i=1}^{N} \left(\prod_{j=1}^{J} \prod_{t=0}^{\tau} f_G\,(G_{it}|.)^{d_{ik}}\, f_{\widehat{V}} \left(\widehat{V}_{it}|. \right) \prod_{t=\tau}^{T_{79}} f_w\,(w_{it}|.)^{d_{ik}} \right) \\
proposal &= 0.75 N \left(\xi_k^{(m-1)}, s_{\xi_k} \right) + 0.25 N\,(\mu_\xi, v_\xi)
\end{aligned}
$$

where:

$$
\begin{aligned}
\mu_\xi &= v_\xi \sum_{i=1}^{N} \sum_{t=0}^{\tau} d_{ikt} \left(\prod_{s=0}^{t} \theta_{d_s k} \right) (U_{ik0}^s + v_{ik0}) \left(\frac{G_{it} - \beta_k^{g\prime} X_i^g}{\sigma_k} \right) \\
v_\xi &= \left[\sum_{i=1}^{N} \sum_{t=0}^{\tau} d_{ikt} \left(\prod_{s=0}^{t} \theta_{d_s k} \right)^2 (U_{ik0}^s + v_{ik0})^2 \right]^{-1}
\end{aligned}
$$

(v) θ_k :

$$\begin{aligned} posterior &\propto 1\left(0 \leq \theta_k \leq \overline{\theta}\right) \prod_{i=1}^{N} \prod_{t=\tau}^{T_{79}} f_w\left(w_{it}|.\right)^{d_{ik}} \prod_{i=1}^{N} \prod_{t=0}^{\tau<2} f_{\widehat{V}}\left(\widehat{V}_{ijt}|.\right) \\ & \prod_{i=1}^{N} \prod_{j=1}^{J} \left(f_{\widehat{V}}\left(\widehat{V}_{ij2}|.\right)^{(1-d_{i02})} \prod_{t=1}^{\tau} f_G\left(G_{it}|.\right) \right)^{d_{ik}} \\ proposal &= N\left(\theta_k^{(m-1)}, v_\theta\right) \end{aligned}$$

(vi) β_k^w :

$$\begin{aligned} posterior &\propto \prod_{i=1}^{N} \left(f_{\widehat{V}}\left(\widehat{V}_{ik2}|.\right)^{d_{ik2}} \prod_{j=1}^{J} \prod_{t=0}^{\tau} f_{\widehat{V}}\left(\widehat{V}_{it}|.\right)^{1(t<2)} \prod_{t=\tau}^{T_{79}} f_w\left(w_{it}|.\right)^{d_{ik2}} \right) \\ proposal &= N\left(\beta_k^{w(m-1)}, v_{\beta_k^w}\right) \end{aligned}$$

(vii) β_k^g :

$$\begin{aligned} posterior &\propto \prod_{i=1}^{N} \left(\prod_{t=0}^{\tau} f_{\widehat{V}}\left(\widehat{V}_{ikt}|.\right)^{d_{ik}1(t>0)} \prod_{t=\tau}^{T_{79}} f_w\left(w_{it}|.\right)^{d_{ik}} \right) \\ proposal &= N\left(\beta_k^{g(m-1)}, v_{\beta_k^g}\right) \end{aligned}$$

(viii) A_k :

$$\begin{aligned} posterior &\propto \prod_{i=1}^{N} \left[\prod_{j=1}^{J} \prod_{t=0}^{\tau} f_{\widehat{V}}\left(\widehat{V}_{it}|.\right)^{1(t<2)} \left(f_{\widehat{V}}\left(\widehat{V}_{ik2}|.\right) \prod_{t=\tau}^{T_{79}} f_w\left(w_{it}|.\right) \right)^{d_{ik2}} \right] \\ proposal &= 0.75N\left(A_k^{(m-1)}, v_A\right) + 0.25 f_{A_k}^{p*} \end{aligned}$$

where $f^{p*}_{A_k} \propto \prod_{i=1}^{N} \prod_{t=\tau}^{T_{79}} f_w \left(w_{it}|.\right)^{d_{ik2}}$

(ix) ρ_k :

$$\begin{aligned} posterior &\propto \prod_{i=1}^{N} \prod_{t=\tau}^{T_{79}} f_w \left(w_{it}|.\right)^{d_{ik2}} \\ proposal &= gamma(a,b) \end{aligned}$$

where

$$\begin{aligned} a &= 1 + \frac{1}{2} \sum_{i=1}^{N} d_{ik\tau} \sum_{t=\tau_i}^{T_{79}} 1 \left(w_{ikt} > 0 \right) \\ b &= \frac{1}{2} \sum_{i=1}^{N} d_{ik\tau} * \\ &\left(\sum_{t=\tau_i}^{T_{79}} 1 \left(w_{ikt} > 0 \right) \left(w_{ikt} - \beta_k^{w\prime} X_i^w - M_\tau^{u(k)} U_{i0}^s - M_\tau^{G(k)\prime} \widetilde{G}_i^\tau - t A_k \right)^2 \right) \end{aligned}$$

(x) β_k^c :

$$\begin{aligned} posterior &\propto \prod_{i=1}^{N} \left(f_{\widehat{V}} \left(\widehat{V}_{i2}|.\right)^{d_{ik2}} \prod_{j=1}^{J} \prod_{t=0}^{\tau} f_{\widehat{V}} \left(\widehat{V}_{it}|.\right)^{1(t<2)} \right) \\ proposal &= 0.75 N \left(\beta_k^{c(m-1)}, v_c \right) + 0.25 f^{p*}_{\beta_k^c} \end{aligned}$$

where $f^{p*}_{\beta_k^c} \propto \prod_{i=1}^{N} f_{\widehat{V}} \left(\widehat{V}_{i2}|.\right)^{d_{ik2}}$

(xi) Latent variables:

$$
\begin{aligned}
f^{G}_{v_{i0}|\Upsilon,Z} &\propto f_v\left(v_{i0}|.\right)\prod_{t=0}^{\tau} f_G\left(G_{it}|.\right) \\
f^{G}_{\widehat{V}_{it}|\Upsilon,Z} &\propto 1\left(\widehat{V}_{ikt} > \max\left\{0; \max_{j\neq k}\widehat{V}_{ijt}\right\}\right) f_{\widehat{V}}\left(\widehat{V}_{it}|.\right) \\
f_{U^s_{i0}|\Upsilon,Z} &\propto f_{U^s}\left(U^s_{i0}|.\right)\prod_{j=1}^{J}\left(\prod_{t=0}^{\tau} f_{\widehat{V}}\left(\widehat{V}_{it}|.\right) f_G\left(G_{it}|.\right)\right)\prod_{t=\tau}^{T_{79}} f_w\left(w_{it}|.\right) \\
f^{p}_{U^s_{i0}} &= N\left(U^{s(m-1)}_{i0}, v_{U^s}\right)
\end{aligned}
$$

