U N I V E R S I T Y O F L A T V I A
F A C U L T Y O F E C O N O M I C S A N D M A N A G E M E N T

ALEXANDER TARVID

MODELLING INDIVIDUAL BEHAVIOUR IN EUROPEAN LABOUR–EDUCATION MARKET SYSTEM

Doctoral Dissertation

Submitted for the Doctoral Degree in Economics, Subfield – Econometrics

Supervisor: Dr.math. Mara Gulbe

R I G A 2 0 1 5

This work has been partially supported by the European Social Fund within the project "Support for Doctoral Studies at the University of Latvia."

BRIEF TABLE OF CONTENTS

TABLE OF TABLES VIII

TABLE OF FIGURES X

TABLE OF ALGORITHMS XII

TABLE OF ABBREVIATIONS XIII

INTRODUCTION 1

1 EUROPEAN LABOUR–EDUCATION MARKET SYSTEM 8

2 LITERATURE ON LABOUR–EDUCATION MARKET SYSTEM 28

3 FACTORS AFFECTING OVEREDUCATION 58

4 EFFECTS OF OVEREDUCATION 102

5 SOCIAL NETWORKS IN LABOUR–EDUCATION MARKET SYSTEM: AGENT-BASED MODELLING 155

CONCLUSIONS AND RECOMMENDATIONS 211

WORKS CITED 216

APPENDIX A TRANSFORMATION OF EDUCATION VARIABLES IN THE ESS DATABASE 238

APPENDIX B ADDITIONAL DESCRIPTIVE STATISTICS FOR MISMATCH 240

APPENDIX C DISCRETE CHOICE MODELLING METHODOLOGY 243

APPENDIX D AGENT-BASED MODELLING METHODOLOGY 255

APPENDIX E COMPUTATIONS 266

TABLE OF CONTENTS

TABLE OF TABLES VIII

TABLE OF FIGURES X

TABLE OF ALGORITHMS XII

TABLE OF ABBREVIATIONS XIII

INTRODUCTION 1

AIMS AND HYPOTHESES 2
THEORETICAL AND METHODOLOGICAL BASIS 3
NOVELTY 3
APPROBATION OF RESULTS 4
STRUCTURE OF THE DISSERTATION 6

1 EUROPEAN LABOUR–EDUCATION MARKET SYSTEM 8

1.1 EUROPEAN HIGHER EDUCATION AREA AND EUROPE 2020 STRATEGY 8
1.2 EMPLOYABILITY OF GRADUATES AND SKILL MISMATCH 12
1.3 ACCESSIBILITY OF HIGHER EDUCATION AND GOVERNMENT FINANCING 20
1.4 MOBILITY OF STUDENTS AND STAFF AND INTERNATIONALISATION 23
1.5 SUMMARY 26

2 LITERATURE ON LABOUR–EDUCATION MARKET SYSTEM 28

2.1 OVEREDUCATION: EXISTENCE, EFFECTS, MEASURES AND INFLUENCING FACTORS 28
2.1.1 THEORY BEHIND MISMATCH: IS OVEREDUCATION REAL? 28
2.1.1.1 Theories Assuming the Labour Market Is Perfect 29
2.1.1.2 Theories Assuming the Labour Market Is Imperfect 30
2.1.2 EFFECTS OF OVEREDUCATION 31
2.1.3 MEASURES OF OVEREDUCATION 34
2.1.4 FACTORS INFLUENCING OVEREDUCATION 34
2.1.4.1 Ability, Academic Performance and Personality 34
2.1.4.2 Sex and Age 36
2.1.4.3 Immigrant Background 37
2.1.4.4 Job and Labour Market Characteristics 38
2.1.4.5 Characteristics of Individual's Education 39
2.2 SOCIAL NETWORKS: ROLE IN THE LABOUR–EDUCATION MARKET SYSTEM 40
2.2.1 A SHORT NOTE ON THE TERMINOLOGY OF GRAPH THEORY 41
2.2.2 CREATING, MAINTAINING AND DISSOLVING SOCIAL TIES 41
2.2.3 REFERRAL HIRING AND INFORMATION DIFFUSION 44

2.2.4 Choices Related to Higher Education 45
2.2.4.1 Personality, Ability and Other Individual Aspects 46
2.2.4.2 Social Environment: Family, Social Class and Social Networks 49
2.2.4.3 Monetary Costs and Benefits 52
2.3 Mechanisms for Finding and Quitting Job in Existing Agent-Based Models of the Labour Market . 53
2.3.1 Choice among Vacancies 53
2.3.2 Choice among Applicants 54
2.3.3 Decision to Quit 56
2.4 Summary 56

3 FACTORS AFFECTING OVEREDUCATION 58

3.1 European Social Survey Data 58
3.2 ISCO-Based Measure of Overeducation: Definition, Benefits and Drawbacks 59
3.3 Unobserved Heterogeneity in Overeducation Models: Personality vs. Ability 61
3.3.1 Data and Methods 62
3.3.2 Results 65
3.3.3 Conclusions 71
3.4 General Model of Overeducation Determinants 72
3.4.1 Data and Methods 72
3.4.2 Explanatory Variables 73
3.4.2.1 Basic Variables 73
3.4.2.2 Personality Variables 75
3.4.2.3 Macro-Level Variables 75
3.4.3 Results 76
3.4.3.1 Effects from Individual Characteristics 76
3.4.3.2 Effects from Family Characteristics 78
3.4.3.3 Effects from Labour Market Experience 78
3.4.3.4 Effects from Macro-Level Variables 83
3.4.4 Conclusions 83
3.5 Field of Study Effects for Tertiary Graduates 84
3.5.1 Data and Methods 84
3.5.2 Results 85
3.5.3 Conclusions 91
3.6 Industry Effects 91
3.6.1 Data and Methods 91
3.6.2 Results 93
3.6.2.1 Average Overeducation Rates in Industries 93
3.6.2.2 General Models 95
3.6.2.3 Industry-Specific Models 96
3.6.3 Conclusions 100
3.7 Summary 100

4 EFFECTS OF OVEREDUCATION 102

4.1 Job Satisfaction Determinants of Tertiary-Educated Employees in European Countries 102

4.1.1 LITERATURE ON THE CAUSES AND EFFECTS OF JOB SATISFACTION 103
4.1.2 DATA AND METHODS 104
4.1.3 RESULTS 107
4.1.4 DISCUSSION 115
4.1.5 CONCLUSIONS 116
4.2 OVEREDUCATION AND THE PROPENSITY TO QUIT: DOES VOCATIONAL EDUCATION HELP? 116
4.2.1 EDUCATION STATISTICS IN THE DATA 117
4.2.2 DIRECT EFFECTS ON THE PROPENSITY TO QUIT 121
4.2.2.1 Methods 121
4.2.2.2 Results 123
4.2.3 EFFECTS ON OVEREDUCATION, OVERSKILLING AND WORKING IN AN EVENING OR WEEKEND JOB 128
4.2.3.1 Methods 128
4.2.3.2 Results 129
4.2.4 DISCUSSION AND CONCLUSIONS 130
4.3 CHOICES RELATED TO GETTING A DOCTORAL DEGREE IN THE BALTICS 131
4.3.1 DOCTORAL EDUCATION IN THE BALTIC COUNTRIES AT A GLANCE 133
4.3.2 DATA 137
4.3.3 RESULTS 138
4.3.3.1 Goals Pursued when Enrolling in Doctoral Studies 138
4.3.3.2 Reasons of Pursuing Particular Goals: Econometric Analysis 144
4.3.3.3 Choosing the Higher Education Institution for Doctoral Studies 148
4.3.4 DISCUSSION AND CONCLUSIONS 151
4.4 SUMMARY 152

5 SOCIAL NETWORKS IN LABOUR–EDUCATION MARKET SYSTEM: AGENT-BASED MODELLING 155

5.1 JOB SATISFACTION AS A UNIFIED MECHANISM FOR AGENT BEHAVIOUR IN A LABOUR MARKET WITH REFERRAL HIRING 155
5.1.1 JOB SATISFACTION MECHANISM 157
5.1.2 MODEL SPECIFICATION 159
5.1.2.1 Job Search 160
5.1.2.2 Dynamics inside Firms 162
5.1.2.3 The Birth and Death of Firms 163
5.1.2.4 Dynamics of Social Networks 164
5.1.3 PARAMETRISATION 165
5.1.4 RESULTS 168
5.1.4.1 Firm-Level Statistics 169
5.1.4.2 Labour-Market Status of Persons 171
5.1.4.3 Wage Distribution 173
5.1.5 DISCUSSION AND CONCLUSIONS 175
5.2 FIELD OF STUDY CHOICE AND MIGRATION: EFFECTS OF SOCIAL NETWORKS 176
5.2.1 MODEL SETUP 177
5.2.1.1 Agents' Characteristics 177
5.2.1.2 Labour Market Dynamics 178
5.2.1.3 Choice of Field of Study 181
5.2.1.4 Social Networks 181

5.2.2 PARAMETRISATION 182
5.2.3 RESULTS 185
5.2.4 CONCLUSIONS 188
5.3 THE EFFECTIVENESS OF ACCESS RESTRICTION TO HIGHER EDUCATION IN DECREASING OVEREDUCATION 188
5.3.1 MODEL SPECIFICATION 189
5.3.1.1 Agents, Jobs and Overeducation 189
5.3.1.2 Social Networks and Higher Education 190
5.3.1.3 Labour Market Dynamics 191
5.3.1.4 Government Policies: OEM and REM 192
5.3.2 ANALYSIS OF REM WITHOUT SOCIAL NETWORKS 193
5.3.3 PARAMETRISATION AND CALIBRATION 194
5.3.3.1 Setting Common Parameters 194
5.3.3.2 Setting Labour Market Dynamics Parameters 195
5.3.3.3 Calibrating the Remaining Parameters 199
5.3.4 RESULTS 199
5.3.5 SENSITIVITY ANALYSIS 203
5.3.5.1 Re-Calibrating to the Cases of France, Spain and the UK 203
5.3.5.2 Varying Restriction Size 206
5.3.6 DISCUSSION AND CONCLUSIONS 207
5.4 SUMMARY 210

CONCLUSIONS AND RECOMMENDATIONS 211

WORKS CITED 216

APPENDIX A TRANSFORMATION OF EDUCATION VARIABLES IN THE ESS DATABASE 238

APPENDIX B ADDITIONAL DESCRIPTIVE STATISTICS FOR MISMATCH 240

APPENDIX C DISCRETE CHOICE MODELLING METHODOLOGY 243

C.1 MAXIMUM LIKELIHOOD ESTIMATION 243
C.2 LOGISTIC REGRESSION 244
C.2.1 ESTIMATION 244
C.2.2 REPORTING 244
C.3 FIXED-EFFECTS LOGISTIC REGRESSION 245
C.3.1 ESTIMATION 246
C.3.2 COMPUTATION TIME 246
C.4 MIXED-EFFECTS LOGISTIC REGRESSION 246
C.4.1 ESTIMATION 247
C.4.2 SPECIFYING THE VARIANCE–COVARIANCE STRUCTURE OF RANDOM EFFECTS 249
C.4.3 COMPUTATION TIME 249
C.4.4 REPORTING 250
C.5 ORDERED DISCRETE CHOICE WITH FEW LEVELS: (GENERALISED) ORDERED LOGIT 250
C.5.1 ESTIMATION OF THE PROPORTIONAL-ODDS ORDERED LOGIT MODEL 251

C.5.2 REPORTING OF THE PROPORTIONAL-ODDS ORDERED LOGIT MODEL 251
C.6 ORDERED DISCRETE CHOICE WITH MANY LEVELS: CENSORED LINEAR REGRESSION 251
C.6.1 ESTIMATION 251
C.7 VARIANCE ESTIMATION FOR CLUSTERED SAMPLES 252
C.8 TYPICAL GOODNESS-OF-FIT STATISTICS FOR LOGIT-TYPE REGRESSIONS 253

APPENDIX D AGENT-BASED MODELLING METHODOLOGY 255

D.1 WHAT IS AGENT-BASED MODELLING 255
D.2 HOW TO DO AGENT-BASED MODELLING 256
D.2.1 FROM GOAL TO IMPLEMENTATION 256
D.2.2 CALIBRATION: SETTING PARAMETERS 257
D.2.3 VALIDATION: CHECKING THE MODEL 257
D.2.4 RUNNING THE MODEL 259
D.2.5 ANALYSING OUTPUT AND ITS SENSITIVITY 259
D.3 THE MOTIVATION BEHIND USING AGENT-BASED MODELLING 259
D.3.1 COMPLEX ADAPTIVE SYSTEMS 259
D.3.2 TRADITIONAL ECONOMIC MODELLING VS. AGENT-BASED MODELLING 261
D.3.3 AGENT-BASED MODELLING VS. OTHER INDIVIDUAL-LEVEL SIMULATION TECHNIQUES 261
D.3.4 AGENT-BASED MODELLING VS. SYSTEM DYNAMICS 264
D.4 THE PROBLEMS OF AGENT-BASED MODELLING 264