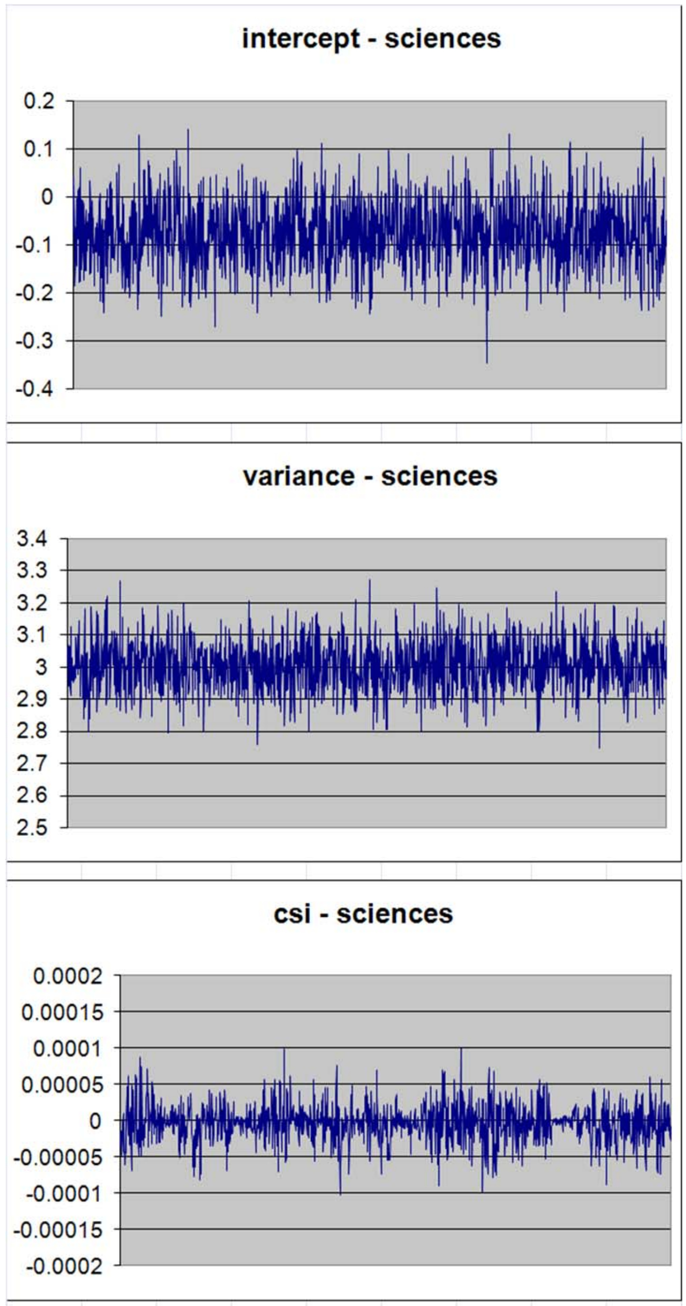

Figure D.1: MCMC - signal, sciences

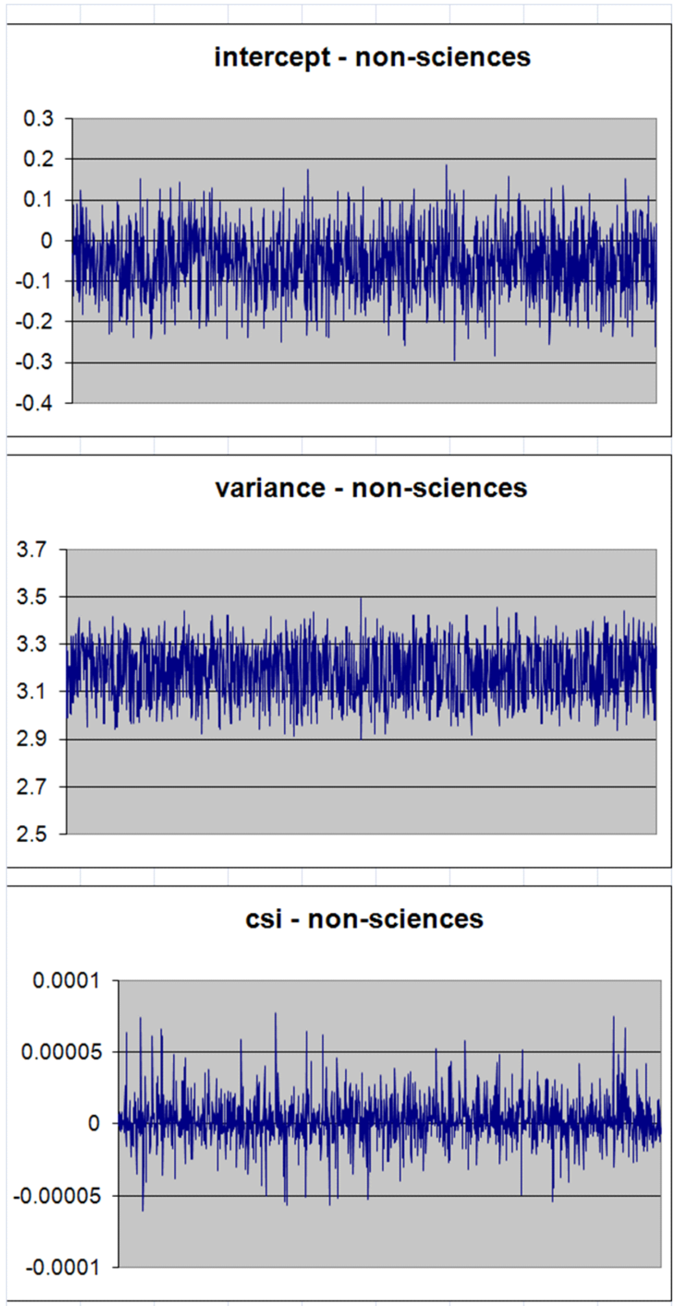

Figure D.2: MCMC - signal, non-sciences

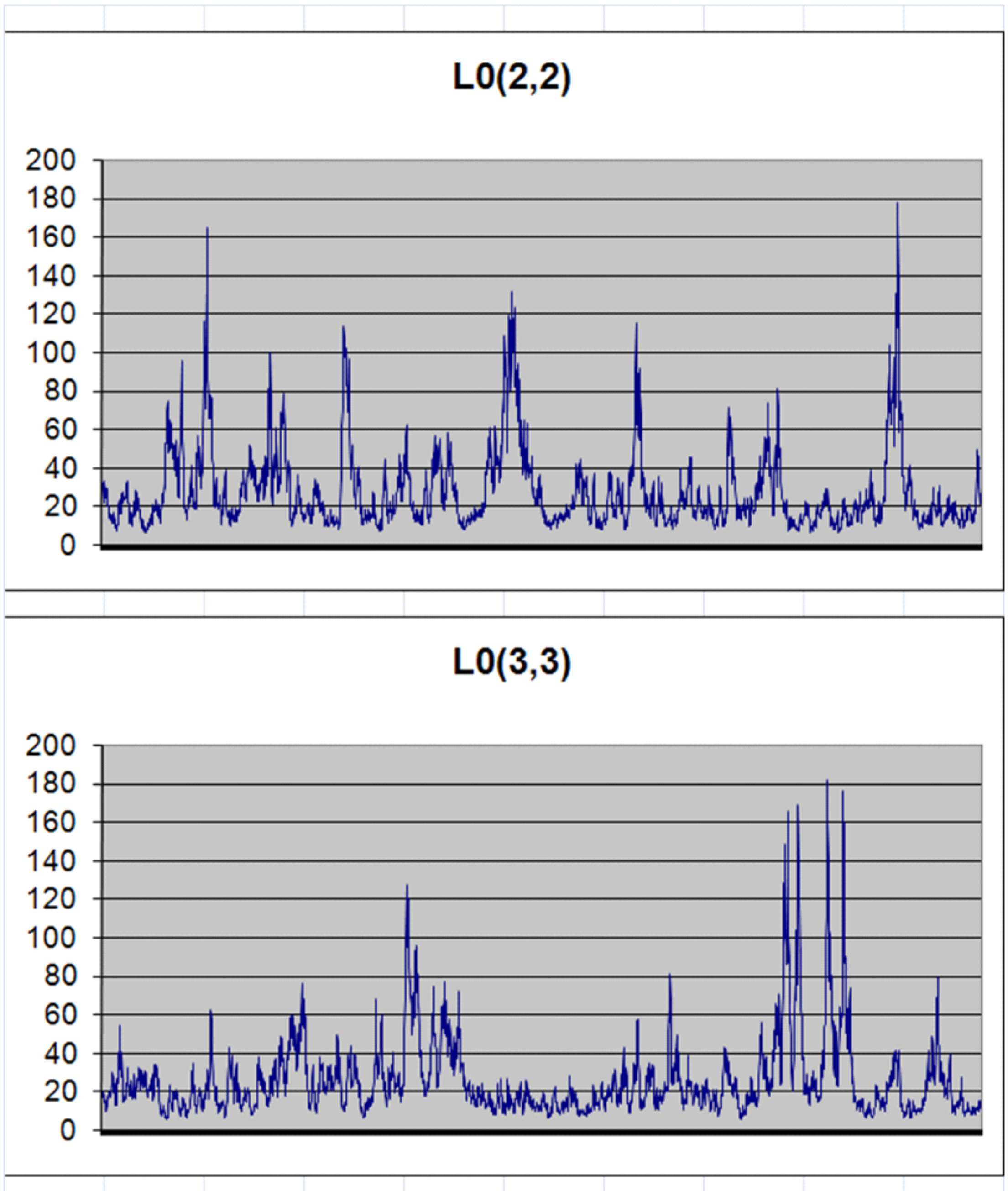

Figure D.3: MCMC - diagonal elements of Λ_0