APPENDIX E COMPUTATIONS 266

TABLE OF TABLES

TABLE 1-1 – NUMBER OF TERTIARY GRADUATES IN EUROPE, 2001–2009, BY FIELD OF STUDY (THOUSANDS) 13
TABLE 1-2 – ANNUAL GROWTH IN THE NUMBER OF TERTIARY GRADUATES IN EUROPE, 2001–2009, BY FIELD OF STUDY 13
TABLE 1-3 – FINDINGS ON THE INCIDENCE OF OVEREDUCATION IN EUROPEAN COUNTRIES, PER CENT 14
TABLE 1-4 – FINDINGS ON THE INCIDENCE OF UNDEREDUCATION IN EUROPEAN COUNTRIES, PER CENT 15
TABLE 2-1 – MEASURES OF OVEREDUCATION 35
TABLE 2-2 – DESCRIPTION OF HOLLAND'S PERSONALITY TYPES AND RELATED FIELDS OF STUDY.. 47
TABLE 3-1 – COUNTRY COVERAGE IN ESS DATA, BY ROUND 59
TABLE 3-2 – CORRESPONDENCE BETWEEN ISCO MAJOR GROUPS AND EDUCATION LEVELS.. 60
TABLE 3-3 – PERSONALITY VARIABLES IN ESS DATA 64
TABLE 3-4 – ODDS RATIOS AFTER LOGIT, RESPONDENT'S INCOME IS AT LEAST HALF OF HOUSEHOLD INCOME 66
TABLE 3-5 – ODDS RATIOS AFTER LOGIT, RESPONDENT'S INCOME IS A VERY LARGE FRACTION OF HOUSEHOLD INCOME.................... 67
TABLE 3-6 – QUALITY OF FIT OF LOGIT REGRESSIONS IN COUNTRY GROUPS, RESPONDENT'S INCOME IS AT LEAST HALF OF HOUSEHOLD INCOME.. 68
TABLE 3-7 – QUALITY OF FIT OF LOGIT REGRESSIONS IN COUNTRY GROUPS, RESPONDENT'S INCOME IS A VERY LARGE FRACTION OF HOUSEHOLD INCOME 69
TABLE 3-8 – GROUPING OF INITIAL PERSONALITY VARIABLES INTO COMPOSITE PERSONALITY VARIABLES 75
TABLE 3-9 – ODDS RATIOS AFTER MIXED EFFECTS LOGISTIC REGRESSION OF OVEREDUCATION 80
TABLE 3-10 – MERGING THE ORIGINAL FIELDS OF STUDY FROM ESS INTO BROAD FIELDS 84
TABLE 3-11 – UNCONDITIONAL DISTRIBUTION OF THE INCIDENCE OF OVEREDUCATION OVER FIELDS OF STUDY, TERTIARY GRADUATES AGED 20–40 86
TABLE 3-12 – ODDS RATIOS AFTER DIFFERENT LOGISTIC REGRESSIONS OF OVEREDUCATION FOR EUROPE 87
TABLE 3-13 – ODDS RATIOS OF FIELD-OF-STUDY EFFECTS AFTER MIXED EFFECTS LOGIT REGRESSIONS OF OVEREDUCATION FOR COUNTRY GROUPS 89
TABLE 3-14 – PREDICTED PROBABILITIES OF OVEREDUCATION OVER FIELDS OF STUDY AFTER MIXED EFFECTS LOGISTIC MODEL, TERTIARY GRADUATES AGED 20–40... 90
TABLE 3-15 – ODDS RATIOS AFTER MIXED EFFECTS LOGIT OF OVEREDUCATION, ALL INDUSTRIES 98
TABLE 3-16 – ODDS RATIOS AFTER MIXED EFFECTS LOGIT OF OVEREDUCATION, BY INDUSTRY 99
TABLE 4-1 – JOB-RELATED FACTORS AFFECTING JOB SATISFACTION FROM ESS ROUND 5 105
TABLE 4-2 – MARGINAL EFFECTS AFTER TOBIT REGRESSIONS OF JOB SATISFACTION: EDUCATION LEVEL EFFECTS 108
TABLE 4-3 – MARGINAL EFFECTS AFTER TOBIT REGRESSIONS OF THE JOB SATISFACTION OF TERTIARY EDUCATED EMPLOYEES AGED 20–60, BY COUNTRY GROUP AND EDUCATION LEVEL... 109
TABLE 4-4 – RANKING OF THE FIVE MOST IMPORTANT JOB-RELATED EFFECTS ON JOB SATISFACTION, BY COUNTRY GROUP AND EDUCATION LEVEL 113
TABLE 4-5 – DESCRIPTIVE STATISTICS ON EDUCATION LEVEL AND TYPE 118
TABLE 4-6 – SAMPLE AVERAGES OF LABOUR MARKET VARIABLES BY EDUCATION LEVEL AND TYPE... 119
TABLE 4-7 – NUMBER OF OVEREDUCATED RESPONDENTS IN EACH OCCUPATION 121
TABLE 4-8 – ODDS RATIOS AFTER RANDOM EFFECTS LOGIT REGRESSIONS OF A HIGH LIKELINESS OF NOT WORKING WITH THE CURRENT EMPLOYER IN 6 MONTHS 124
TABLE 4-9 – COMPARISON OF OVEREDUCATION EFFECTS: AVERAGE EFFECT VS. EFFECT BY TYPE OF EDUCATION............................ 125
TABLE 4-10 – COMPARISON OF TYPE OF MISMATCH EFFECTS: OVEREDUCATION VS. OVERSKILLING IN 2010 125
TABLE 4-11 – MARGINAL EFFECTS AFTER MULTIVARIATE PROBIT MODELS OF OVEREDUCATION, OVERSKILLING AND EVENING OR WEEKEND JOB 129
TABLE 4-12 – ODDS RATIOS AFTER COUNTRY-SPECIFIC LOGISTIC REGRESSIONS OF HAVING LABOUR-MARKET PRIMARY GOALS......... 145
TABLE 4-13 – ODDS RATIOS AFTER SEX-SPECIFIC GENERALISED ORDERED LOGISTIC REGRESSIONS OF THE TYPE OF GOAL: CONSTRAINED PART OF THE MODELS 146
TABLE 4-14 – ODDS RATIOS AFTER SEX-SPECIFIC GENERALISED ORDERED LOGISTIC REGRESSIONS OF THE TYPE OF GOAL: UNCONSTRAINED PART OF THE MODELS 147
TABLE 5-1 – DISTRIBUTION OF CAREER OPPORTUNITIES 165

Table 5-2 – Distribution of Job Variety 165
Table 5-3 – Empirical Data on Wage Returns on Year of Experience 167
Table 5-4 – Distribution of Annual Workforce Change 167
Table 5-5 – Parameter Values 168
Table 5-6 – Results of Kolmogorov–Smirnov Test on the Equality of Firm Size Distribution 170
Table 5-7 – Number of Firms at End-of-Year 260 170
Table 5-8 – Statistics on Unemployment Duration 173
Table 5-9 – Statistics on On-the-Job Search Duration 173
Table 5-10 – Median Wages by Job Type and Firm Size at End-of-Year 260 174
Table 5-11 – $\mathcal{F}$-characteristics and Their Mappings to the Fields of Study Used in the Model 183
Table 5-12 – The Distribution of the Levels of $\mathcal{F}$-characteristics in the Population 183
Table 5-13 – Wage Intervals of Professions 183
Table 5-14 – $\mathcal{H}$-characteristics Used in the Model 184
Table 5-15 – Other Parameters 184
Table 5-16 – Statistics of Employees with High-Quality Match with Their Current Profession 186
Table 5-17 – Overeducation Rates by Field of Study 186
Table 5-18 – Marginal Effects of Social Network Characteristics on Choosing Fields of Study after Multinomial Logistic Regression 187
Table 5-19 – Share of Emigrants from Country by Education 187
Table 5-20 – Agent–Job Correspondence 190
Table 5-21 – Common Parameter Values 195
Table 5-22 – Distribution of the Length of the Longest Unemployment Spell Experienced by the Respondent 197
Table 5-23 – Labour Market Parameters 197
Table 5-24 – Calibrated Parameter Values 199
Table 5-25 – REM Reform Efficiency 201
Table 5-26 – Labour Market Parameters for Spain, France and the United Kingdom 203
Table 5-27 – Calibrated Parameters for Europe, Spain, France and the United Kingdom 203
Table 5-28 – REM Reform Efficiency in Europe, France, Spain and the United Kingdom 206
Table 5-29 – Average Popularity of Higher Education in the Social Network of Agent by Its Level of Education, by Country 206
Table 5-30 – REM Reform Efficiency by Restriction Size 206
Table A-1 – Shares of Primary, Secondary and Tertiary Graduates in the ESS Database 239
Table B-1 – Countries with Extreme Mismatch Incidence for Non-natives, Disabled and Living outside Big Cities, Total Sample 240
Table B-2 – Countries with Extreme Mismatch Incidence for Non-natives, Disabled and Living outside Big Cities, Youth Subsample (15–29) 241
Table B-3 – Countries with Extreme Mismatch Incidence for Non-natives, Disabled and Living outside Big Cities, Mature Subsample (30+) 242
Table C-1 – Structures of the Random Effects Variance Matrix in Stata 249
Table D-1 – Comparison of Approaches to Building Simulations 258
Table D-2 – Definitions of a Complex Adaptive System 260
Table D-3 – Comparison of Traditional Economic Modelling with Agent-Based Modelling 262

TABLE OF FIGURES

Figure 1-1 – History of the Bologna Process ... 9
Figure 1-2 – Relationships between EHEA and Europe 2020 Strategies ... 11
Figure 1-3 – Dynamics of Average Incidence of Mismatch: Young (15–29) vs. Mature (30+) ... 17
Figure 1-4 – Dynamics of Average Incidence of Mismatch: Males vs. Females ... 18
Figure 1-5 – Differences between the Overeducation Incidence of Immigrants and Fully Natives ... 19
Figure 1-6 – Proportion of Students who Pay Tuition Fees for Higher Education, 2011–2012 ... 22
Figure 1-7 – Annual Expenditure on Public Higher Education Institutions per Student in EUR PPS in 2010 ... 24
Figure 1-8 – Annual Expenditure on Public Higher Education Institutions per Student as % of GDP per Capita in 2010 . 24
Figure 1-9 – Balance of International Student Flows in 2010 ... 25
Figure 1-10 – Structural View of the Remaining Chapters ... 27
Figure 2-1 – Example of Undirected and Directed Graphs and Their Main Statistics ... 41
Figure 2-2 – Holland's Personality Types ... 48
Figure 3-1 – Relationship between Tertiary Attainment and ISCO-based Overeducation ... 60
Figure 3-2 - Odds Ratios of Age Groups after Mixed Effects Logit Regression of Overeducation ... 77
Figure 3-3 – Odds Ratios of Firm Size after Mixed Effects Logit Regression of Overeducation ... 79
Figure 3-4 – Industry Positions on Overeducation, Averaged over Countries ... 93
Figure 3-5 – Country-Specific Industry Positions on Overeducation vs. Share of ISCO 4–9: Countries with Two Distinct Industry Groups ... 94
Figure 3-6 – Country-Specific Industry Positions on Overeducation vs. Share of ISCO 4–9: Countries with Continuous Positive Relationship ... 94
Figure 3-7 – Country-Specific Industry Positions on Overeducation vs. Share of ISCO 4–9: Countries with No Relationship ... 95
Figure 3-8 – Odds Ratios of Overeducation for Different Industries Relative to Manufacturing, Without and With Macro Variables ... 96
Figure 3-9 – Odds Ratios of Overeducation for Different Industries Relative to Manufacturing, by Sex ... 97
Figure 4-1 – Occupational Structure in the Sample by Educational Track ... 119
Figure 4-2 – Demand for Doctoral Studies in the Baltic Countries, 1998–2012 ... 134
Figure 4-3 – Population Pyramid in the Baltic Countries, 1 Jan 2014 ... 135
Figure 4-4 – Popularity of Fields of Study at Doctoral Level in the Baltic Countries, 1998 vs. 2012 ... 136
Figure 4-5 – Goals Pursued when Enrolling in Doctoral Studies, by Country ... 139
Figure 4-6 – Definition of Goal Types ... 140
Figure 4-7 – Goals Pursued when Enrolling in Doctoral Studies, by Country and Field of Study ... 143
Figure 4-8 – Most Important Factors when Choosing an HEI for Doctoral Studies, by Country and Goal Type ... 148
Figure 4-9 – Most Important Information Sources about the HEI for Doctoral Studies, by Country and Goal Type ... 150
Figure 5-1 – Example of Creating of a Consideration List from the Awareness List ... 161
Figure 5-2 – Distribution of Firm Size at End-of-Year 260 Depending on the Composition of Job Satisfaction ... 169
Figure 5-3 – Average Job Satisfaction at End-of-Year 260 ... 170
Figure 5-4 – Average Number of Friends at End-of-Year 260 ... 171
Figure 5-5 – Beveridge Curve at End of Years 251–260 ... 172
Figure 5-6 – Unemployment, Entrepreneurship and Agents Involved in On-the-Job Search for at Least One Month as Percentage of All Agents ... 172
Figure 5-7 – Histograms of Wages at End-of-Year 260 ... 174
Figure 5-8 – Determination of Employment Possibilities and Salary in the Skilled Labour Market ... 179
Figure 5-9 – The Process of Finding a Job for University Graduates ... 180
Figure 5-10 – Demand for Fields of Study in Country 1 ... 185
Figure 5-11 – Selected Examples of Emigration Patterns ... 186
Figure 5-12 – Overeducation Rate Dynamics ... 200
Figure 5-13 – Agents Not Admitted to the University from Fresh Graduates Desiring to Continue Studies ... 201

Figure 5-14 – Average Share of Students and Tertiary Educated Agents among Agent's Friends 202
Figure 5-15 - Distribution of Overeducation Rate over Runs 45 Years after Restricting Access to Higher Education, by Country 205

TABLE OF ALGORITHMS

ALGORITHM 5-1 – MONTHLY ACTIONS IN THE LABOUR MARKET 159
ALGORITHM 5-2 – CHOOSING THE FIELD OF STUDY 181

TABLE OF ABBREVIATIONS

Abbreviation	Meaning
ABM	Agent-based modelling
AGQ	Adaptive Gaussian quadrature
AI	Artificial intelligence
AIC	Akaike information criterion
BA	Base model with ability variable
BIC	Schwarz's Bayesian information criterion
BP	Base model with personality variable
BPM	BP with macro-level variables
BPM-PRE	BPM with personality random effects
CAS	Complex adaptive system
cdf	Cumulative distribution function
CEE	Central and Eastern Europe
EEA	European Economic Area
EHEA	European Higher Education Area
ERA	European Research Area
ESS	European Social Survey
EU	European Union
FOS	Field of study
FSU	Former Soviet Union
GDP	Gross domestic product
GOT	General Occupational Themes
HEI	Higher education institution
HQ	High-qualified
HQM	High-quality match
ICT	Information and telecommunication technology
ILO	International Labour Organisation
ISCO	International Standard Classification of Occupations
ISCED	International Standard Classification of Education
JS	Job satisfaction
LAA	Latin America, Africa and Asia
LEMS	Labour–education market system
LQ	Low-qualified
LR	Likelihood ratio
LTU	Long-term unemployment
ML	Maximum likelihood
MSM	Microsimulation model
MTU	Medium-term unemployment
OECD	Organisation for Economic Cooperation and Development
OEM	Open education market
OLS	Ordinary least squares
ORU	Over-, required, undereducation
PPS	Purchasing power standard
R&D	Research and development
REM	Restricted education market
ROC	Receiver operating characteristic
ROC AUC	ROC area under curve
SES	Socio-economic status
SD	System dynamics
Std. dev.	Standard deviation
UK	United Kingdom
US	United States of America

INTRODUCTION

In the latest decades, the EU has been actively engaging into the expansion of higher education. The proportion of individuals with tertiary education in the EU rose from 17 per cent in 2000 to 25 per cent in 2012 (Eurostat data). On the one hand, education expansion is a positive development – as reflected, for instance, by prevalent sizeable differences in the unemployment risk of individuals with primary, secondary and tertiary education levels. The unemployment rate of the primary-educated relative to that of the tertiary-educated increased from 2.4 times in 2000 to 3.0 times in 2013, while the relative rate for the secondary-educated dropped from around 2.0 in 2000 to 1.5 in 2013 (Eurostat data). These numbers clearly show that tertiary graduates are better able to find jobs in the labour market than individuals with lower levels of education are. This, however, does not mean that the jobs they find are appropriate.

Indeed, there is a worrying tendency of growing mismatch in European labour markets, which has risen in policy agendas and attracted major attention from such organisations as Cedefop (2012), European Commission (DG Employment, Social Affairs and Inclusion, 2012), International Labour Organisation (2013a), OECD (Quintini, 2011) and World Economic Forum (2014). All of them express concern with current situation and trends, among which is the trend of growing overeducation. This means that individuals increasingly more often work in jobs that require lower education levels than they have. In other words, the expansion of higher education came together with dropping quality of match between individuals' education and job requirements. For that reason, the current generation of young people, who typically are "graduates from expensive and outmoded education systems graduating with high debts and mismatched skills" (World Economic Forum, 2014, p. 10), was recently referred to as *generation at risk* (International Labour Organisation, 2013a). At the same time, the International Labour Organisation (2013a) stresses that official statistical databases do not include comparable data on skills and competencies, which hampers the ability to analyse the quality of match between graduates and jobs.

Various imperfections in the labour market give rise to mismatch, and these are well documented at theoretical and empirical levels. However, the reasons of rising overeducation lie not only in the labour market, but also in the education market. In particular, the expansion of higher education was partly allowed by the reduction in academic and social selection into universities (Convert, 2005), which may have had a serious impact on the distribution of the quality of graduates.

Overeducation, thus, appears to stem from the deficiencies of the link between the labour market and the education market, coupled with well-known rigidities and other imperfections of these markets. It is then necessary to consider both markets together as a single system, and in this dissertation, these will be referred to as ***labour–education market system*** (LEMS). The mentioned link works both ways. The workforce available in the labour market consists of the output of the education market. At the same time, individuals base their decisions on education partly on their perspectives in the labour market after graduation. Both directions of the link between the markets will be considered in this dissertation.

The decisions individuals make in both markets, as their other decisions, are typically not based on the detailed analysis of full information. Referral hiring is one of the well-known empirical examples of that. This is a practice of making hiring decisions largely depending on social networks, which are increasingly prevalent nowadays and affecting many of the decisions of individuals. Social networks are used not only as a source of information, but also as a guide to use that information. Inquiring the opinions of known people (e.g., friends, acquaintances or family members) is relatively fast and, because of trust to the source of information, appears valid to the inquirer. The flip side is that the decision may be far from what could be theoretically expected based on a fully rational view of the individual.

Conventional econometric methods, while applicable in case of studying the determinants and effects of overeducation, are of limited use in studying how social networks distort LEMS-related decisions of individuals. Agent-based modelling (ABM) is an attractive alternative in this case. While several published papers attempt to model labour market dynamics, albeit with rather restrictive assumptions (but still more realistic than those in mathematical models), I am unaware of any published study using ABM to model LEMS with an elaborate mechanism of systematic influence of social networks on individual decision-making. This dissertation will contribute to filling this gap.

Aims and Hypotheses

The aim of this dissertation is to **develop conventional econometric and agent-based models of individual behaviour in European labour–education market system with a focus on overeducation and the influence of social networks.**

To reach this aim, I set the following tasks:

- Analyse the current state and latest developments in European LEMS, with special emphasis on overeducation
- Analyse the existing research on LEMS
- Analyse the existing research on agent-based modelling
- Analyse the determinants of overeducation
- Analyse the labour-market and educational outcomes influenced by overeducation
- Develop agent-based models of LEMS with elaborate individual decision-making mechanisms influenced by social networks

I put forward the following hypotheses:

- Personality is at least as important as ability in influencing the labour-market position of individuals, and in particular, their exposure to overeducation
- Agent-based models with enhanced social networking component in agents' decision-making is a good alternative to conventional mathematical methods in modelling LEMS

Theoretical and Methodological Basis

The dissertation is based on world theoretical and empirical literature on the issues related to labour market, education market, individual decision-making, social networks and modelling methodologies.

The main methods used in the dissertation are conventional econometric methods (logistic regression, (generalised) ordered logistic regression, tobit regression, fixed- and mixed-effects logistic regression) and agent-based modelling.

The empirical part of the dissertation mainly uses European Social Survey data, but other data sources, such as TREE for Switzerland and my own survey among doctoral students in the Baltics, are used in specific sections. In background and theoretical parts, numerous other sources are also used, such as Eurostat, various legal documents and periodic literature.

Novelty

The dissertation contributes to the existing literature by:

- Developing an income-based measure of ability and a three-factor measure of personality traits and comparing their relative power in explaining overeducation risk

- Developing conventional econometric models to analyse the general determinants of overeducation in European countries and the less-studied effects from field of study and industry on overeducation risk
- Developing conventional econometric models to analyse (1) the effects of overeducation on the decisions and position of individuals in LEMS as part of a more general analysis of the determinants of job satisfaction for tertiary graduates, (2) the role of vocational education in mitigating the effects from overeducation on the propensity to quit the job (case of Switzerland) and (3) the motivation of graduates to continue studies at doctoral level and the process of choosing a higher education institution for doctoral studies (case of the Baltic States)
- Developing an agent-based model of LEMS with an elaborate mechanism of job satisfaction depending on social networks, which guides agent behaviour during the application for vacancies, selection among different job offers and quitting the job
- Developing an agent-based model of LEMS with (1) the choice of field of study depending on the interactions within social networks, (2) the model of the quality of match between the agent and the job and (3) a possibility to emigrate
- Developing an agent-based model of LEMS with the choice to continue studies at university level depending on the interactions within social networks and using the model to compare the efficiency of restricting access to higher education with the results of *laissez-faire* as possible policy responses to rising overeducation

Approbation of Results

The results of this dissertation appeared in 13 publications, of which there is one book, 7 journal articles and three book chapters. In addition, they were presented in 12 international and four local conferences. The dissertation contains elements based on the results of a project with the International Labour Office.

Publications in books:

- Tarvid, A. (2015). *Agent-Based Modelling of Social Networks in Labour–Education Market System.* New York: Springer.

Publications in journals:

- Tarvid, A. (2016, forthcoming). Job Satisfaction as a Unified Mechanism for Agent Behaviour in a Labour Market with Referral Hiring. *International Journal of Computational Economics and Econometrics*.

- Tarvid, A. (2015). The Role of Industry in the Prevalence of Overeducation in Europe. *Procedia Economics and Finance, 30*, 876–884*.* doi:10.1016/S2212-5671(15)01337-4
- Tarvid, A. (2015). Job Satisfaction Determinants of Tertiary Graduates in Europe. *Procedia Economics and Finance, 24*, 683–691, doi:10.1016/S2212-5671(15)00674-7
- Tarvid, A. (2015). The Effectiveness of Access Restriction to Higher Education in Decreasing Overeducation. *Economic Analysis and Policy, 45,* 11–26. doi:10.1016/j.eap.2014.12.003
- Tarvid, A. (2014). PhD Degree in Latvia: Individual's Internal Need or Demand from Labour Market? *Humanities and Social Sciences Latvia, 22*(2), 109–121.
- Tarvid, A. (2014). Motivation to Study for PhD Degree: Case of Latvia. *Procedia Economics and Finance, 14,* 585–594. doi:10.1016/S2212-5671(14)00747-3
- Tarvid, A. (2013). Unobserved heterogeneity in overeducation models: Is personality more important than ability? *Procedia Economics and Finance, 5*, 731–740. doi:10.1016/S2212-5671(13)00084-1

Publications in book chapters:

- International Labour Organization. (2013). *Global Employment Trends for Youth 2013: A generation at risk*. Geneva: International Labour Office. Retrieved from http://www.ilo.org/wcmsp5/groups/public/---dgreports/---dcomm/documents/publication/wcms_212423.pdf
- Tarvid, A. (2013). Field of Study Choice and Migration: Effects of Social Networks. In A. Koch & P. Mandl (Ed.), *Modeling Social Phenomena in Spatial Context (Geosimulation Series, Vol. 2)* (pp. 125–138)*.* Münster: LIT Publisher.
- Tarvid, A. (2012). Effects of the Field of Higher Education on the Prevalence of Over-education in European Countries. In J. E. Kesner (Ed.), *Education: Evaluation, Reform and Policy* (pp. 167–184). Athens: Athens Institute for Education and Research.

Other publications:

- Tarvid, A. (2012). *Job Satisfaction Determinants of Tertiary-Educated Employees in European Countries*. ATINER's Conference Paper Series, No. ECO2012-0257, Athens Institute for Education and Research. Retrieved from http://www.atiner.gr/papers/ECO2012-0257.pdf
- Tarvid, A. (2011). Job Satisfaction Modelling in Agent-Based Simulations. In A. Bruzzone, M. A. Piera, F. Longo, P. Elfrey, M. Affenzeller, & O. Balci (Ed.), *The 23rd European Modeling & Simulation Symposium* (pp. 158–165). Rende, Italy: DIPTEM Università di Genova.

Presentations in international conferences:

- A. Tarvid. *Job Satisfaction Determinants of Tertiary Graduates in Europe*. International Conference on Applied Economics 2015. Kazan Federal University, Kazan, Russia. July 2015.
- A. Tarvid. *The Role of Industry in the Prevalence of Overeducation in Europe*. 3rd Economics and Finance Conference. IISES, Rome, Italy. April 2015.
- A. Tarvid. *Reasons of Getting a Doctoral Degree in the Baltics: A Cross-Country Comparison.* LEER Workshop on Education Economics. KU Leuven, Leuven, Belgium. April 2015.
- A. Tarvid. *Job Satisfaction as a Unified Mechanism for Agent Behaviour on a Labour Market with Referral Hiring*. 10th Conference of the European Social Simulation Association. Universitat Autonoma de Barcelona, Barcelona, Spain. September 2014.
- A. Tarvid. *Motivation to Study for PhD Degree: Case of Latvia*. International Conference on Applied Economics. Technological Institute of Western Macedonia, Chania, Greece. July 2014.
- A. Tarvid. *Overeducation and the Propensity to Quit: Does Vocational Education Help? Case of Switzerland in 2007–2010*. BIGSSS International Conference 2014: Social Stratification and Social Policy. BIGSSS International Graduate School of Social Sciences, Bremen, Germany. July 2014.

- A. Tarvid. *Battling with Overeducation: Will Restricting Access to Higher Education Work in the Presence of Social Networks?* 9th Conference of the European Social Simulation Association. Warsaw School of Economics, Warsaw, Poland. September 2013.
- A. Tarvid. *Unobserved Heterogeneity in Overeducation Models: Is Personality More Important than Ability?* International Conference on Applied Economics 2013. Bahçeşehir University, Istanbul, Turkey. June 2013.
- A. Tarvid. *Field of Study Choice and Migration: Effects of Social Networks*. 8th Conference of the European Social Simulation Association. University of Salzburg, Salzburg, Austria. September 2012.
- A. Tarvid. *Job Satisfaction Determinants of Tertiary-Educated Employees in European Countries.* 7th Annual International Symposium on Economic Theory, Policy and Applications. Athens Institute for Education and Research, Athens, Greece. July 2012.
- A. Tarvid. *Job Satisfaction Modelling in Agent-Based Simulations*. The 23rd European Modeling & Simulation Symposium. University of Genoa & University of Calabria, Rome, Italy. September 2011.
- A. Tarvid. *Effects of the Field of Higher Education on the Prevalence of Over-education in European Countries*. 6th Annual International Symposium on Economic Theory, Policy and Applications. Athens Institute for Education and Research, Athens, Greece. July 2011.

Presentations in local conferences:

- A. Tarvid. *The Effectiveness of Access Restriction to Higher Education in Decreasing Overeducation*. LU 73rd Conference. The University of Latvia, Riga, Latvia. February 2015.
- A. Tarvid. *Job Satisfaction Determinants of Tertiary-Educated Employees in Europe after the Crisis*. LU 71th Conference. The University of Latvia, Riga, Latvia. January 2013.
- A. Tarvid. *Analysis of the Factors Determining the Job Satisfaction of Tertiary Graduates in European Countries.* LU 70th Conference. The University of Latvia, Riga, Latvia. February 2012.
- A. Tarvid. *Review of Agent-Based Modelling Applications in Economics.* LU 69th Conference. The University of Latvia, Riga, Latvia. February 2011.

Participation in research projects:

- External Collaborator. *Global Employment Trends for Youth 2013*. International Labour Office. Oct 2012 – Apr 2013.

Structure of the Dissertation

The dissertation consists of an introduction, five chapters and conclusions. Its length without appendices is 237 pages. It contains 74 tables, 45 figures and two algorithms.

Chapter 1 sets the context of the dissertation by describing the current situation with the European labour–education market system.

Chapter 2 presents the literature review. It covers the topics of overeducation and social networks in LEMS.

Chapter 3 presents the results on factors affecting overeducation, specifically concentrating on the effects from personality, field of study and industry.

Chapter 4 presents findings on the determinants of the job satisfaction of tertiary graduates, the effects of educational track on the propensity to quit and the motivation to get a doctoral degree as part of the study on the effects from overeducation.

Chapter 5 applies agent-based modelling to creating advanced simulations of an abstract labour–education market system. Simulations are parameterised using data on Europe. Different aspects are studied, including job satisfaction, choice of field of study and policy response to high overeducation.

There are five appendices in the dissertation. Appendix A discusses the transformation of the original variables defining education levels in the European Social Survey data. Appendix B contains tables with additional statistics on overeducation and undereducation in Europe. The following two appendices contain the methodological background for readers who would like to learn more about the methods used in this dissertation. Appendix C discusses the conventional econometric methods used in discrete choice modelling. Appendix D discusses agent-based modelling, including the motivation behind its use, its comparison with standard economic modelling and other individual-level simulation approaches, and its problems. Finally, Appendix E contains the proofs of two propositions used in Section 5.3.

1 EUROPEAN LABOUR–EDUCATION MARKET SYSTEM

The education system and the labour market are closely intertwined. Hence, analysing them separately does not seem to be right. One has to at least recognize and explicitly show the interdependencies of these two systems. This applies not only to theoretical modelling or empirical research, but also to reviewing the governance principles of both systems.

For this reason, this chapter considers these systems as two views of a single labour–education market system (LEMS). It starts at the strategic level, explicitly linking European higher education strategy with European labour market strategy. Then the three priorities proposed for European higher education at the Ministerial Conference in 2012 are reviewed.

1.1 European Higher Education Area and Europe 2020 Strategy

Since 1998, European higher education system has been experiencing a major reform, the Bologna Process (see Figure 1-1 for a description of its main events). Its main goal is to provide (1) “the educational component necessary for the construction of a Europe of knowledge within a broad humanistic vision and in the context of [mass] higher education systems” and (2) “lifelong access to learning that supports the professional and personal objectives of a diversity of learners” (Sursock & Smidt, 2010, p. 9).

In 2010, the Process solemnly ended with the launch of European Higher Education Area (EHEA). This does not mean that the reforms stopped or all initial objectives have been reached. Indeed, while two-tier degree structure has been implemented by more than 95% of HEIs in the participating countries and there is some progress in quality and mobility enhancement (Sursock & Smidt, 2010), there is still a long way to go.