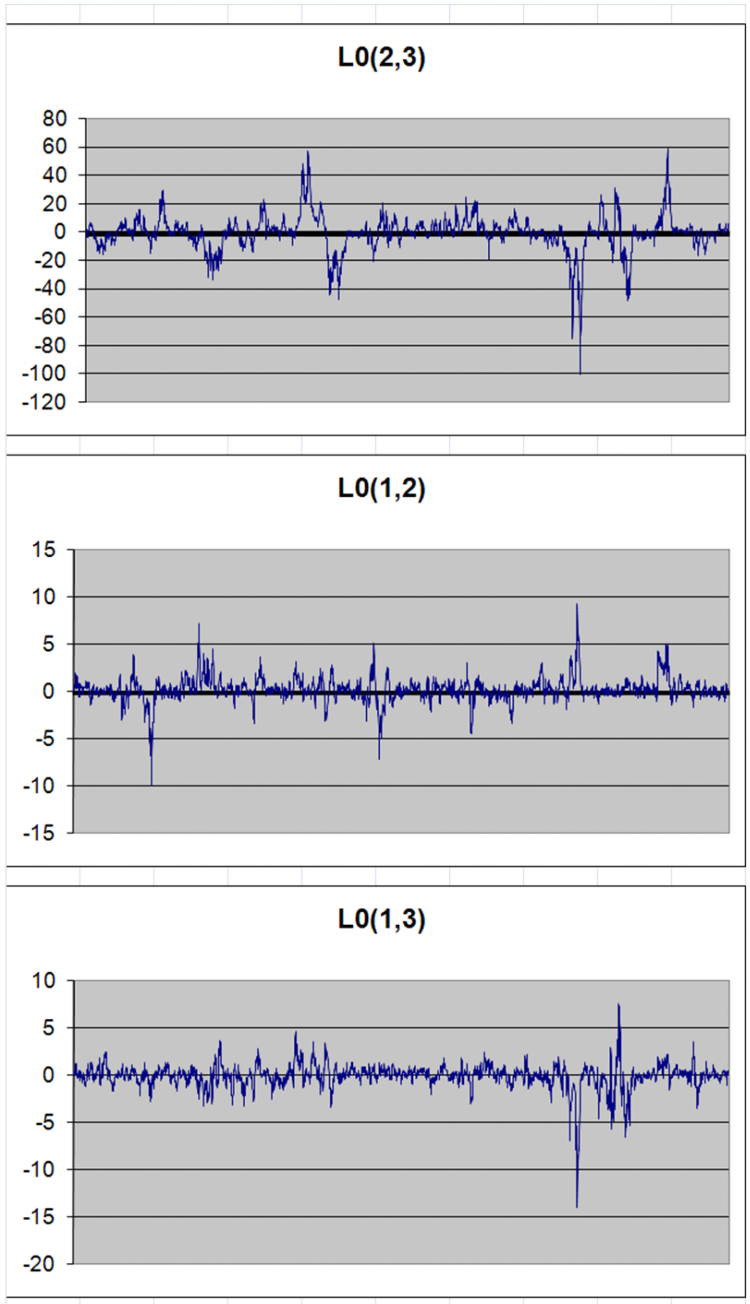

Figure D.4: MCMC - off-diagonal elements of Λ_0

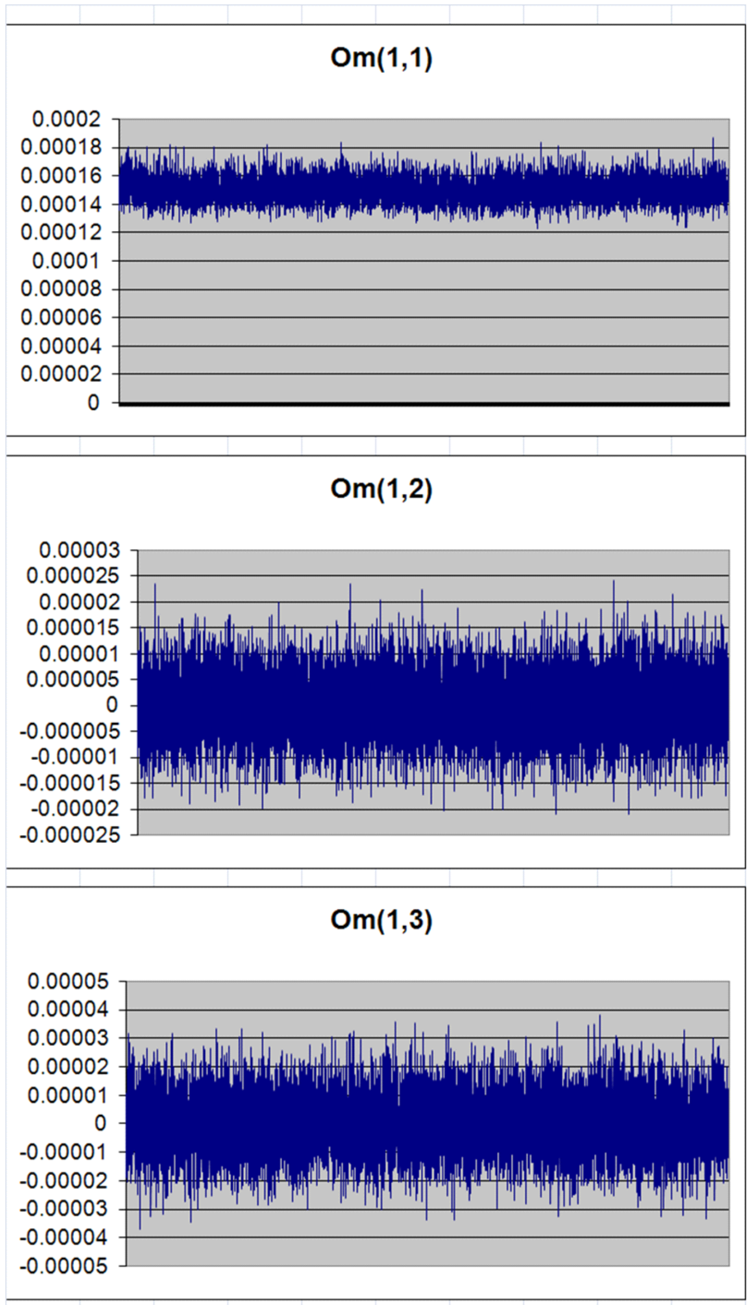

Figure D.5: MCMC - elements of Ω_0

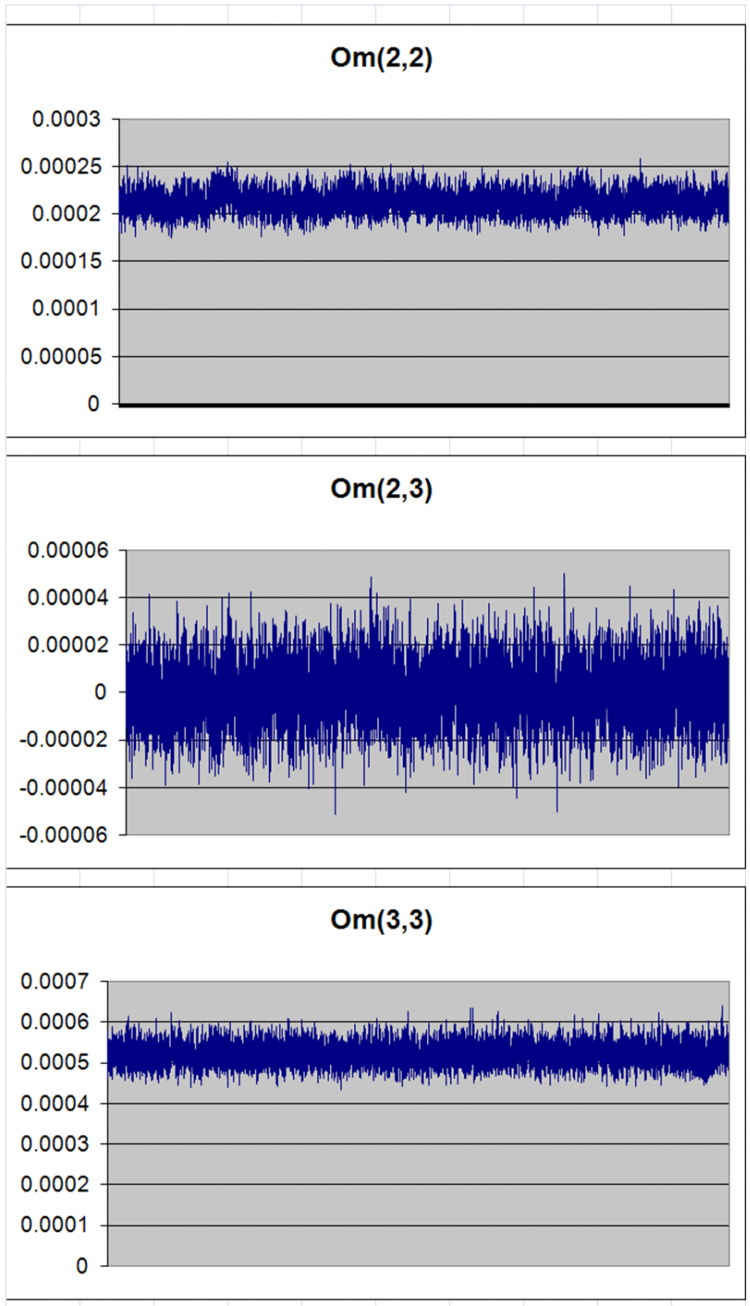

Figure D.6: MCMC - elements of Ω_0 (cont.)