In particular, the Bucharest Ministerial Conference held in 2012 suggested three priorities (Romanian Bologna Secretariat, n.d.):

- Accessibility of higher education
- Employability of graduates
- Mobility of students and staff

On the other hand, the European University Association (Sursock & Smidt, 2010) proposed to focus on

- Framing European higher education strategy within a broad vision of future society

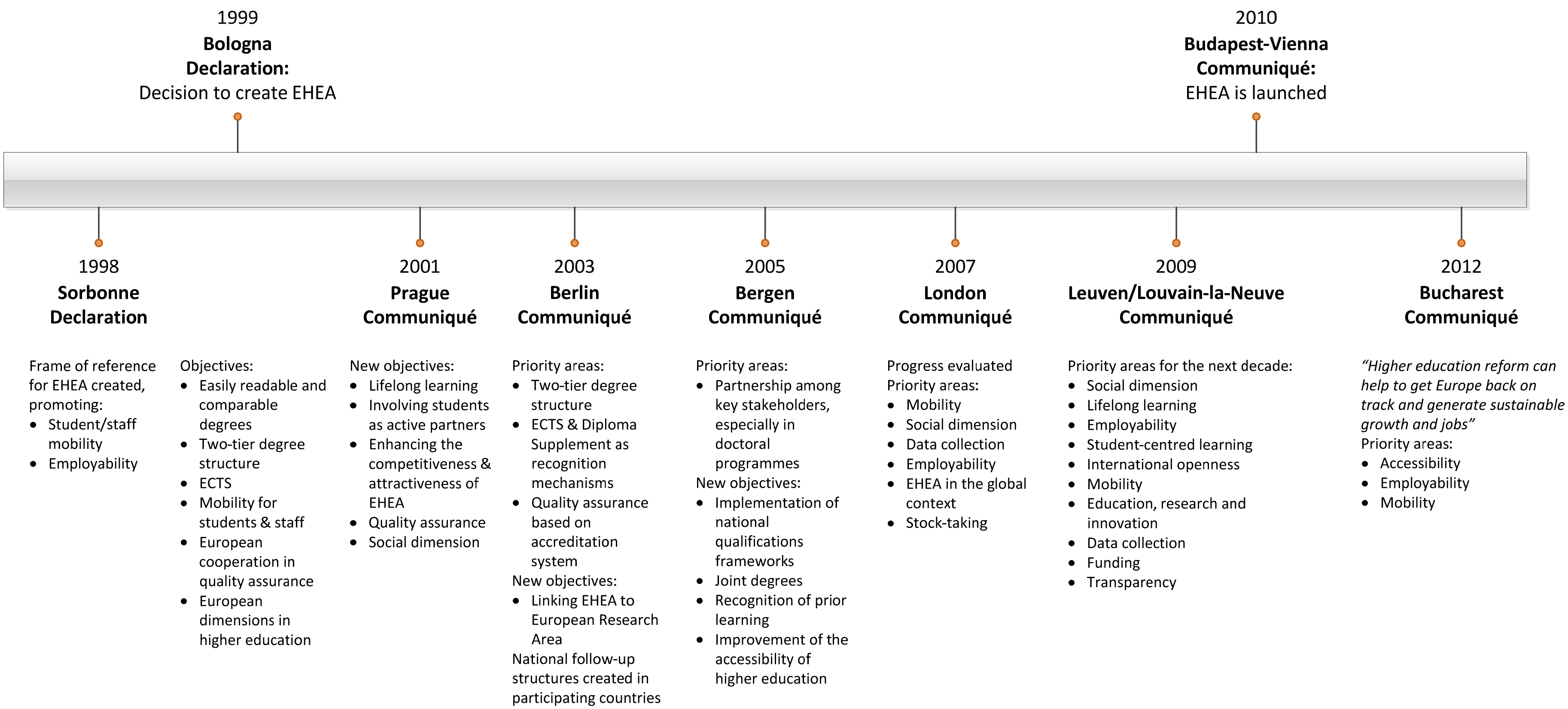

Figure 1-1 – History of the Bologna Process

Source: compiled from Davies (2008), Romanian Bologna Secretariat (n.d.), Veiga, Amaral and Mendes (2008)

- Engaging all stakeholders (including students) in quality assurance
- Enhancing the cooperation between European countries in their internationalisation strategies
- Strengthening the links between EHEA and European Research Area[1] to create a strong interface among education, research and innovation

Several reasons are noted for the implementation of the reform being sluggish. Firstly, there is certain scepticism in the academic community of some countries towards the reform (Garcia-Garcia, Gonzales, & Argüelles, 2009; Veiga, Amaral, & Mendes, 2008), which is a result of poor communication with some of the major stakeholders (Sursock & Smidt, 2010). Secondly, there is too much emphasis on goals measuring form, or state, as opposed to goals measuring substance, or embeddedness (Veiga, Amaral, & Mendes, 2008). Sursock and Smidt mean the same when talking about a "fragmented and instrumental view of education that has not always facilitated understanding in institutions of the important links between the various elements" (2010, p. 9). Thirdly, not enough data crucial in further planning of the educational reform are being collected (Sursock & Smidt, 2010).

To sum up, because the measurable goals of the reform are set at a high level and the motivation behind them and their interrelations are not properly communicated to stakeholders, HEIs tend to implement them in a way that makes them seem progressing towards these high-level goals but does not require implementing substantial internal changes. Hence, it leads to fulfilling the formal goals but not the vision of the reform itself. Moreover, the lack of detailed data that should have been collected by the responsible authorities, such as Eurostat or statistics offices of the participating countries, hampers the further planning of the reform and the assessment of its progress.

There are also warnings that while common learning outcomes (in the sense that students graduating similar study programmes in different countries end up acquiring similar knowledge and skills) are desirable, it is necessary to allow a considerable variation in the nature and form of degree programmes and the learning experience at the tertiary level across countries, unless

---

[1] The main aim of the European Research Area (ERA) is to maximise the return on investment in research by increasing the efficiency and effectiveness of European public research system. The ERA reform has five priorities:

- More effective national research systems
- Optimal transnational co-operation and competition
- An open labour market for researchers
- Gender equality and gender mainstreaming in research
- Optimal circulation and transfer of scientific knowledge

The benefits from ERA come through improved researchers' mobility, European collaboration in R&D and improvements in research infrastructure. (European Commission, 2013; Mitsos, et al., 2012) the primary and secondary education systems are also aligned and cultural differences accommodated (Byrne, et al., 2012). This stresses the dependencies between all three levels of education and reinforces the need of a systemic view of education as opposed to focusing exclusively on one particular level. In the same vein, but perhaps with a rather different motivation, the European University Association stresses the need to respect institutional diversity by making HEIs relate the change agenda to their specific mission and objectives in the further implementation of the Bologna Process (Sursock & Smidt, 2010).

The Bucharest Communiqué of 2012 acknowledges that "higher education should be at the heart of [the] efforts to overcome the [current economic] crisis" (EHEA Ministerial Conference, 2012, p. 1). The question whether higher education should indeed be the central element in anti-crisis policy or it is just one entry in the list of important elements is out of scope of this text. What is important is that this is not the first place where education policy is explicitly related to economic policy. The notion of education as the key for economic and social progress of Europe in official rhetoric dates back to the Treaty of Rome of 1957 (Panitsidou, Griva, & Chostelidou, 2012). Veiga, Amaral and Mendes (2008) argue that there were strong connections between the Lisbon Strategy and the Bologna Process already in the beginning of the 2000s.

These connections are inherited by the mentioned guidelines for further development of the EHEA and the targets of Europe 2020 Strategy, the successor of the Lisbon Strategy, as shown in Figure 1-2. The Europe 2020 Strategy aims at creating a framework for making the growth of the EU smart, sustainable and inclusive. The figure shows that EHEA targets are closely associated

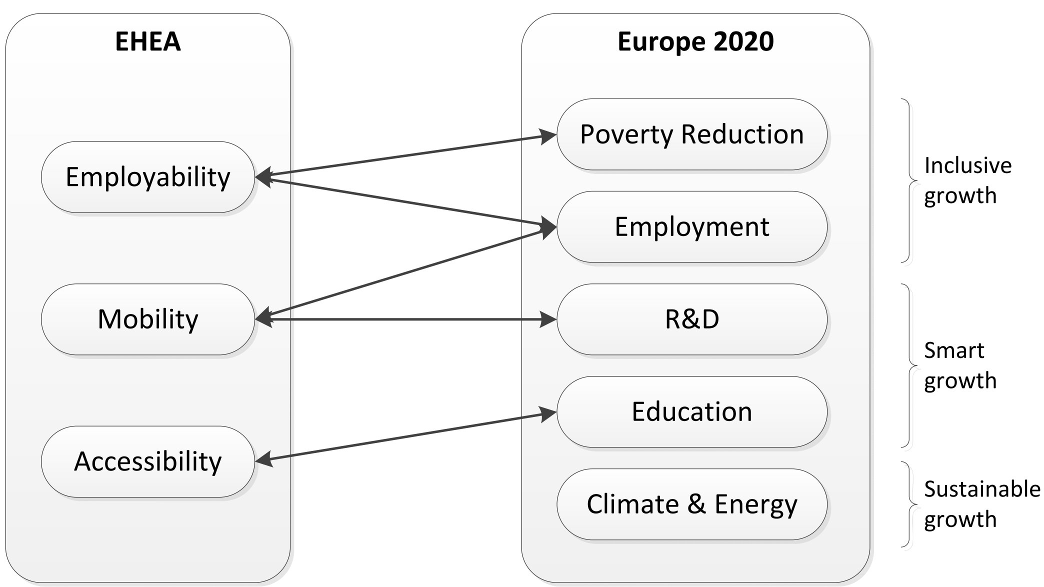

Figure 1-2 – Relationships between EHEA and Europe 2020 Strategies

Source: compiled from European Commission (2013), Romanian Bologna Secretariat (n.d.)

with the targets for smart and inclusive growth. Increasing the employability of graduates through providing them skills needed in the labour market would allow reaching full employment and, hence, reducing poverty. Enhancing the mobility of students and staff would boost R&D through the exchange of experience and positively affect the employment level through making graduates ready to move and work in other countries where their skills are more needed. Finally, making higher education more accessible would raise the overall education level of society.

For future prosperity of Europe, all three targets of EHEA are important. Unfortunately, as the last decade has shown, reaching them is a quite complicated task. The following sections will discuss the observable effects from employability, accessibility and mobility, as well as the challenges of motivating or preventing reaching these targets.

1.2 Employability of Graduates and Skill Mismatch

Improving the employability of graduates is, perhaps, the most challenging task of all three, and is predicted to rise in importance (Gül, Gül, Kaya, & Alican, 2010). There is some evidence that employers accept masters and doctors with relative ease, but bachelors face certain challenges with respect to employability (Sursock & Smidt, 2010).

Dividing long tertiary studies into shorter cycles to improve graduates' time-to-market, creating a system of degrees and study programmes comparable across all participating countries and ensuring the quality of higher education are all important factors in enhancing the chances of finding a job. However, a major challenge for LEMS is ensuring the alignment of the skills of the working-age population with the requirements of the labour market.

This alignment requires two actions, which are important but drastically difficult to implement:

- Ensuring that the contents of the study programmes correspond to the realities of the near-future labour market
- Keeping the demand for study programmes in line with the demand for particular skills in the near-future labour market

One of the major problems in taking these actions is time lags. Firstly, it takes several years for students to graduate HEIs. Hence, both actions require forecasting what skills will be demanded in the future. These forecasts will never be accurate enough, especially in the rapidly evolving fields of study. Secondly, HEIs are quite inert, and it may require years for them to introduce substantial changes in curricula. Thirdly, because skill requirements in some industries,

Table 1-1 – Number of Tertiary Graduates in Europe, 2001–2009, by Field of Study (thousands)

Field of Study	2001	2002	2003	2004	2005	2006	2007	2008	2009
Non-Technical Fields	**1328**	**1404**	**1473**	**1654**	**1821**	**1869**	**1919**	**2064**	**2010**
Social Sciences, Business, Law	990	1058	1110	1259	1383	1391	1427	1544	1517
Arts & Humanities	339	346	363	395	438	477	492	520	493
Technical Fields	**714**	**750**	**804**	**821**	**863**	**879**	**904**	**944**	**934**
Engineering, Manuf., Constr.	408	429	458	467	485	492	511	529	546
Science, Math., Computing	306	321	346	354	378	388	393	415	389
Other Fields	**900**	**950**	**1017**	**1058**	**1136**	**1202**	**1246**	**1302**	**1315**
Health & Welfare	403	428	470	495	544	572	599	626	663
Teacher Training & Educ. Sci.	334	351	365	373	387	418	422	437	407
Services	112	117	125	133	146	147	158	169	176
Agriculture & Veterinary	51	54	56	57	60	65	67	71	69

Source: Eurostat data on EU-27 countries excluding Luxembourg and Greece and including Iceland and Norway, own calculations.

Table 1-2 – Annual Growth in the Number of Tertiary Graduates in Europe, 2001–2009, by Field of Study

Field of Study	2001	2002	2003	2004	2005	2006	2007	2008	2009
Non-Technical Fields	**5%**	**6%**	**5%**	**12%**	**10%**	**3%**	**3%**	**8%**	**–3%**
Social Sciences, Business, Law	5%	7%	5%	13%	10%	1%	3%	8%	–2%
Arts & Humanities	3%	2%	5%	9%	11%	9%	3%	6%	–5%
Technical Fields	**5%**	**5%**	**7%**	**2%**	**5%**	**2%**	**3%**	**4%**	**–1%**
Engineering, Manuf., Constr.	6%	5%	7%	2%	4%	1%	4%	4%	3%
Science, Math., Computing	3%	5%	8%	3%	7%	3%	1%	5%	–6%
Other Fields	**11%**	**6%**	**7%**	**4%**	**7%**	**6%**	**4%**	**5%**	**1%**
Health & Welfare	8%	6%	10%	5%	10%	5%	5%	4%	6%
Teacher Training & Educ. Sci.	13%	5%	4%	2%	4%	8%	1%	4%	–7%
Services	17%	5%	7%	6%	9%	1%	7%	7%	4%
Agriculture & Veterinary	–1%	5%	4%	2%	4%	9%	3%	5%	–2%

Source: Eurostat data on EU-27 countries excluding Luxembourg and Greece and including Iceland and Norway, own calculations.

like Information Technology, change rapidly, HEIs are reluctant to teach the latest skills and instead choose to focus on fundamental skills, leaving it to graduates to get further education in some courses or to employers to provide on-the-job training.

However, even if the contents of study programmes are appropriate, the challenge of matching the supply and demand of skills in the labour market remains. For instance, Tables 1-1 and 1-2 show that the non-technical fields of study were gaining much more popularity than the technical fields in the first decade of the 2000s. Will the labour market be able to absorb all these graduates? How to ensure that applicants to HEIs base their field-of-study decisions on the forecasted needs of the labour market and not on what is currently popular? These are the major questions policy-makers face. While a useful concept for re-training older employees, the lifelong learning programme does not help in matching the supply and demand for younger population.

Table 1-3 – Findings on the Incidence of Overeducation in European Countries, per cent

Country	General Population	Male	Female	Youth [a]
Austria	58.0			1.1 – 10.6
Belgium	10.5 – 54.2			2.0 – 59.0
Czech Republic	50.0	17.4	12.7	1.5 – 9.3
Denmark	34.0			
Estonia	39.0			2.2 – 8.4
Finland	11.1 – 27.0	10.3	14.5	3.3 – 14.1
France	28.0	11.2	17.6	4.4 – 13.9
Germany	11.8 – 60.6	12.3 – 15.6	10.7 – 19.1	2.2 – 12.6
Greece	32.0	26.8	15.0	
Hungary	37.0	23.6	19.8	
Iceland	30.0			
Ireland	33.0			
Italy	13.9 – 71.5	14.9 – 21.3	12.8 – 18.4	4.0 – 19.0
Latvia	43.0			
Lithuania	31.0			
Luxembourg	27.0			
Netherlands	11.2 – 39.0	8.7 – 11.5	12.2 – 13.6	2.9 – 41.7
Norway	16.6 – 34.0			2.5 – 20.4
Poland	13.9 – 29.0			
Portugal	12.6 – 33.0	16.1	14.8	3.4 – 6.5
Romania	25.0			
Slovakia	49.0			
Slovenia	36.0			
Spain	13.8 – 37.2	23.2	24.0	6.5 – 24.8
Sweden	27.0			
Switzerland	14.9 – 13.4	13.3 – 15.0	13.5 – 14.7	
UK	13.0 – 36.8	19.1 – 25.0	20.5 – 27.0	13.7 – 53.0

Source: compiled from compiled from Baert, Cockx and Verhaest (2013), Barone and Ortiz (2010), Bauer (2002), Blázquez and Budría (Blázquez & Budría, 2012), Brynin and Longhi (2009), Budría (2011), Büchel and Battu (2003), Büchel and van Ham (2003), Cainarca and Sgobbi (2012), Chevalier (2003), Croce and Ghignoni (2012), Cutillo and Di Pietro (2006), Dekker, de Grip and Heijke (2002), Frei and Sousa-Poza (2012), Ghignoni and Verashchagina (2014), Groot and van den Brink (2000), Hartog (2000), Jauhiainen (2011), Jensen, Gartner and Rässler (2010), Karakaya, Plasman and Rycx (2007), Kiersztyn (2013), Mavromaras, McGuinness, O'Leary, Sloane and Fok (2010), McGuinness and Bennett (2007), Murillo, Rahona-López and del Mar Salinas-Jiménez (2012), Ortiz and Kucel (2008), Ramos and Sanromá (2013), Sánchez-Sánchez and McGuiness (2015), Støren and Wiers-Jenssen (2010), Sutherland (2012), Verhaest and Omey (2010; 2012), Wirz and Atukeren (2005)

The table includes estimates starting from the 1990s only.

* Defined as aged under 31. Also included are studies on recent (less than 5 years) graduates.

Failing to propose perfect policies to match skill supply and demand leads to skill mismatch. Depending on some nuances, this phenomenon is referred to as over-/undereducation, over-/underskilling and over-/underqualification. I will refer to the first of these concepts throughout this text. An ***overeducated*** (***undereducated***) individual is defined as having more (less) education than required in his current job. Depending on how they measure mismatch, studies show that generally in European countries, about 10 per cent to one-third of the labour force are overeducated and around 20 per cent are undereducated, the total mismatch, thus, being around 30 to 50 per cent of the labour force. Tables 1-3 and 1-4 summarise the relevant statistics. When comparing tables' columns, it is necessary to bear in mind that, while male and female intervals

Table 1-4 – Findings on the Incidence of Undereducation in European Countries, per cent

Country	General Population	Male	Female	Youth [a]
Austria				8.4 – 30.6
Belgium	25.8 – 32.4			5.4 – 25.5
Czech Republic		17.8	25.6	11.1 – 17.8
Estonia				18.4 – 33.1
Finland		39.4	37.9	10.9 – 26.3
France		44.9	41.4	14.4 – 15.4
Germany	12.1	10.4 – 18.8	15.6 – 21.5	6.3 – 25.9
Greece		21.8	25.6	
Hungary		19.9	24.9	
Italy	17.1	17.7 – 24.7	16.3 – 32.8	11.7 – 22.5
Netherlands	12.0	3.8 – 16.7	2.1 – 14.3	5.3 – 25.2
Norway				11.6 – 29.1
Portugal	17.0 – 38.0	16.6	18.9	22.6 – 50.8
Spain	11.0 – 25.6	33.3	27.8	7.1 – 23.8
Switzerland	1.9	2.0	1.8	
UK	17.0	40.6	43.7	5.5 – 26.1

Source: compiled from Bauer (2002), Cainarca and Sgobbi (2012), Frei and Sousa-Poza (2012), Ghignoni and Verashchagina (2014), Groot and van den Brink (2000), Hartog (2000), Karakaya, Plasman and Rycx (2007), Murillo, Rahona-López and del Mar Salinas-Jiménez (2012), Sánchez-Sánchez and McGuiness (2015), Verhaest and Omey (2012)

The table includes estimates starting from the 1990s only.
* Defined as aged under 31. Also included are studies on recent (less than 5 years) graduates.

generally are comparable because studies on male subpopulation typically also study female subpopulation, the results for youth should be compared with the results for general population cautiously. Empirical research has concentrated on general population; thus, the variation of measures employed for analysis is much higher than in the case of youth.

In this text, mismatch is measured by the ISCO-based measure. Its precise definition, the discussion of its strengths and weaknesses, and the comparison to other measures are postponed until Section 3.2. At this moment, it is important to show the major trends in mismatch.[2]

Figure 1-3 shows that there are pronounced trends in mismatch. In general, overeducation has been increasing and undereducation decreasing over the last decade. From the point of view of labour market participants, this is a worrying tendency. Indeed, undereducation means that a person without formal qualifications for a position, nevertheless, is hired on it. The figure shows that the share of such observations has been falling rapidly, indicating the presence of a pressure on individuals to invest in formal education. On the other hand, the stable upward trend in overeducation shows that individuals who invested in formal education are increasingly failing to find a job matching this education and are forced to occupy less demanding positions. Overall,

[2] The trends are based on European Social Survey data, which will be introduced in Section 3.1.

both trends mean that it becomes more difficult to find a good return on the investment in education.

Comparing the young (aged 15–29) with the mature (aged 30 and above) labour market participants, one can observe significant disadvantages for the former. On average, they face both higher overeducation risk and lower undereducation risk. In other words, they more frequently occupy lower-level positions, whatever their education level.

Figure 1-4 shows the average mismatch dynamics separately for young/mature males and young/mature females. Females are more frequently overeducated and less frequently undereducated than males of their age group. Moreover, gender differences are stable over time. At the same time, both young males and females face a higher overeducation risk and a lower undereducation risk than the mature of the same sex, as was the case in the total population. The figure also shows that the differences between the two generations in the exposure to overeducation are decreasing for males, but not for females.

One might assume that the disadvantaged groups, such as the non-natives, the disabled and those living outside big cities, are hurt by facing higher overeducation and lower undereducation incidence than the natives, the healthy and those living in big cities, respectively. Tables in Appendix B show relevant comparisons.

In general, the disabled are not at disadvantage with respect to mismatch. However, there are sizeable differences in the exposure of the young disabled individuals in around one-third of countries. Specifically, they are hurt with both higher exposure to overeducation and lower to undereducation in Latvia. In Finland, Ireland and Poland, they are more frequently overeducated; while in Czech Republic, Hungary, Lithuania, Slovakia, Switzerland and Turkey less frequently undereducated than those without serious health problems.

In contrast, there are few troubling cases when inhabitants of small cities or rural areas are at disadvantage, compared to those living in big cities. A notable example is Lithuania, where inhabitants of both small cities and rural areas face a higher overeducation risk than those living in big cities.

Minorities and second-generation immigrants rarely are at a greater disadvantage than the natives. This is not the case with first-generation immigrants, especially from Central and Eastern Europe (CEE). Immigrants from Latin America, Africa and Asia also face disadvantages, but mostly from lower undereducation. On the contrary, immigrants originating in CEE are subject to higher risk of overeducation in almost all countries. With respect to young CEE immigrants, however, this holds only for around half of the countries.

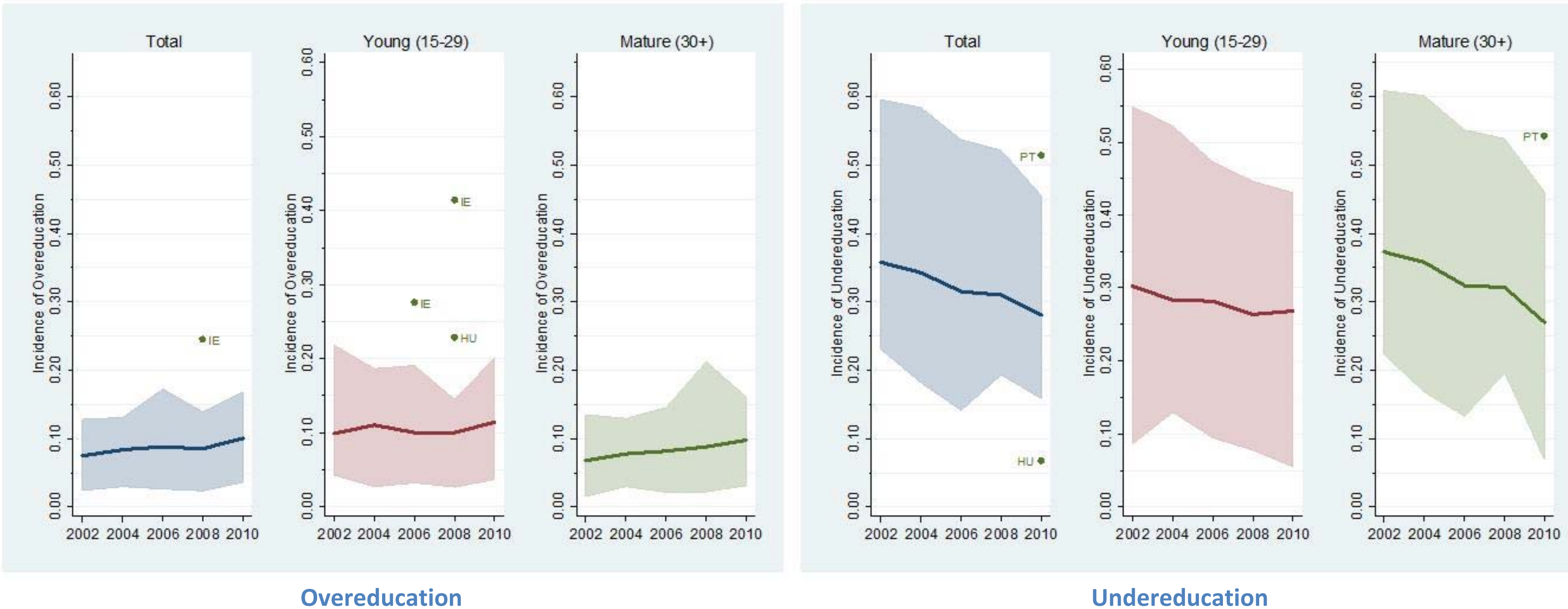

Figure 1-3 – Dynamics of Average Incidence of Mismatch: Young (15–29) vs. Mature (30+)

Source: European Social Survey data, own calculations

Based on data from countries appearing in all five European Social Survey rounds (Belgium, Denmark, Finland, France, Germany, Hungary, Ireland, the Netherlands, Norway, Poland, Portugal, Slovenia, Spain, Sweden, Switzerland and the UK); the shaded area shows the range of data. Averages are unweighted. Labelled points outside shaded areas represent countries that have significantly different mismatch incidence from other countries in that round. These outliers have incidence either above $p_{75} + 1.5 \times IQR$ or below $p_{25} - 1.5 \times IQR$, where p_{25} and p_{75} are, respectively, 25th and 75th percentiles of the incidence distribution in a given round and IQR is the interquartile range (i.e., $p_{75} - p_{25}$). The outliers in a given European Social Survey round do not participate in the calculation of the averages in that round.

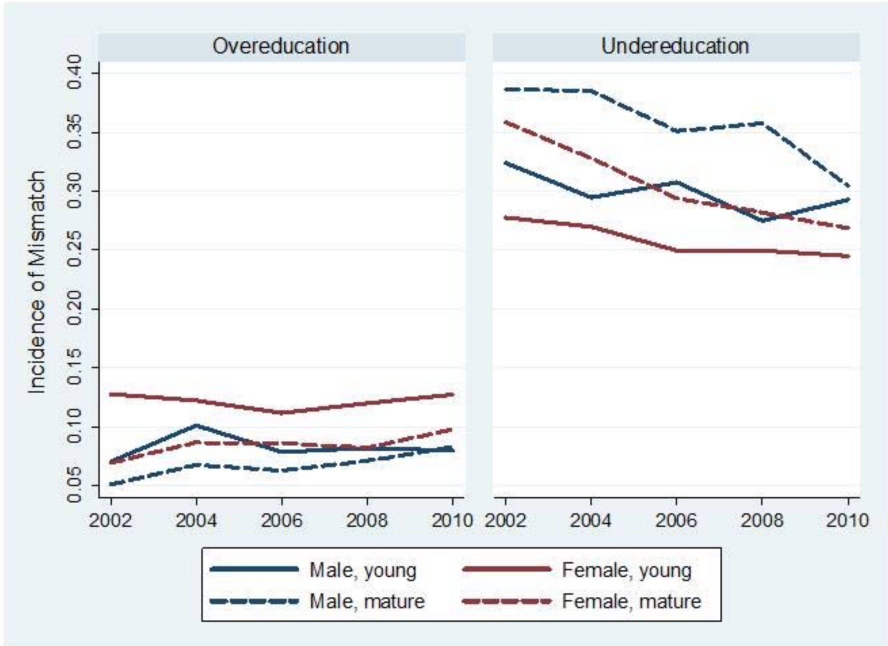

Figure 1-4 – Dynamics of Average Incidence of Mismatch: Males vs. Females

Source: European Social Survey data, own calculations

Figure 1-5 shows differences in the exposure to overeducation risk of immigrants and the fully natives in selected countries. In Cyprus, Finland, France, Israel and Ireland, the difference among the young is greater than among the mature. However, in most countries, either young and mature immigrants are affected equally or, in the majority of cases, mature immigrants are hurt more. Actually, in several countries, young immigrants are hurt *less* than young fully natives, while mature immigrants are hurt *more* than mature fully natives (Austria, Denmark, Germany, Hungary and the Netherlands). In addition, in Slovenia, the differences for both the young and the mature are negative (but for the mature, they are close to zero).

Mortensen and Vilella-Vila (2012) note that there is a considerable excess supply of low-, medium- and high-skilled labour force in 2013. However, they argue, mismatch will be reduced until near-zero by 2020; in particular, there will be full match of the high-skilled. This does not seem plausible, given current trends.