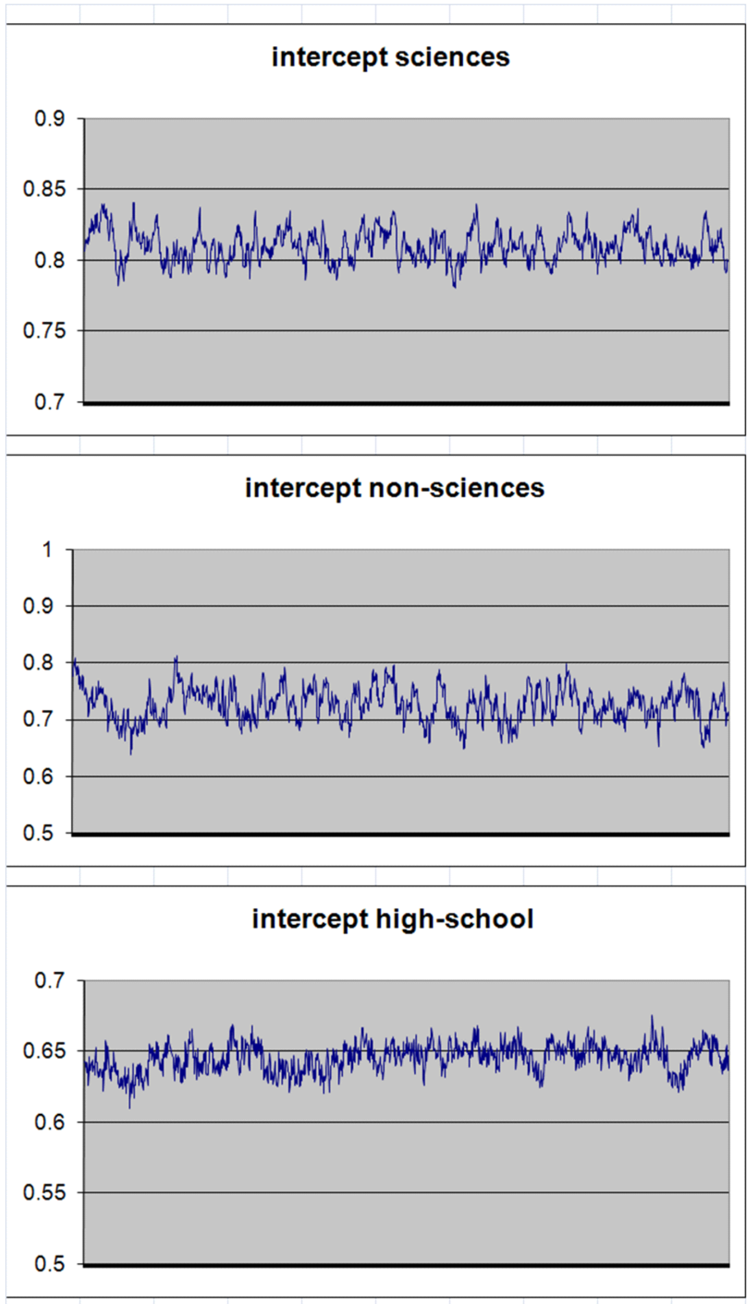

Figure D.7: MCMC - intercepts of the wage equations

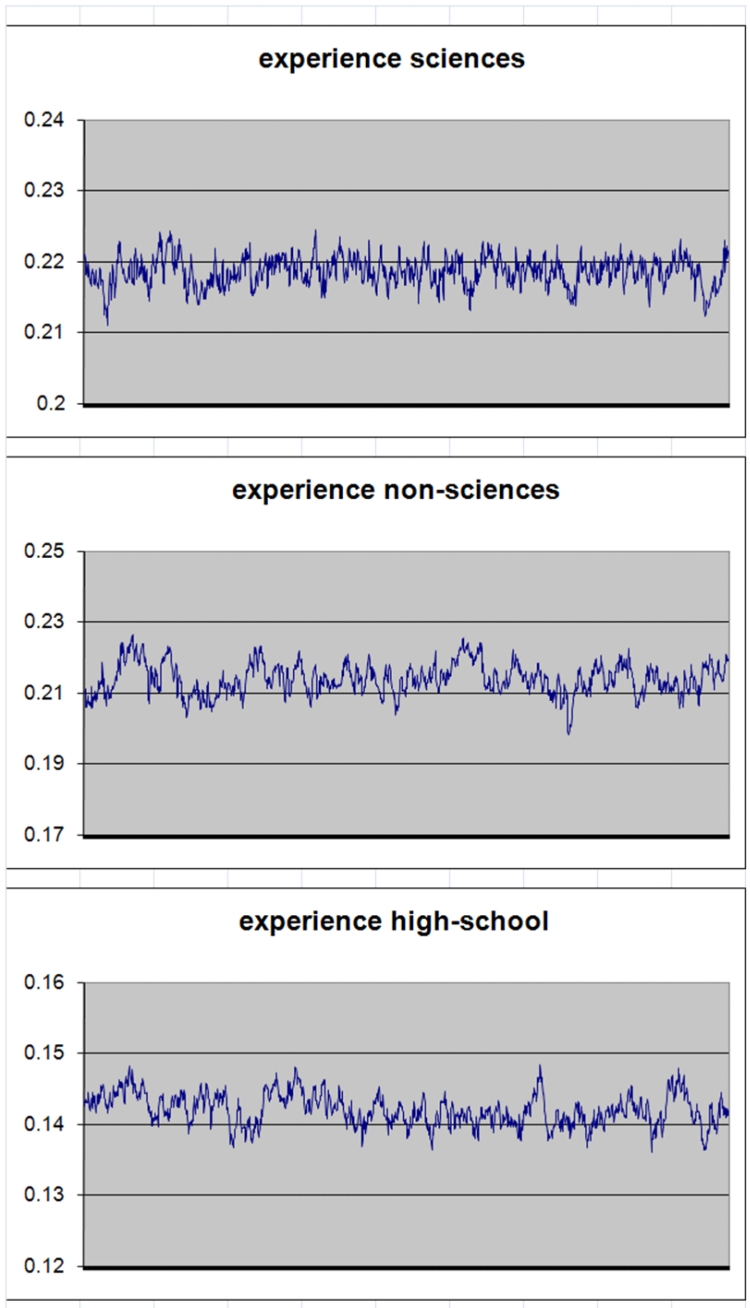

Figure D.8: MCMC - coefficients on experience in the wage equations

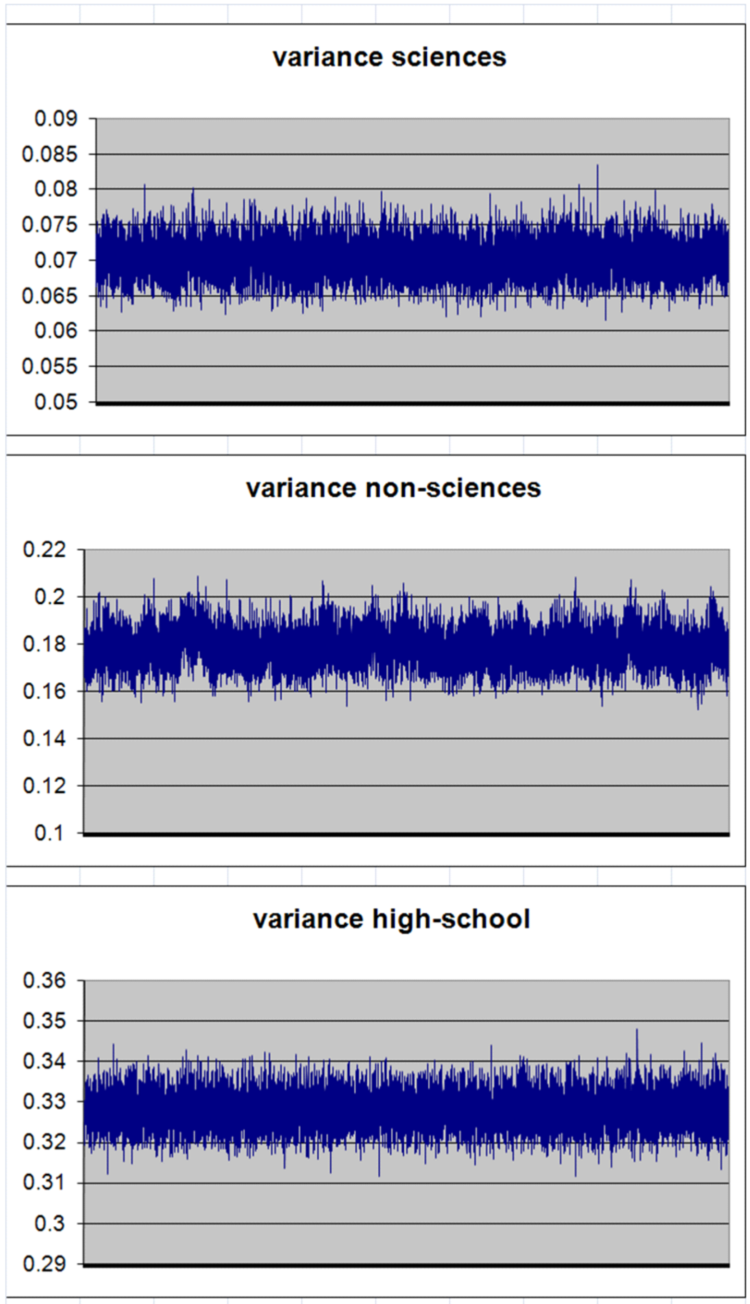

Figure D.9: MCMC - variances of the wage equations

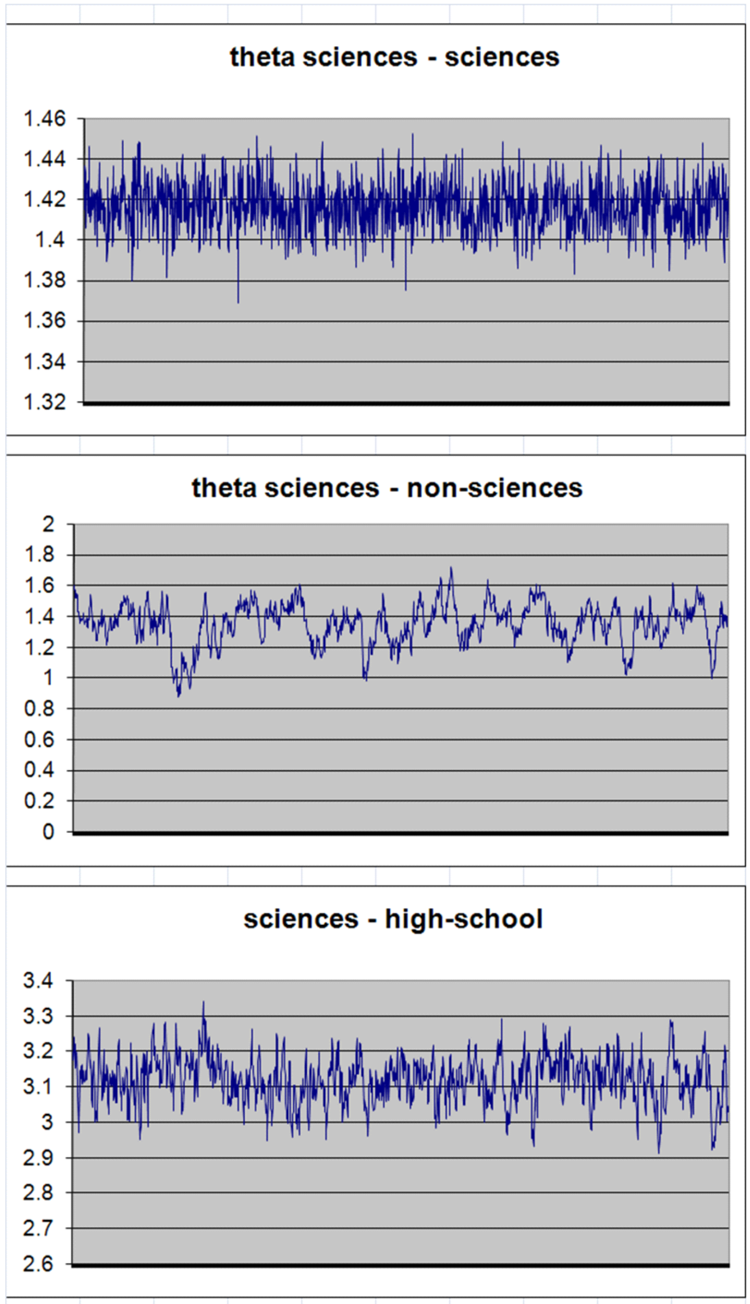

Figure D.10: MCMC - human capital accumulation rates

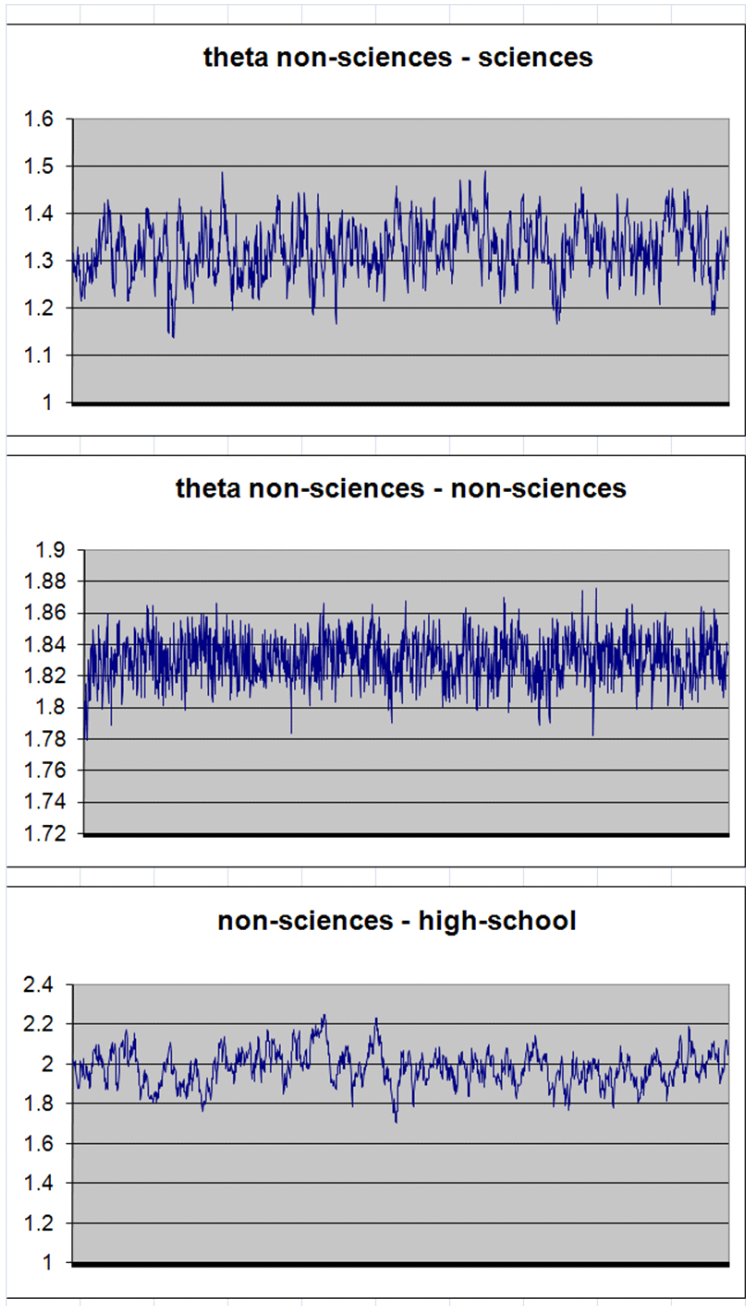

Figure D.11: MCMC - human capital accumulation rates (cont.)

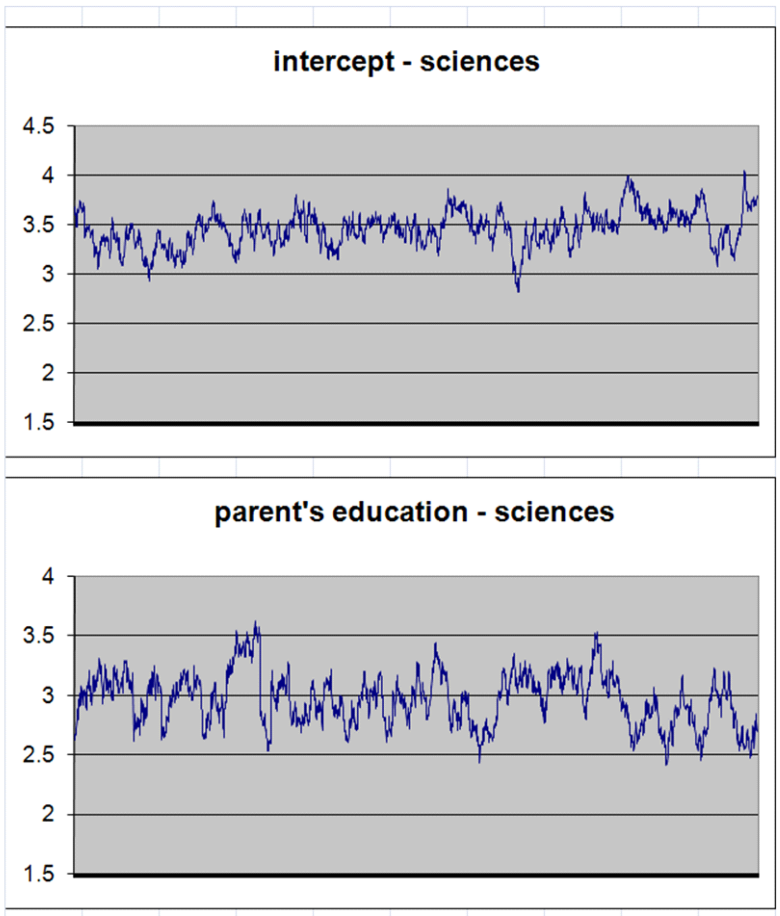

Figure D.12: MCMC - cost equation, sciences

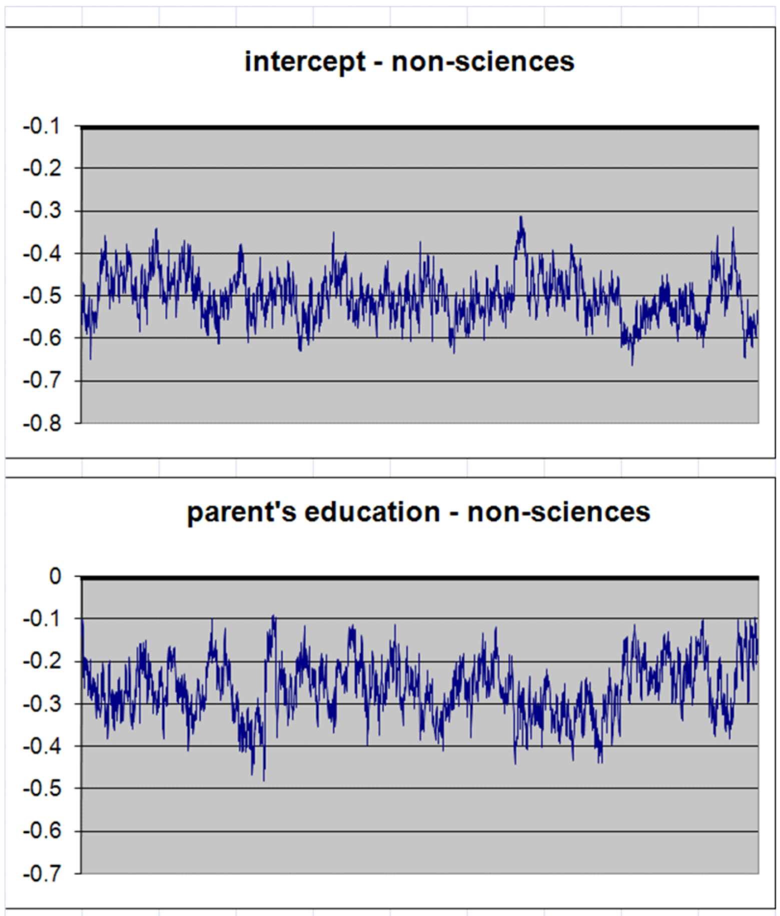

Figure D.13: MCMC - cost equation, non-sciences

APPENDIX E

SOME ROBUSTNESS CHECKS

In the empirical exercise realized in Section 3 above, we mentioned that the parameters (S, δ) of the model were exogenously fixed. In the subsample of white males, we could perform some robustness checks in order to verify if our conclusions would still hold if different values were chosen for them. The results displayed in the tables below suggest that most of the estimated parameters would not have their averages changed if we perturbed the original model in order to raise S from 10 to 100, or diminish δ from 0.9 to 0.75. The most sensitive parameters were the coefficients of the cost equation, especially in the case were S was changed (as expected). The variance matrices display also some variation, but the diagonals keep similar values and the off-diagonal elements are still of much smaller magnitude if compared to the diagonal elements.