Several ways of dealing with mismatch have been proposed. Some researchers propose to restructure the economy to take advantage of a large supply of university graduates (Murillo, Rahona-López, & del Mar Salinas-Jiménez, 2012), thus, reshaping the labour market to fit whatever the education market outputs. Others propose to reshape the education system to fit whatever the labour market demands (Guironnet & Peypoch, 2007) by, among others, increasing the selectivity of access to universities or motivating students to choose educational tracks that

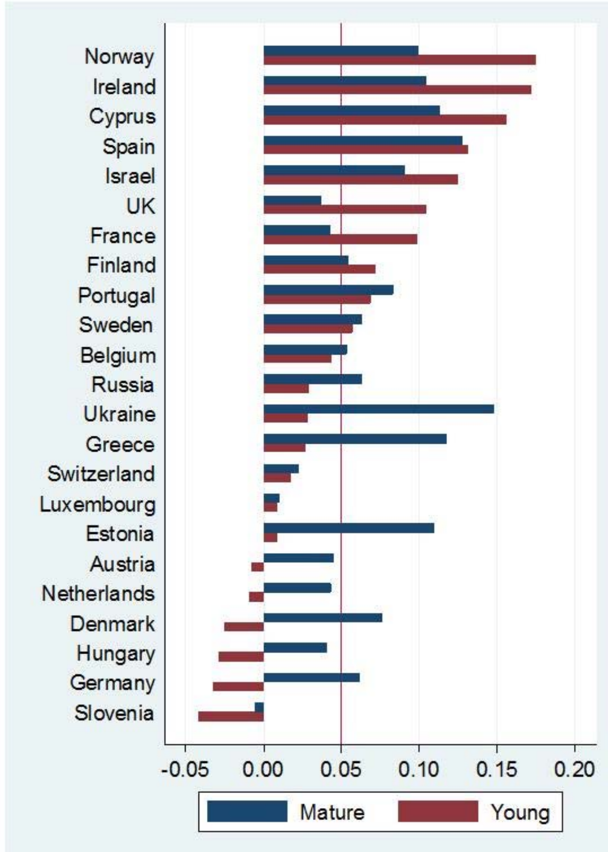

Figure 1-5 – Differences between the Overeducation Incidence of Immigrants and Fully Natives

Source: European Social Survey data, own calculations

To be read: In Norway, young immigrants have a 17 percentage points higher overeducation incidence than young fully natives (i.e., who themselves and whose both parents were born in Norway). Countries sorted by the differences among the young. For each country, data from all available European Social Survey Rounds are pooled. Only countries having at least 15 immigrants among the young and the mature are included.

fit with employers' demands (Cainarca & Sgobbi, 2012). Still others propose to improve the link between the education system and the labour market (Budría, 2011), changing both systems to fit each other.

As already noted, managing higher education separately from the labour market does not seem to be a proper solution. Nor does the other extreme – giving higher education into full slavery to the labour market. Since universities are a crucial element in innovation systems and a cornerstone of a knowledge society, they should have both (1) close contacts with businesses to be able to identify the needs for research, the necessary competencies and the number of graduates and (2) independence in defining the areas of (fundamental) research they consider important for seeking new knowledge and understanding, but which may not lead to any immediate applicability or demand from the market (Hansen & Lehmann, 2008).

Given that progress is rapid in many partitions of knowledge, could key general competencies that should be taught to all students, irrespective of their field of study, be distinguished? A recent study shows that a panel of international experts agreed that the most important competencies future students should be taught are systemic thinking and handling of complexity, anticipatory thinking and critical thinking (Rieckmann, 2012). This should help them adapt to the changing realities of our world, and at the same time make them act in a manner consistent with the sustainable growth objective.

As a side note: The European Parliament and the Council of the EU have developed their list of eight key competencies for lifelong learning. It includes communication in the mother tongue, communication in foreign languages, mathematical competence and basic competences in science and technology, digital competence, learning to learn, social and civic competence, sense of initiative and entrepreneurship, and cultural awareness and expression (European Parliament & EU Council, 2006). Together, it is hoped, they would prepare individuals for more flexible employment schemes, such as multiple entries in and exits from the labour market, in a more dynamic economy. It remains to be seen whether these eight competencies turn out to be more important than the three proposed by international experts.

1.3 Accessibility of Higher Education and Government Financing

The 2001 reform of Italian higher education system, which introduced the two-tier degree structure, increased the diversity of fields of study and simplified study contents, led to a 15% increase in the probability of going to university, mostly through motivating individuals with good schooling ability but unfavourable family background to continue studies (Cappellari & Lucifora, 2009). Nevertheless, making higher education accessible for everyone, irrespective of one's socio–economic background, is hard to imagine without a considerable government financing. Indeed, there are policy instruments available to enhance the accessibility of higher education to certain disadvantaged groups such as low-income, minorities or married individuals (Christou & Haliassos, 2006). Nordic countries are the most successful from OECD countries in providing targeted public support for adult education for such individuals (Tuijnman, 2003).

The technological advancements have recently led to the mushrooming of distance learning. In particular, massive open online courses provided by organisations such as Udacity, Coursera and edX and world top universities such as Stanford and MIT give everyone the opportunity to get top-level knowledge without major transaction costs. Unfortunately, there is evidence that distance learning does not increase the participation of socioeconomically disadvantaged groups (Gül, Gül, Kaya, & Alican, 2010). Thus, the development of distance learning should not result in abolishing targeted government aid.

The main current challenge for European HEIs is meeting increasing expectations for education high in quality and diversity at times of financial austerity (Gül, Gül, Kaya, & Alican, 2010). In their theoretical paper, Arcalean and Schiopu (2010) proved that while any government, irrespective of its size, should devote more resources to primary and secondary, rather than tertiary, education, countries with higher taxes (such as the EU Member States) should be more generous at funding tertiary education.

Nevertheless, facing the serious economic crisis in Europe, many countries traditionally having free higher education started discussing the introduction of tuition fees. In particular, it refers to Denmark, Sweden and Finland, with the first two countries actually introducing tuition fees for non-EU/EEA students in 2006 and 2011, respectively (Weimer, 2011).

As a result, in two years, the number of full-degree international students dropped 45 per cent in Denmark (Myklebust, 2010). In Sweden, this led to a drastic 58 per cent reduction in the number of international students enrolled in international master's degree programmes in the first year of introducing fees (Simpson, 2011). Contrary to Niklas Traneus, the head of the project Study Destination Sweden, who believed that within a few years, many Swedish HEIs would attract increasing numbers of international students (Myklebust, 2011), the overall drop in the number of non-EU/EEA students over the following two years amounted to 80 per cent (Valtonen & Lampinen, 2013).

In 2012, legislators in Finland agreed to change legislation to allow universities to charge fees for international students (Myklebust & Dobson, 2013). Student unions in Finland fiercely opposed this initiative, commenting that "tuition fees and Nordic education are essentially incompatible" and "drawing any parallels between Finland and, for instance, the UK or the United States is mindless due to the differences in social structure, culture and language" (Valtonen & Lampinen, 2013).

In general, HEIs charge fees for education in the majority of countries, although there is wide variation across countries regarding the share of students who have to pay (see Figure 1-6). Only the Scandinavian countries and Austria have free higher education for both bachelors and masters, while Cyprus, Greece and Malta provide free education only at bachelor level. A few countries have no tuition fees, but ask for administrative fees; these include Iceland, Germany (except for Bavaria, Hamburg and Lower Saxony, where students are asked to pay), Poland, Czech

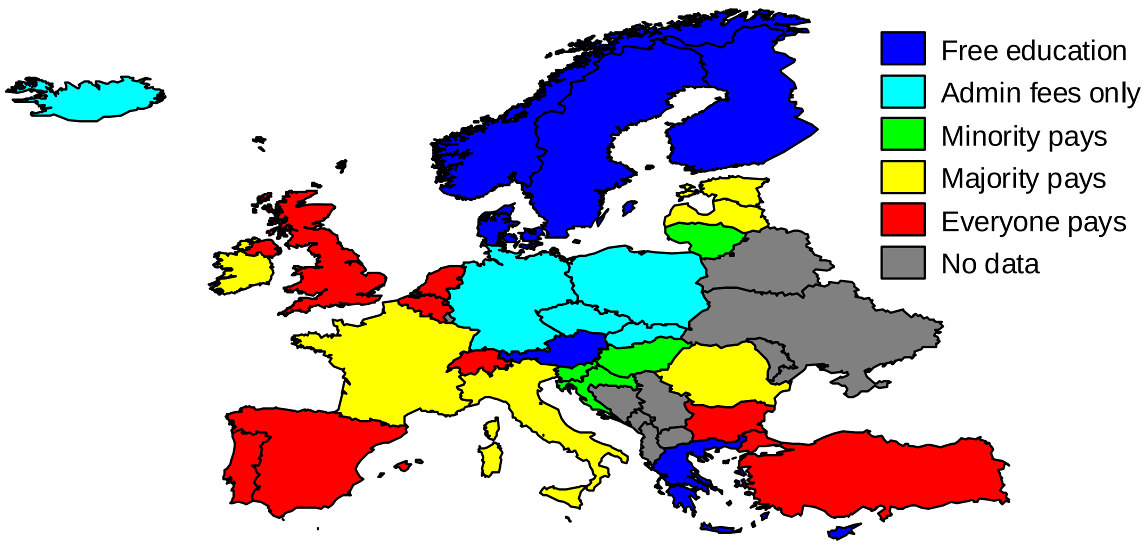

Figure 1-6 – Proportion of Students who Pay Tuition Fees for Higher Education, 2011–2012

Source: compiled based on Eurydice (2012); data on Switzerland taken from Rectors' Conference of the Swiss Universities (2013); data on Spain taken from Ministerio de Educación, Cultura y Deporte (2012)

Data refer to public or government-dependent private HEIs (but not to private HEIs) and cover bachelor and master's students (but not doctoral students).
Country-specific notes:

- Cyprus, Greece, and Malta: studies are free at bachelor level, but master's students have to pay
- Scotland: Bachelor studies are free for students from Scotland and the EU, but not from other parts of the UK; master's students pay tuition fees
- Denmark: part-time bachelors pay tuition fees
- Belgium: data shown for Flemish community; in French community, majority of students pay
- Germany: all students pay tuition fees in Bavaria, Hamburg and Lower Saxony
- Czech Republic: fees are charged at master's level
- Iceland: data shown for public HEIs; government-dependent private HEIs (where around 20 per cent of students study) charge tuition fees

Republic and Slovakia. In other countries, higher education is chargeable, and these countries differ on the amount of government support for students – (1) how many students get it and (2) whether they pay for tuition from the support they get (e.g., loans in the UK) or the government waives tuition fees (e.g., state-funded places in Latvia and Estonia).

Hot debates have been going over decreasing EU funding of R&D in universities. While the European Parliament proposed to invest in research funding €100 billion, the European Commission's proposal amounted to €80 billion. Fearing that subsequent negotiations might further decrease the funding, a Petition for the Attention of the EU Heads of State or Government has been signed by more than 150,000 individuals (Initiative for Science in Europe, 2012). As a result of active participation of different communities in discussions on this issue, European countries agreed that "the funding for Horizon 2020[3] and ERASMUS for all programmes will represent a real growth compared to 2013 level" (European Council, 2013, p. 7), meaning that no cuts in financing are planned. Sometimes it seems that the governing bodies of the EU and its Member States are forgetting their goals for smart, inclusive and sustainable growth and the crucial role of education and research for it when facing the consequences of lavish government spending.

Typically, the governments of European countries spend on financing public HEIs €5000 to €15,000 per student (in PPS) or an equivalent of 30 per cent to 50 per cent of per-capita GDP per student (see Figures 1-7 and 1-8, respectively). Countries can be grouped geographically on their spending patterns.

Scandinavian countries and Cyprus are spending the most. In absolute terms, Norway, Sweden and Switzerland spend over €15,000 and Cyprus over €20,000 per student; while in relative terms, Sweden and Finland spend 50%–60% and Cyprus over 80% of GDP per capita per student. On the other side of the spectrum, Latvia's government is one of the least willing to spend on higher education by both criteria (less than €5000 and only 20%–30% of GDP per capita per student). It is accompanied by Bulgaria and Romania on the absolute measure and by Iceland, Italy and Slovakia on the relative measure.

Tertiary education needs government financing, at least to do fundamental research and provide fair access to higher education to everyone. There are concerns that falling government financing may lead to the marketization and privatisation of higher education, putting the long-term interests of humanity and the fundamental principle of fairness at risk, with unknown consequences (Gül, Gül, Kaya, & Alican, 2010). Nevertheless, there are arguments for increasing the efficiency of HEIs, fostering competition in the education market and substituting public financing of study programmes that are not "public goods" with private funding by increasing tuition fees and providing income-contingent loans (Jacobs & van der Ploeg, 2006).

1.4 Mobility of Students and Staff and Internationalisation

With the world becoming increasingly connected, ensuring that students spend some time during their studies abroad, and academic staff regularly goes to foreign HEIs to exchange

[3] Horizon 2020 is a single financial instrument of research funding, combining all research and innovation funding currently provided through other bodies, such as Framework Programmes for Research and Technical Development, the Competitiveness and Innovation Framework Programme, and the European Institute of Innovation and Technology (European Commission, 2012). Essentially, it is one of the instruments from the toolkit of implementing the Europe 2020 Strategy.

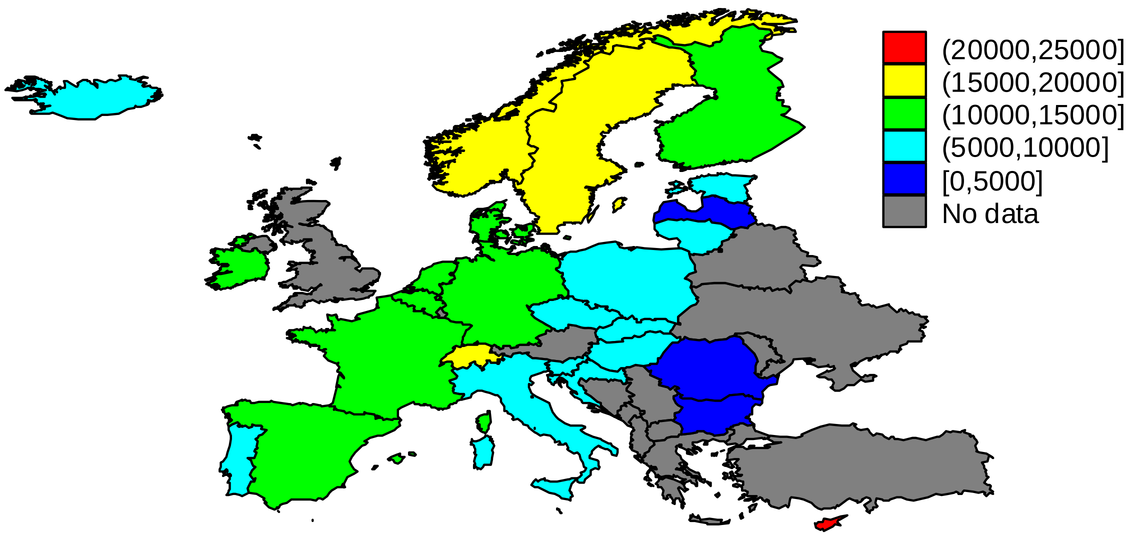

Figure 1-7 – Annual Expenditure on Public Higher Education Institutions per Student in EUR PPS in 2010

Source: compiled based on Eurostat data

Data for Estonia and Germany taken from 2009.

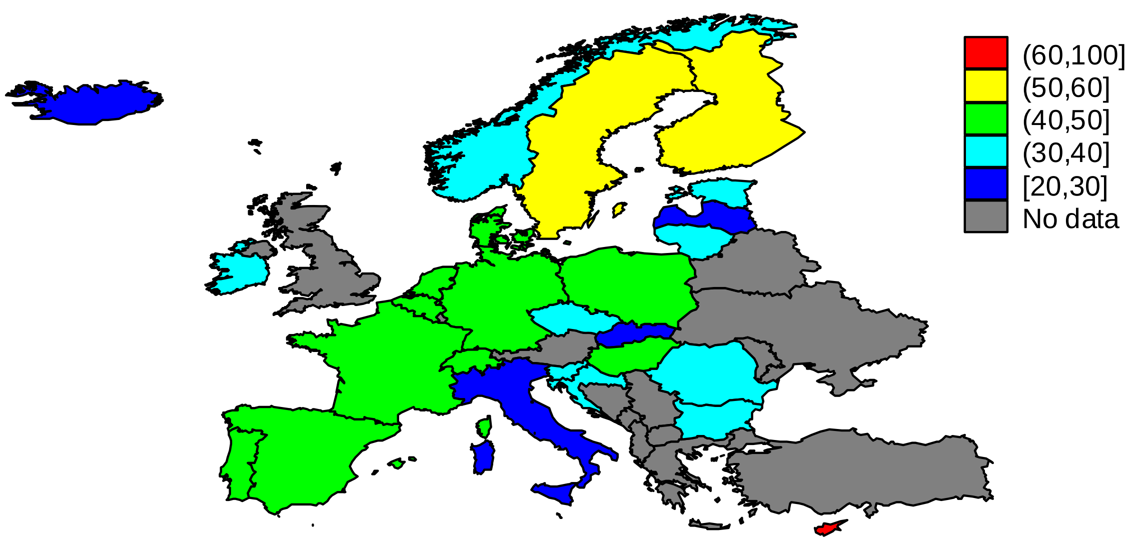

Figure 1-8 – Annual Expenditure on Public Higher Education Institutions per Student as % of GDP per Capita in 2010

Source: compiled based on Eurostat data

Data for Estonia and Germany taken from 2009.

experience, is important for international competitiveness. Internationalisation was the third most important change driver for HEIs in 2008–2010 and is expected to become the most important in 2011–2015, with more institutions integrating international partnerships into their strategies and around one-third placing them at the heart of their strategy (Sursock & Smidt, 2010).

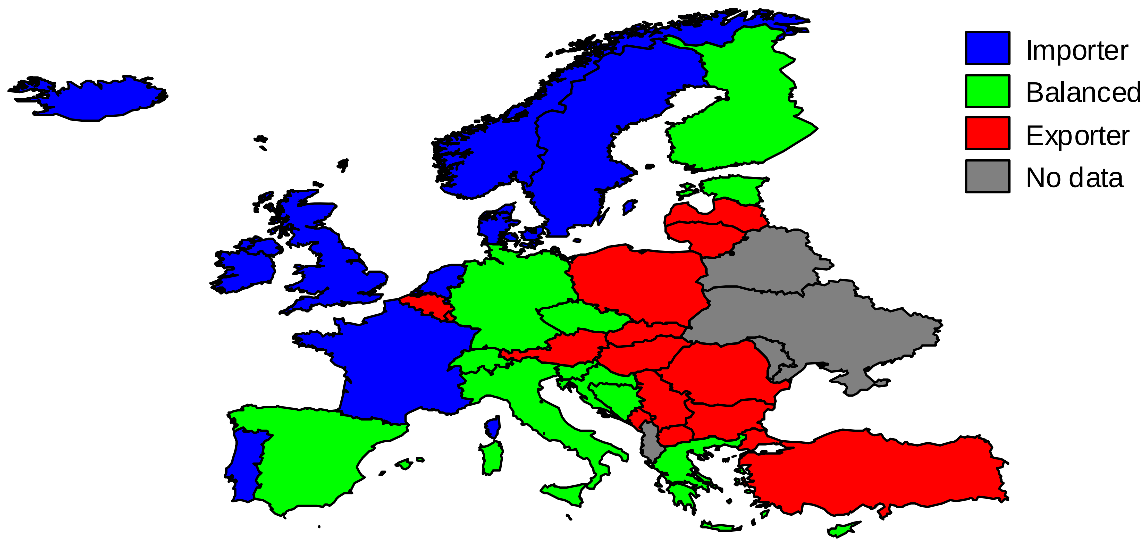

Figure 1-9 – Balance of International Student Flows in 2010

Source: Sursock and Smidt (2010, p. 78, Map 13)
Data for Bulgaria taken from year 2007 (Sursock & Smidt, 2010, p. 78, Map 12).

Latest data show that both student and, to a lesser extent, staff mobility has been continuously increasing (Sursock & Smidt, 2010). As shown in Figure 1-9, there are clear imbalances in student mobility patterns, with Eastern countries serving as exchange student exporters and Western countries as their importers. This division of countries has been fairly stable in the first decade of the 2000s (Sursock & Smidt, 2010).

The most important benefit mobility gives to students is intercultural experience, which enhances their careers both upon returning to the home country and if they decide to work abroad (Button, Green, Tenhnah, Johansson, & Baker, 2005; Hansen & Lehmann, 2008; Tamtam, Gallagher, Olabi, & Naher, 2012). For academic staff, it is an opportunity to develop new courses, possibly jointly with foreign academics (Hansen & Lehmann, 2008), and to engage in inter-collaborative research with colleagues from foreign HEIs (Davies, 2008).

Several obstacles to student mobility have been noted in the literature. Firstly, students from poorer families are less likely to participate in exchange programmes (Davies, 2008). Secondly, as English is the *de facto* language of instruction for exchange students, the quality of the output from increased mobility depends crucially on the knowledge of English by both students and instructors; unfortunately, there is evidence that the proficiency in English is still insufficient (Davies, 2008; Tamtam, Gallagher, Olabi, & Naher, 2012). Thirdly, academic calendars lack harmonisation across Europe (Sursock & Smidt, 2010), making it difficult to coordinate short-term student exchange. Finally, misunderstanding of the nature of the mobility priority and lack of support from academic staff leads to problems in the recognition of study abroad (Sursock & Smidt, 2010).

Thus, HEIs should provide preparation and support for both students and staff to reap the maximum benefit from international experience (Davies, 2008), which requires that mobility, especially for students, is central to the internationalisation strategy of HEIs (Sursock & Smidt, 2010).

The negative side of mobility is that for students with experience of studies abroad, especially in wealthier countries, it works as an additional motivator to emigrate (Davies, 2008). One could, though, argue that the tendency to emigrate to wealthier countries exists independently of whether there is wide participation in foreign exchange programmes. In any case, the benefits of increased mobility far outweigh the risks of brain drain.

1.5 Summary

In this chapter, it was discussed that:

- Education and labour markets are mutually dependent and complementary, providing a motivation for considering both of them as parts of the labour–education market system (LEMS)
- Although seeming to be unstable during crises, there is an understanding at the strategic level of the EU that the Union’s future prosperity depends crucially on LEMS
- Higher education reforms aiming at creating European Area of Higher Education (EHEA) were started, but while EHEA was launched in 2010, the goals of the reforms have not yet been achieved
- The key priority areas are the employability of graduates, accessibility of higher education and mobility of students and staff
- While problems remain in all three areas, the most problematic area appears to be employability, where the key struggle is increasing overeducation, indicating that graduates are increasingly unable to find a job matching their education level

Chapters 3 and 4 analyse the problem of overeducation further. The former chapter studies the factors that influence the risk of overeducation, while the latter provides evidence of some of its negative consequences on the labour-market and educational situation of an individual. Chapter 5 presents the results of agent-based modelling of LEMS taking into account the effects

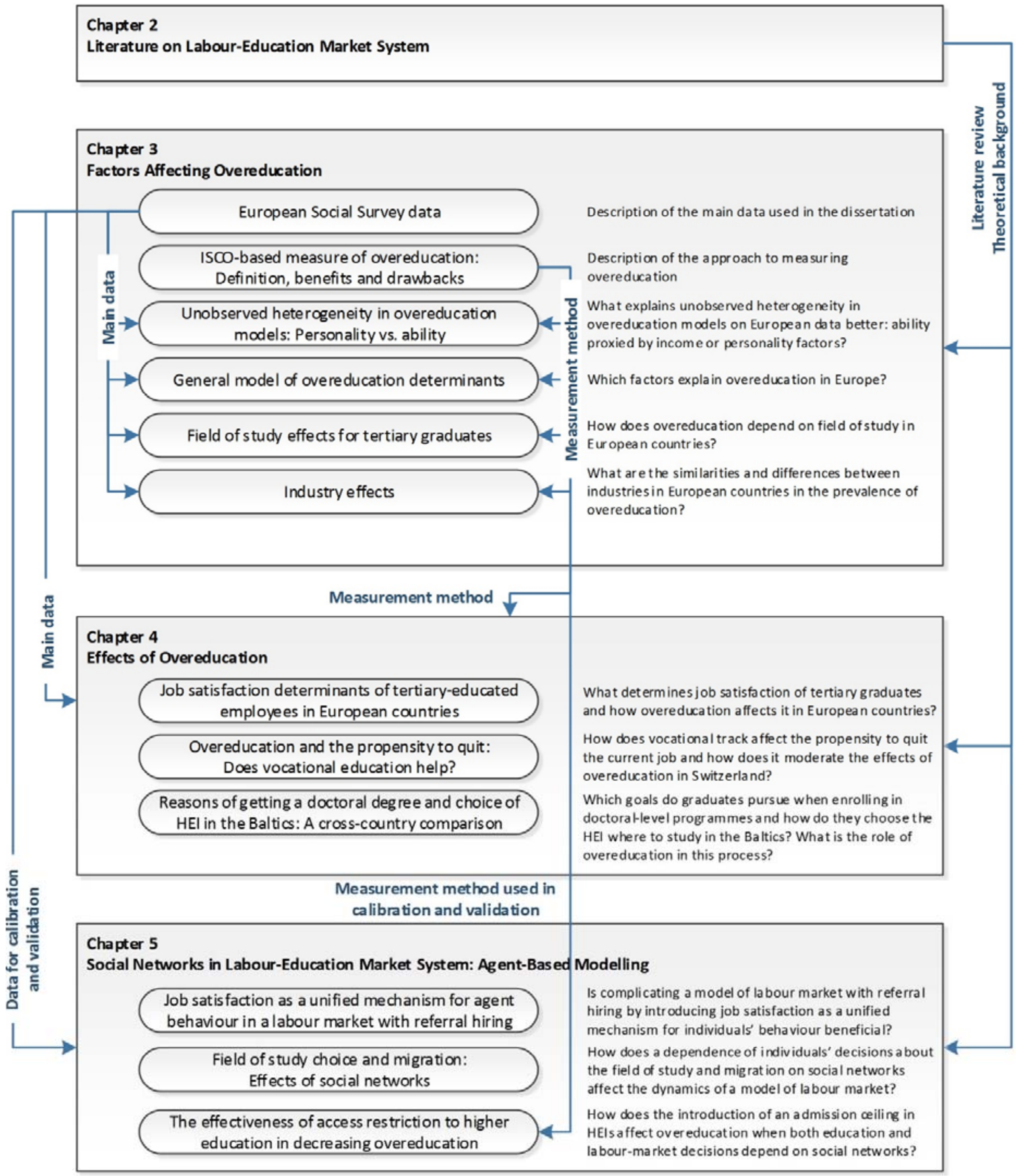

Figure 1-10 – Structural View of the Remaining Chapters

from social networks. Figure 1-10 gives a general overview of the remaining chapters of the dissertation.

2 LITERATURE ON LABOUR–EDUCATION MARKET SYSTEM

This chapter provides the review of theoretical and empirical literature on the issues studied in later chapters. It starts with a section on overeducation, summarising the debates on its existence and measurement, as well as causes and effects. Section 2.2 describes the theoretical grounding and empirical evidence of the effects of social environments on LEMS. Section 2.3 analyses how the events of finding and losing job are modelled in the agent-based modelling literature. The last section concludes.

2.1 Overeducation: Existence, Effects, Measures and Influencing Factors

In Section 1.2, it was shown that the current state and the recent dynamics of mismatch – both overeducation and undereducation – are troubling. However, whether some policy actions are required to correct for these failures depends on the seriousness of the consequences of mismatch. This section will show that the effects on different aspects of the quality of life and employment of individuals are negative and sizeable.

Nevertheless, the section will start with a brief review of the theoretical debate on the question of the reality of overeducation. Then empirical labour-market effects are described. The section continues with a review and comparison of measures of mismatch used in the literature. Finally, the major groups of factors influencing the risk of mismatch are described.

2.1.1 Theory behind Mismatch: Is Overeducation Real?

On a theoretical level, the existence of education/qualification/skill mismatch represents an inefficiency of the LEMS. Indeed, overeducation means that firms are not fully utilising the productivity of their workers, while undereducation means that firms are not operating at their productive frontier by employing less productive workers than they should. The failure can be both on the demand side (labour market) and on the supply side (education market). Hence, the inefficiency exists in the *system* of these two markets, and not exclusively in the labour market.

Nevertheless, theories explaining the (non-)existence of mismatch generally focus on the labour market, assuming that the education market is given and cannot be changed. They differ in their assumptions about the labour market. Two groups of theories can be distinguished.

The first group is formed by theories explaining mismatch as a temporal phenomenon or a result of imperfect measurement of the extent of match. The underlying assumption behind these models is that the labour market is a perfect market, quickly dealing with any imperfections using its powerful market forces. In effect, these theories claim that mismatch does not exist.

The second group of theories also focuses on the labour market, but assumes that it is an imperfect market, where mismatch is a persistent market failure. They then propose different explanations of why market forces cannot correct it.

It should be stressed again that none of these theories look on the specifics of the education market. In particular, they do not look on whether the education system is stratified, whether access to higher education is selective, whether there are large differences in the demand for fields of study that are not aligned with the demand in the labour market and other important particularities that policy-makers have to take into account.

2.1.1.1 Theories Assuming the Labour Market Is Perfect

The ***human capital theory*** (Becker, 1993) assumes that firms look on worker's productivity, which is determined by their level of human capital, and set the wage in accordance with individual's marginal product. This human capital is a sum of formal education and training (or labour market experience). Note that the extent of mismatch is typically measured by formal education, not accounting for training or experience. The theory then argues that the overeducated do not receive training from the firm, the matched receive training and the undereducated receive more training, so that in the end, the productivity of all workers is the same. If the potential productivity of the overeducated is still higher than that of the matched after training, the firm may change its production processes to reap the full benefits of this productivity. Otherwise, because the overeducated are not paid according to their potential marginal product, they will quit to a more appropriate job. As in a typical neoclassical framework, timing is not important, so immediate adjustment is assumed.

Related to human capital theory are arguments about ***unobserved heterogeneity***, which assume that the labour market is working perfectly, but we have imperfect instruments for measuring the match between an individual and a job. Hence, our measures show that the employee is mismatched in his position just because we cannot take into account some of his unobserved or intangible characteristics, which, together with the observed ones, make him matched.

The ***job mobility theory*** (Sicherman, 1991), also called ***career mobility theory***, associates overeducation with a state before moving up the career ladder. Workers agree to spend some time in mismatched positions to accumulate the right amount of skills to get a match. They, thus, sacrifice immediate increase in pay for larger wages in the future. If they do not receive the expected promotion, however, they will quit their job. Again, it is assumed that market forces will quickly adjust the position of the overeducated by promoting them or forcing them to quit.

2.1.1.2 Theories Assuming the Labour Market Is Imperfect

The ***job-competition theory*** (Thurow, 1975), sometimes referred to as ***credentialism***, assumes that formal education does not allow firms to assess workers' productivity. However, it allows them to evaluate how much training is required for workers to reach target productivity. Firms then are more likely to employ individuals in whose training they would have to invest less. Thus, a higher level of education allows individuals to get a higher rank in the job queue. Once the individual is the first in the queue, he is assigned the job and paid the pre-defined wage for that job. This means that the more people invest in a higher education, the less likely it is that one with a lower education receives the position. Thus, credential inflation is generated, hence the second name of the theory.