Table E.1: Robustness checks of major-specific parameters

	Sciences			Non-Sciences			High-School		
	default	S=100	δ=0.75	default	S=100	δ=0.75	default	S=100	δ=0.75
σ	3.002	3.001	3.005	3.185	3.182	3.177	-	-	-
ξ *	-0.036	-0.025	-0.030	0.020	0.016	0.002	-	-	-
βg	-0.079	-0.084	-0.084	-0.055	-0.055	-0.054	-	-	-
βw	0.810	0.840	0.875	0.726	0.780	0.834	0.645	0.656	0.655
A	0.219	0.233	0.238	0.214	0.258	0.263	0.142	0.133	0.133
ρ	0.070	0.069	0.069	0.178	0.171	0.170	0.328	0.327	0.327
θj,hs	3.119	3.260	3.212	1.978	2.305	2.295	-	-	-
θj,sc	1.417	1.421	1.421	1.322	1.458	1.461	-	-	-
θj,ns	1.349	1.679	1.713	1.831	1.811	1.806	-	-	-
$\beta^c(1)$	3.449	11.004	2.816	2.958	11.216	2.790	-	-	-
$\beta^c(2)$	-0.502	-1.687	-0.634	-0.266	-1.273	-0.506	-	-	-

$^*(\times 10^4)$

Table E.2: Robustness checks of other parameters

	Ω (x104)			Λ		
	default	S=100	δ=0.75	default	S=100	δ=0.75
(1,1)	5.189	4.919	4.973	-	-	-
(1,2)	0.000	0.000	0.002	0.128	0.206	-0.216
(1,3)	0.004	0.005	-0.001	-0.145	-0.370	0.465
(2,2)	1.499	1.447	1.447	27.459	29.926	21.993
(2,3)	0.000	0.000	-0.001	-0.549	2.314	-1.787
(3,3)	2.122	2.076	2.088	25.926	33.958	41.282

APPENDIX F

IDENTIFICATION

In this appendix, I provide a formal proof of identification of most parameters of the model presented in Section 2, the only exceptions being the matrix Ω_0 and the cost coefficients $\left\{\beta_j^c\right\}_{j=1}^J$. The notation in this appendix will be:

$$D_i^t = \left(\left\{d_{ijs}\right\}_{j=1}^J \vdots_{s=0}^t\right)$$

$$\overline{G}_{jt,X,D} = E\left(G_{it}|X_i^w, D_i^t\right) \qquad \overline{G}^2_{jt,X,D} = E\left(G^2_{it}|X_i^w, D_i^t\right)$$

$$\overline{w}_{jt,X,D} = E\left(w_{it}|X_i^w, D_i^t\right) \qquad \overline{w}^2_{jt,X,D} = E\left(w^2_{it}|X_i^w, D_i^t\right)$$

$$\overline{U}^s_{jt,X,D} = E\left(U^s_{ijt}|X_i^w, D_i^{t-1}, d_{ijt}=1\right) \qquad \overline{U}^{s2}_{jt,X,D} = E\left(U^{s2}_{ijt}|X_i^w, D_i^{t-1}, d_{ijt}=1\right)$$

$$V^{w_{jt}}_{X,D} = var\left(w_{it}|X_i^w, D_i^t\right) \qquad V^{U_{jt}}_{X,D} = var\left(U^{s2}_{ijt}|X_i^w, D_i^{t-1}, d_{ijt}=1\right)$$

$$V^{G_t}_{X,D} = var\left(G_{it}|X_i^w, D_i^t\right) \qquad C^{wG}_{jt,D,X} = cov\left(G_{it}, w_{ijt}|D_i^2, X_i\right)$$

1. $\left\{A_j, \rho_j\right\}_{j=0}^J$

If we have two periods of income observations of a given subsample of individuals with observables $\left(X_i^w, D_i^t\right)$, we can form:

$$\begin{aligned}
\overline{w}_{jt,X,D} &= \beta_j^w X_i^w + \overline{U}^s_{j\tau,X,D} + A_j\left(t-\tau_i\right) \\
\overline{w}_{jt+1,X,D} &= \beta_j^w X_i^w + \overline{U}^s_{j\tau,X,D} + A_j\left(t+1-\tau_i\right) \\
&\Rightarrow \quad A_j = \overline{w}_{jt+1,X,D} - \overline{w}_{jt,X,D}
\end{aligned}$$

$$\rho^2 = \frac{var\left(w_{jt+1} - w_{jt}|X_i^w, D_i^t\right)}{2}$$

2. $\{\varphi_j\}_{j=1}^{J}$

Suppose I have access to two subsamples of individuals with histories $D_i^{2*} = \left(D_i^{1*}, d_{ij2} = 1\right)$ and $D_i^{2**} = \left(D_i^{1**}, d_{ij2} = 1\right) : D_i^{1*} \neq D_i^{1**}$. From the motion of U_{it}^s and the formulae for w_{ijt}, G_{ijt}

$$\begin{aligned}
\overline{w}_{j3,X,D} &= \beta_j^w X_i^w + \overline{U}_{j3,X,D}^s \\
&= \beta_j^w X_i^w + \frac{\xi_j^2 \lambda_{jj2}}{1+\xi_j^2 \lambda_{jj2}} \frac{\left[\overline{G}_{j3,X,D} - \beta_j^G X_i^g\right]}{\varphi_j} + \frac{\theta_{jj}}{1+\xi_j^2 \lambda_{jj2}} \overline{U}_{j2,X,D}^s \\
\overline{G}_{j3,X,D} &= \beta_j^G X_i^g + \varphi_j \overline{U}_{j2,X,D}^s \\
&\Rightarrow \overline{w}_{j3,X,D} - \frac{\overline{G}_{j3,X,D}}{\varphi_j} = \left(\beta_j^w - \frac{\beta_j^G}{\varphi_j}\right) X_i \qquad \text{(A1)} \\
\varphi_j &= \frac{\overline{G}_{j3,X,D^*} - \overline{G}_{j3,X,D^{**}}}{\overline{w}_{j3,X,D^*} - \overline{w}_{j3,X,D^{**}}}
\end{aligned}$$

In this case, I not only uniquely identify the set of parameters $\{\varphi_j\}_{j=1}^{J}$, but also the combination $\left(\beta_j^w - \frac{\beta_j^G}{\varphi_j}\right)$, since the left hand side of (A1) is observable and I still can vary the coordinates X_i. The key for identification here (and hereafter) is that after the beginning of the last period of college, no further decision takes place, and I can observe two outcomes $\left(G_{i3}\text{ and } w_{ij3}\right)$ for the same subsample of individuals with decision history D_i^2.

4. $\{\sigma_j\}_{j=1}^{J}$

Now, let's compute the observed second moments in a subsample of graduates in major j. From the equations

$$\begin{aligned}
\frac{G_{ij3} - \beta_j^g X_i}{\varphi_j} &= \theta_{jj} U_{ij2}^s + \theta_{jj} v_{ij2} + \frac{\sigma_j}{\varphi_j} \varepsilon_{i2} \\
w_{ij3} &= \beta_j^w X_i + \theta_{jj} U_{ij2}^s + \frac{\xi_j^2 \lambda_{jj2}}{1+\xi_j^2 \lambda_{jj2}} \left(\theta_{jj} v_{ij2} + \frac{\sigma_j}{\varphi_j} \varepsilon_{i2}\right) + \rho_j \eta_{i3}
\end{aligned}$$

we find the relation:

$$\sigma_j^2 \left[1 + \frac{\xi_j^2 \lambda_{jj2}}{1+\xi_j^2 \lambda_{jj2}}\right] = V_{X,D}^{G_t} - \varphi_j^2 \left(V_{X,D}^{w_{jt}} - \rho_j^2\right) \qquad \text{(AV(1))}$$

and from the equation

$$w_{ij3} - \frac{\left(G_{ij3} - \beta_j^g X_i\right)}{\varphi_j} - \beta_j^w X_i = -\frac{1}{1+\xi_j^2 \lambda_{jj2}} \left(\frac{\sigma_j}{\varphi_j} \varepsilon_{i2} + \theta_{jj} v_{ij2}\right) + \rho_j \eta_{i3}$$

we find

$$\sigma_j^2 \left(\frac{1}{1+\xi_j^2 \lambda_{jj2}}\right) = \varphi_j^2 \left(V_{X,D}^{w_{jt}} - \rho_j^2\right) + V_{X,D}^{G_t} - 2\varphi_j C_{jt,D,X}^{wG} \qquad \text{(AV(2))}$$

which implies (just by adding up AV(1) and AV(2)):

$$\sigma_j^2 = V_{X,D}^{G_t} - \varphi_j C_{jt,D,X}^{wG}$$

and that σ_j^2 is identified, since we have already identified φ_j. Furthermore, notice we can identify the combination $\xi_j^2 \lambda_{jj2}(D)$ [1], for every decision history D.

5. Λ_0 and $\{\theta_j\}_{j=1}^J$

It is immediately to notice λ_{000} is non-identified, as it does not appear in any equation of the model.

The next step is to isolate the subsample of individuals who have never changed majors during college, and compute the covariance between grades in periods 3 and 4, and the covariance between wages and grades in period 3:

1. $\xi_j^2 \lambda_{jj2} = \frac{\sigma_j^2}{\varphi_j^2\left(V_{X,D}^{w_{jt}} - \rho_j^2\right) + V_{X,D}^{G_t} - 2\varphi_j C_{jt,D,X}^{wG}} - 1$

$$
\begin{aligned}
cov\left(G_{ij3}, G_{ij2}|D^3, X_i\right) &= cov\left[\begin{array}{c} \theta_{jj}\varphi_j\left(U^s_{ij2}+v_{ij2}\right), \\ \varphi_j\left(U^s_{ij2}+v_{ij2}\right)+\sigma_j\varepsilon_{i2} \end{array} |D^2_i, X_i\right] \\
\frac{cov\left(G_{ij3}, G_{ij2}|D^3, X_i\right)}{\varphi_j} &= \theta_{jj}\varphi_j\lambda_{jj2}+\theta_{jj}\varphi_j V^{U_{jt}}_{X,D} \\
&\quad +\theta_{jj}\sigma_j cov\left(U^s_{ij2}, \varepsilon_{i2}|D^2_i, X_i\right)
\end{aligned}
$$

On the other hand:

$$
\begin{aligned}
cov\left(w_{ij3}, G_{ij2}|D^2, X_i\right) &= cov\left(U^s_{ij3}, \varphi_j U_{ij2}+\sigma_j\varepsilon_{i2}|D^2, X_i\right) \\
&= cov\left(\begin{array}{c} \theta_{jj}U^s_{ij2}+\left(\frac{\theta_{jj}\lambda_{jj2}\xi^2_j}{1+\xi^2_j\lambda_{jj2}}\right)v_{ij2}, \\ \varphi_j U^s_{ij2}+\varphi_j v_{ij2}+\sigma_j\varepsilon_{i2} \end{array} |D^2, X_i\right) \\
&= \frac{\theta_{jj}\varphi_j\xi^2_j\lambda^2_{jj2}}{1+\xi^2_j\lambda_{jj2}}+\theta_{jj}\varphi_j V^{U_{jt}}_{X,D} \\
&\quad +\theta_{jj}\sigma_j cov\left(U^s_{ij2}, \varepsilon_{i2}|D^2, X_i\right)
\end{aligned}
$$

which implies:

$$
\theta_{jj}\lambda_{jj2} = \left(\frac{cov\left(G_{ij3}, G_{ij2}|D^3, X_i\right)}{\varphi^2_j} - \frac{cov\left(w_{ij3}, G_{ij2}|D^2, X_i\right)}{\varphi_j}\right)\left(1+\xi^2_j\lambda_{jj2}\right)
$$

and because we had already identified $\xi^2_j\lambda_{jj2}$, we now identify $\theta_{jj}\lambda_{jj2}$. However, we also know that $\frac{\xi^2_j\lambda_{jj2}}{\theta_{jj}\lambda_{jj2}} = \theta_{jj}\left(\frac{\varphi_j}{\sigma_j}\right)^2$, or $\theta_{jj} = \frac{\xi^2_j\lambda_{jj2}}{\theta_{jj}\lambda_{jj2}}\left(\frac{\sigma_j}{\varphi_j}\right)^2$, and therefore θ_{jj} is identified (and so is λ_{jj2})

Now, from the law of motion of Λ_{it}, we know that choice $d_{ijt} = 1$ implies the elements of Λ_{it} to be of the type:

$$\lambda_{kmt+1} = \theta_{jk}\theta_{jm}\left[\lambda_{kmt} - \frac{\xi_j \lambda_{kjt}\lambda_{mjt}}{(1+\xi_j \lambda_{jjt})}\right]$$

In the subsample of non-switchers, this means that:

$$\begin{aligned}
\lambda_{jj3}\left(d_{ij0} = d_{ij1} = d_{ij2} = 1\right) &= \lambda_{jj3}(j,j,j) \\
&= \frac{\theta_{jj}^4 \lambda_{jj0}}{1+\left(1+\theta_{jj}^2\right)\xi_j \lambda_{jj0}} \\
&\Rightarrow \lambda_{jj0}(j,j,j) = \frac{\lambda_{jj3}}{\theta_{jj}^4 - \left(1+\theta_{jj}^2\right)\xi_j \lambda_{jj3}}
\end{aligned}$$

therefore λ_{jj0} is identified.

Turning now to the subsample of individuals whose decision history contains a chage of majors in the last period of college, we have the following relations:

$$
\begin{aligned}
cov\left(G_{ik2}, G_{ij3}|D^2, X_i\right) &= cov\left(\begin{array}{c} \varphi_k U^s_{ik2} + \varphi_k v_{ik2} + \sigma_k \varepsilon_{i2}, \\ \theta_{jj}\varphi_j U^s_{ij2} + \theta_{jj}\varphi_j v_{ij2} \end{array} |D^2, X_i\right) \\
&= \theta_{jj}\varphi_j\varphi_k\lambda_{jk2} + \\
&\quad \theta_{jj}\varphi_j cov\left(\varphi_k U^s_{ik2} + \sigma_k\varepsilon_{i2}, U^s_{ij2}|D^2, X_i\right) \\
cov\left(G_{ik2}, w_{ij3}|D^2, X_i\right) &= cov\left(\begin{array}{c} \varphi_k U^s_{ik2} + \varphi_k v_{ik2} + \sigma_k \varepsilon_{i2}, \\ \theta_{jj} U^s_{ij2} + \left(\frac{\theta_{jj}\lambda_{jj3}\xi_j}{1+\xi_j\lambda_{jj2}}\right) v_{ij2} \end{array} |D^2, X_i\right) \\
&= \left(\frac{\theta_{jj}\lambda_{jj2}\xi_j}{1+\xi_j\lambda_{jj2}}\right)\varphi_k\lambda_{jk2} + \\
&\quad \theta_{jj} cov\left(\varphi_k U^s_{ik2} + \sigma_k\varepsilon_{i2}, U^s_{ij2}|D^2, X_i\right) \\
&= \left(\frac{\theta_{jj}\lambda_{jj2}\xi_j}{1+\xi_j\lambda_{jj2}}\right)\varphi_k\lambda_{jk2} + \frac{cov\left(G_{ik2}, G_{ij3}|D^2, X_i\right)}{\varphi_j} \\
&\quad - \frac{\theta_{jj}\varphi_j\varphi_k\lambda_{jk2}}{\varphi_j}
\end{aligned}
$$

$$
\lambda_{jk2} = \left(1 + \xi_j\lambda_{jj2}\right)\left(\frac{cov\left(G_{ik2}, G_{ij3}|D^2, X_i\right)}{\theta_{jj}\varphi_k\varphi_j} - \frac{cov\left(G_{ik2}, w_{ij3}|D^2, X_i\right)}{\theta_{jj}\varphi_k}\right)
$$

So far we have identified λ_{jk2} and λ_{jj2} for all possible decision histories. If we restrict our attention to the subsample of people who have chosen (k, k, j), we observe

$$
\begin{aligned}
\lambda_{jk2}(k, k, j) &= \frac{\theta^2_{kk}\theta^2_{kj}\lambda_{kj0}}{1 + \left(1 + \theta^2_{kk}\right)\xi_k\lambda_{kk0}} \\
&\Rightarrow \theta^2_{kj}\lambda_{kj0} = \frac{\left[1 + \left(1 + \theta^2_{kk}\right)\xi_k\lambda_{kk0}\right]\lambda_{jk3}(k, k, k, j)}{\theta^2_{kk}}
\end{aligned}
$$

which means $\theta_{kj}^2\lambda_{kj0}$ is identified. We also know that

$$\begin{aligned}
\lambda_{kk3}(k,k,k) &= \frac{\theta_{kk}^4\lambda_{kk0}}{1+\left(1+\theta_{kk}^2\right)\xi_k\lambda_{kk0}} \\
\frac{\lambda_{jj3}(k,k,j)}{\lambda_{jj0}} &= \theta_{kj}^4\left[1-\frac{\left(1+\theta_{kk}^2\right)\xi_k}{1+\left(1+\theta_{kk}^2\right)\xi_k\lambda_{kk0}}\frac{\lambda_{kj0}^2}{\lambda_{jj0}}\right] \\
\theta_{kj}^4 &= \frac{\lambda_{jj3}(k,k,k,j)}{\lambda_{jj0}}+\frac{\xi_k\lambda_{kk0}\left(1+\theta_{kk}^2\right)}{\lambda_{jj0}}\frac{\lambda_{jk3}^2(k,k,j)}{\lambda_{kk3}(k,k,k)}
\end{aligned}$$

After identifying θ_{kj}^4, we can use it to identify λ_{kj0}^2 :

$$\lambda_{kj0}^2=\frac{\left(\theta_{kj}^4\lambda_{jj0}-\lambda_{jj3}(k,k,j)\right)\theta_{kk}^4\lambda_{kk0}}{\theta_{kj}^4\left(1+\theta_{kk}^2\right)\xi_k\lambda_{kk3}(k,k,k)}$$

and because $\theta_{kj}>0$, we also know that $\operatorname{sign}\left(\lambda_{kj0}\right)=\operatorname{sign}\left(\lambda_{jk3}(k,k,j)\right)$, and therefore λ_{kj0} is identified (and so is the whole matrix Λ_0, except λ_{000}).

The next step is to identify the parameters $\theta_{jk}: j\neq k$. Call

$$R=1+\xi_j\lambda_{jj0}+\left(1+\theta_{kk}^2\right)\theta_{jk}^2\xi_k\lambda_{kk0}\left[1+\xi_j\lambda_{jj0}\left(1-\rho_{kj0}^2\right)\right]$$

In the subsample of students with decision history (j,k,k), we observe

$$\begin{aligned}
\lambda_{kk3}(j,k,k) &= \theta_{jk}^2\theta_{kk}^4\lambda_{kk0}\left[\frac{1+\xi_j\lambda_{jj0}\left(1-\rho_{kj0}^2\right)}{R}\right] \\
\theta_{jk}^2 &= \frac{\lambda_{kk3}(j,k,k)\left(1+\xi_j\lambda_{jj0}\right)}{\left[1+\xi_j\lambda_{jj0}\left(1-\rho_{kj0}^2\right)\right]\left[\theta_{kk}^4-\lambda_{kk3}(j,k,k)\left(1+\theta_{kk}^2\right)\xi_k\right]\lambda_{kk0}}
\end{aligned}$$

where the RHS is already identified, and because θ_{jk} must be positive, it is also identified.

6. β_j^w and β_j^G :

In order to disentangle the combination $\left(\beta_j^w - \frac{\beta_j^G}{\varphi_j}\right)$ (already identified, together with φ_j), let's compute the moments:

$$
\begin{aligned}
E\left(w_{ij3}G_{ik2}|D^2, X_i\right) &= \frac{\varphi_k\theta_{jj}\xi_j\lambda_{jj2}}{1+\xi_j\lambda_{jj2}}\lambda_{jk2} + \beta_j^{w\prime}X_i\overline{G}_{k2,X,D} + \beta_k^{G\prime}X_i\overline{w}_{j3,X,D} \\
&\quad -\beta_k^{G\prime}X_iX_i'\beta_j^w + \theta_{jj}E\left(U_{ij2}^s\left(\varphi_kU_{ik2}^s + \sigma_k\varepsilon_{i2}\right)|D^2, X_i\right) \\
E\left(G_{ij3}G_{ik2}|D^2, X_i\right) &= \varphi_j\theta_{jj}\varphi_k\lambda_{jk2} + \beta_j^{G\prime}X_i\overline{G}_{j3,X,D} - \beta_k^{G\prime}X_iX_i'\beta_j^G \\
&\quad +\beta_k^{G\prime}X_iE\left(G_{ij3}|D^2, X_i\right) \\
&\quad +\varphi_j\theta_{jj}E\left(U_{ij2}^s\left(\varphi_kU_{ik2}^s + \sigma_k\varepsilon_{i2}\right)|D^2, X_i\right)
\end{aligned}
$$

$$
\begin{aligned}
&\beta_k^{G\prime}X_i\left(\overline{w}_{j3,X,D} - \frac{\overline{G}_{j3,X,D}}{\varphi_j}\right) + \left(\frac{\varphi_j-1}{\varphi_j}\right)\beta_k^{G\prime}X_iX_i'\beta_j^w \\
= \; & cov\left(w_{ij3}, G_{ik2}|D^2, X_i\right) - \frac{cov\left(G_{ij3}, G_{ik2}|D^2, X_i\right)}{\varphi_j} + \frac{\theta_{jj}\varphi_k\lambda_{jk2}}{1+\xi_j\lambda_{jj2}}
\end{aligned}
$$

In this equation, the RHS is observed/ identified. If I take subsamples with two distinct histories and same X_i, D^2, D^{2*}, call Δ^* the difference operator between moments conditional in different histories (i.e. $\Delta^*F\left(., D^2\right) = F\left(., D^2\right) - F\left(., D^{2*}\right)$) and subtract the second equation from the first, I get rid of the squared term $\frac{(\varphi_j-1)}{\varphi_j}\beta_k^{G\prime}X_iX_i'\beta_j^w$:

$$\beta_k^{G\prime} X_i = \left(\Delta^* \overline{w}_{j3,X,D} - \frac{\Delta^* \overline{G}_{j3,X,D}}{\varphi_j} \right)^{-1} \left[\begin{array}{c} \theta_{jj} \varphi_k \Delta^* \left(\frac{\lambda_{jk2}(D^2)}{1+\xi_j \lambda_{jj2}(D^2)} \right) \\ +\Delta^* cov \left(w_{ij3}, G_{ik2} | D^2, X_i \right) \\ -\varphi_j^{-1} \Delta^* cov \left(G_{ij3}, G_{ik2} | D^2, X_i \right) \end{array} \right]$$

This is a linear equation on the unknown vector $\beta_k^{G\prime}$. I can identify its value by varying the coordinates of X_i. The value of $\beta_j^{w\prime}$ is therefore also identified.■

REFERENCES

[1] ARCIDIACONO, P. (2004) "Ability Sorting and Returns to College Major". *Journal of Econometrics*, 121

[2] BECKER, G. (1962) "Investment in Human Capital: A Theoretical Analysis". *Journal of Political Economy*, v.70, n.5

[3] BECKER, G. (1964) **Human Capital**. Chicago: University of Chicago Press

[4] BEN PORATH, Y. (1967) "The Production of Human Capital and the Life Cycle of Earnings" *Journal of Political Economy*, v.75, n.4

[5] CAMERON, S. and J. Heckman (1998) "Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males". *Journal of Political Economy*, v.106, n.2

[6] CAMERON, S. and J. Heckman (2001) "The Dynamics of Educational Attainment for Black, Hispanic and White Males". *Journal of Political Economy*, v.109, n.3

[7] CUNHA, F., J.Heckman and S. Navarro-Lozano (2005) "Separating Uncertainty from Heterogeneity in Life Cycle Earnings". *Oxford Economic Papers* v.57, n.2.

[8] DeGroot, M. (2004) **Optimal Statistical Decisions**. Willey Classics Library. 1^{st} ed: McGraw Hill, 1970

[9] GEWEKE, J. (1989) "Bayesian Inference in Econometric Models using MonteCarlo Integration". *Econometrica*, v.57, n.6

[10] GITTINS, J.C. (1979) "Bandit Processes and Dynamic Allocation Indeces" *Journal of the Royal Statistical Society. Series B (Methodological)*, v.41, n.2

[11] GITTINS, J.C. and D.M. Jones (1979) "A Dynamic Allocation Index for the Discounted Multiarmed Bandit" *Biometrika*, v.66, n.3

[12] HOTZ, V.J. and R. Miller (1993) "Conditional Choice Probabilities and the Estimation of Dynamic Models". *The Review of Economic Studies*, v.60, n.3.

[13] IMAI, S., N. Jain, and A. Ching (2007): "Bayesian Estimation of Dynamic Discrete Choice Models". *Mimeo.*

[14] JOVANOVIC, B. (1979) "Job Matching and the Theory of Turnover". *Journal of Political Economy*, v.87, n.5

[15] KEANE, M. and K. Wolpin (1994) "The Solution and Estimation of Discrete Choice Dynamic Programming Models by Simulation and Interpolation: Monte Carlo Evidence". *The Review of Economics and Statistics*, v.76, n.4

[16] GAMMERMAN D, Lopes H.F. (2006) **Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference**. Texts in statistical science series, 2nd edn. Springer, New York..

[17] MARSCHAK, J. (1953) "Economic Measurements for Policy and Prediction". *In* Studies in econometric methods, eds. T. Koopmans and W.C. Hood. Cowles Comission.

[18] MILLER, R. (1984) "Job Matching and Occupational Choice" *Journal of Political Economy*, v.92, n.6

[19] MINCER, J. (1958) "Investment in Human Capital and Personal Income Distribution," *Journal of Political Economy*, v.66, n.4

[20] MINCER, J. (1974) **Schooling, Experience and Earnings**, New York, NBER Press

[21] NEAL, D. (1999) "The Complexity of Job Mobility Among Young Men" *Journal of Labor Economics*, v.17, n.2

[22] NORETS, A. (2006) "Inference in Dynamic Discrete Choice Models with Serially Unobserved State Variables". *Mimeo*

[23] ROY, A.D. (1951) "Some Thoughts on the Distribution of Earnings". *Oxford Economic Papers, New Series*, v.3, n.2.

[24] RUST, J. (1987) "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher". *Econometrica*, v.55, n.5.

[25] WILLIS, R. and S. Rosen (1979) "Education and Self-Selection". *Journal of Political Economy*, v.87, n.5.

[26] ZELLNER, A. (1971) **An Introduction to Bayesian Inference in Econometrics**. Willey Classics Library.