The ***signalling theory*** (Spence, 1973) introduces the costs of education. In its setup, individuals have two types of characteristics: immutable indices (e.g., gender, race) and alterable signals (e.g., education level). Firms assess candidates' productivity based on their indices and signals. Individuals invest in signals that maximise the difference between their benefits (e.g., future wages) and costs (time, psychological and monetary costs). It is then argued that for education to be a valuable signal, allowing firms to differentiate between lower-productivity and higher-productivity workers, its costs have to be negatively correlated with employee's productive capabilities. When this is not the case, everyone will invest in the same education, and firms will have to apply different selection criteria.

The ***assignment theory*** (Tinbergen, 1956; Sattinger, 1993) assumes that the productivity realised by the individual in the job is the maximum of his own productivity (which depends on his qualifications) and the productivity of the job itself (the complexity of the job, the technology used in it). Then individuals with higher qualifications are assigned to more complex jobs, and those with lower qualifications – to less complex jobs. In this framework, mismatch can be explained in at least three ways. Firstly, it is a mistake in the assignment process, arising due to its complexity (Büchel, 2001). Secondly, individual's utility may be maximised in positions where he is mismatched, because his utility also depends on other variables, which compensate for the effect on utility from mismatch (Ortiz & Kucel, 2008). Thirdly, the balance between the supply of and demand for individuals with different skills is important (Quintini, 2011), as if there is excess supply of the highly skilled, these will fill the lower-skilled positions.

The ***job-search theory*** (Stigler, 1961; 1962) introduces the costs of job seeking. In its setup, individuals search for jobs until the marginal costs of continuing search equal its marginal benefits (discounted lifetime earnings). Because of its costs, the longer the search continues, the more willing the individual becomes to accept a lower wage level, i.e., he decreases his reservation wage with time. Thus, his reservation wage may fall to the level of jobs for which he is overeducated if he is unable to find anything matching him better.

The ***technological change theory*** (Mendes de Oliveira, Santos, & Kiker, 2000) argues that due to technological change, employers have to hire new workers with better qualifications than their current workers have. Because of inherent costs, firms cannot change their whole workforce immediately. Thus, the periods where individuals employed in similar positions have different qualifications may be lengthy. One of the core assumptions behind this theory is that firms are not willing to invest in upgrading the skills of current workforce to the level of the newly hired.

The ***theory of differential overeducation*** (Frank, 1978) assumes that individuals search for jobs that minimise their extent of overeducation, as it maximises their income. For married couples, the process of income maximisation is more complicated. Frank assumes that the husband first optimises his income (i.e., minimises his overeducation), taking into account the jobs available in the global market. Then, the wife optimises her income, but her search is restricted to the jobs available in the local market where her husband found his job. Hence, the theory argues, the probability of overeducation is higher for wives than for husbands.

2.1.2 Effects of Overeducation

For the most part, empirical research has been concerned with the effects of overeducation on wages. Mainly, the ORU[4] specification, introduced by Duncan and Hoffman (1981), which decomposes the wage effect of education into a part required by the job and the part stemming from over/undereducation, has been used[5]. The general form of the ORU equation is

$$\ln w = \alpha + \beta X + \beta_r Y_r + \beta_o Y_o + \beta_u Y_u + \varepsilon,$$

where w are wages, Y_r is the required education to perform a given job, $Y_o = \max(Y_a - Y_r, 0)$ and $Y_u = \max(Y_r - Y_a, 0)$, where Y_a is the actual education (measured in years), while X

[4] ORU is an abbreviation of Over-, Required, Undereducation, reflecting the structure of the model.

[5] Alternatively, studies have used the model with over- and undereducation dummies; this model is usually attributed to Verdugo and Verdugo (1989).

contains all other relevant controls. Studies using ORU generally agree that $\beta_r > \beta_o > 0 > \beta_u$ (Rubb, 2003a; see also Hartog, 2000). For the overeducated, this means higher wages than for the well-matched *in the same job* ($\beta_o > 0$), while lower than for the well-matched *at the same education level* ($\beta_r > \beta_o$). On the contrary, the undereducated earn less than the well-matched *in the same job* ($\beta_u < 0$) but more than the well-matched *at the same education level* ($\beta_r + \beta_o > 0$). The relationship between the three coefficients continues to hold even if instead of the whole market, a single firm is analysed (Groeneveld & Hartog, 2004).

The standard models of overeducation effects on wages, however, can lead to biased results if individual ability is not controlled for. Thus, several studies reviewed the consistency of the above-mentioned result when the model is re-structured to control for unobserved heterogeneity. Using panel data methods, it is found that this decreases wage differences substantially (Bauer, 2002; Carroll & Tani, 2013; Frenette, 2004; Korpi & Tåhlin, 2009; Tsai, 2010; Verhaest & Omey, 2012), but the relationship between coefficients found in standard ORU models still holds (Korpi & Tåhlin, 2009). In some cases, controlling for heterogeneity did not lead to a substantial reduction in the wage effect of mismatch (Carroll & Tani, 2013; Mavromaras, McGuinness, O'Leary, Sloane, & Fok, 2010).

Several studies use information on wages to proxy individual ability, the underlying assumption being that a higher wage reflects higher unobserved ability. Approximating ability by residuals in a wage equation and comparing the returns on education for individuals with different ability, Chevalier (2003) found that the wage penalty for low-ability overeducated is more than twice higher than that for high-ability overeducated.[6] Other studies approximated ability using individual’s position in wage distribution. Results are mixed. Budría (2011) found that wage effects are more pronounced at upper segments of wage distribution (hence, high-ability workers are hurt more), while McGuinness and Bennett (2007) report that the wage impact is quite pervasive and constant along the whole wage distribution of females, while only low- and middle-ability, but not high-ability, males face wage penalties.

At the same time, the overeducated were not found to enjoy a higher wage growth than the well-matched (Korpi & Tåhlin, 2009; Groeneveld & Hartog, 2004, have the same result for the labour market external to the firm). Taking into account a high persistence of overeducation over time (Baert, Cockx, & Verhaest, 2013; Blázquez & Budría, 2012; Diem & Wolter, 2014; Frei & Sousa-Poza, 2012; Kiersztyn, 2013; Rubb, 2003b), one can conclude that the “overeducated are

[6] His model setup will be reviewed more closely in Section 2.1.4.

penalized early on by an inferior rate of return to schooling, from which they do not recover" (Korpi & Tåhlin, 2009, p. 183).

Regarding other labour-market consequences, overeducation was found to boost career movement (Groeneveld & Hartog, 2004), upward mobility (Dekker, de Grip, & Heijke, 2002), wage growth in firm's internal labour market (Groeneveld & Hartog, 2004) and individual productivity (Büchel, 2002). The first two results are in line with career mobility theory (Sicherman, 1991), which states that more educated individuals are more likely to move up the career ladder.

However, if overeducated graduates do not see enough career opportunities, they become uncommitted to their workplace (Blenkinsopp & Scurry, 2007), and evidence shows that the overeducated are more likely to engage in job search (Wald, 2005). Overeducated graduates are generally less satisfied with their jobs (Sánchez-Sánchez & McGuiness, 2015). Not only does overeducation affect general job satisfaction, but it also decreases different types of job satisfaction, such as extrinsic, intrinsic and social (Peiró, Agut, & Grau, 2010). The overeducated also report much more symptoms of depression than the well-matched (Bracke, Pattyn, & Von dem Knesebeck, 2013). These negative effects were earlier interpreted as "work-related deprivation associated with unfulfilled expectations" (Johnson & Johnson, 2000, p. 551). Undereducation, however, was not found to affect job satisfaction (Peiró, Agut, & Grau, 2010).

Agut, Peiró and Grau (2009) study whether overeducation changes the tendency of employees to engage in extra-role behaviours. According to the social exchange theory (Blau, 1964), if employees perceive a balance between the inputs to the job and the outputs received, they reciprocate by engaging in behaviours that go beyond what is determined by contract. Specifically, Agut, Peiró and Grau study employee engagement into two types of extra-role behaviours. The first is job content innovation, "the development of new work procedures and methods for performing job tasks efficiently" (p. 161); the second is career-enhancing strategies, when "individuals develop work objectives and plans, seek advice and information from others about training or work assignments, or develop skills by doing varied job assignments" (p. 162) to increase their individual career success. In line with theory, perhaps because the overeducated perceive that they are treated unfairly, it is found that they decrease participation in both job-content innovation and career-enhancing strategies, compared to the well-matched.

2.1.3 Measures of Overeducation

Generally, the concept of skill mismatch means a discrepancy between skills possessed by an individual and skills required by the job. In particular, the concept of overeducation (undereducation) means having more (less) education than required by the employer. The measurement of this concept, though, proved to be quite controversial. As noted by the ILO, because "skills and competencies *per se* are not measured by the regular statistical programmes of most countries," in empirical studies "skill proxies are used, such as qualifications, years of schooling and occupations" (International Labour Organisation, 2013a, p. 24, italics added). Table 2-1 describes four approaches used in the literature. Each of them has its own advantages and disadvantages, and the literature has not agreed on a single "correct" measure.

In addition to these traditional measures, some advanced indicators have been proposed recently. For instance, Betti, D'Agostino and Neri (2011) create an indicator based on six work-related dimensions, use factor analysis to identify latent factors and then employ fuzzy logics to assign individuals to the overeducated group based on their performance on these factors.

As already seen in Tables 1-3 and 1-4, different measures can lead to strongly different incidence rates. Naturally, this also leads to differences in model estimates in which overeducation is used. For instance, Chevalier and Lindley (2009) show that genuinely and apparently overeducated have large differences in reasons for accepting their jobs. Analysing the sensitivity of estimation results of regressions modelling the reasons of overeducation to different measurements, Verhaest and Omey (2010), thus, recommend using both objective and subjective measures in empirical research.

2.1.4 Factors Influencing Overeducation

Five broad categories of factors have been found to affect the risk of overeducation: (1) ability, academic performance and personality; (2) sex and age; (3) immigrant background; (4) job and labour market characteristics; and (5) characteristics of individual's education. These are reviewed in turn.

2.1.4.1 Ability, Academic Performance and Personality

A diploma *per se* never guarantees a good knowledge, skills or efficiency of its holder. Thus, measures of individual's ability or academic performance, ideally, should be included in models of overeducation. Unfortunately, the former is not directly measurable, while the latter rarely is found in data. However, scarce research available on this topic shows that "poor-quality"

Table 2-1 – Measures of Overeducation

Idea	Advantages	Disadvantages	Examples of Studies
Normative Measure			
Use a pre-determined mapping between the job and the required education level [a]	• Easily measurable • Objective	• Assumes constant mappings over all jobs of a given occupation • Creating and updating a thorough mapping is costly	Chevalier (2003) Sutherland (2012)
Statistical Measure			
The overeducated are those with education level higher by some ad-hoc value than the mean or mode of the sample within a given occupation [b]	• Easily measurable • Objective • No updating needed: always corresponds to the sample	• Assumes constant mappings over all jobs of a given occupation • Sensitive to cohort effects • Results depend on the level of aggregation of occupations	Fernández & Ortega (2008) Jauhiainen (2011) Ortiz & Kucel (2008)
Self-assessment			
The respondent is asked about their perceptions of the extent their education or skills are used in their job [c]	• Always up-to-date • Corresponds with requirements in the local firm	• Subjective bias: respondents may overstate job requirements, inflate their status or reproduce actual hiring standards	Di Pietro (2002) Frenette (2004) Støren & Wiers-Jenssen (2010) Wirz & Atukeren (2005)
Income-ratio			
Overeducation is a *continuous* variable measured by comparing actual and potential income [d]	• Reflects that a goal of investment into education is maximising income	• An indirect measure, can be influenced by many other factors	Jensen, Gartner, & Rässler (2010) Guironnet & Peypoch (2007)

Source: Developed by the author based on Hartog (2000), Quintini (2011)

[a] The mapping is typically defined by the Dictionary of Occupational Titles (US-specific) or ISCO.

[b] Usually, the ad-hoc value is one standard deviation, but studies also use 80th percentile of the distribution as a cut-off value.

[c] These measures are based not only on a single question about perceived overeducation or skill underutilisation, but also on indices comprised of several such questions (see, e.g., Barone & Ortiz, 2010).

[d] This approach actually connects overeducation to another failure in the labour market – underpayment.

graduates do face a higher risk of overeducation (Barone & Ortiz, 2010; Chevalier, 2003; Verhaest & Omey, 2010).

In this regard, the work of Chevalier (2003), who studies a sample of recent graduates aged under 25, is frequently cited. He uses the ISCO-based criterion to define the overeducated. Then, based on a self-assessment question, he defines two subtypes: *apparently overeducated*, who have similar unobserved skills as the well-matched and, hence, who have high chances to move to a well-matched job over time; and *genuinely overeducated*, who have a much lower skill endowment and, hence, will find it much more difficult to get a well-matched job. Unobservable skills are measured by residuals in a regression with log-wages as a dependent variable (the approach, thus, resembles approximating ability via position in the wage distribution). The overeducation model with these residuals as an independent variable shows that the unobserved skills of the genuinely overeducated are substantially worse than other graduates'.

Sometimes, ability is approximated by parental education. In this case, it was also shown that higher parental education (Barone & Ortiz, 2010) and, in particular, that of individual's father (Lianos, Asteriou, & Agiomirgianakis, 2004; Verhaest & Omey, 2010), corresponds to lower overeducation risk.

There are few studies on how personality traits affect the likelihood of mismatch. Blázquez and Budría (2012) show that individuals with high conscientiousness, extraversion, external locus of control and low openness to experience face lower probability of entering overeducation or remaining in that state.

2.1.4.2 Sex and Age

Empirical evidence about the effect from sex has been mixed, with approximately equal number of studies concluding that women have a higher skills mismatch risk than men (Aleksynska & Tritah, 2013; Baert, Cockx, & Verhaest, 2013; Betti, D'Agostino, & Neri, 2011; Karakaya, Plasman, & Rycx, 2007; Ramos & Sanromá, 2013; Tani, 2012; Verhaest & Omey, 2010; Verhaest & Van der Velden, 2013) as those finding no difference across sex (Blázquez & Budría, 2012; Büchel & van Ham, 2003; Chevalier, 2003; Chevalier & Lindley, 2009; Frei & Sousa-Poza, 2012; Frenette, 2004; Støren & Wiers-Jenssen, 2010; Wirz & Atukeren, 2005); a few studies result in men being at a relative disadvantage (DG Employment, Social Affairs and Inclusion, 2012; Kiersztyn, 2013).[7] This first of these results is mainly attributed to women's traditional role in family. Indeed, it was found that having a small kid substantially increases woman's exposure to the overeducation risk (Barone & Ortiz, 2010; Wirz & Atukeren, 2005). Another possible explanation is that fields of study are still frequently divided into "male" and "female" ones, with "female" fields (like arts and humanities) being more exposed to overeducation, as shown below.

The literature disagrees on the effect on overeducation from age (or labour market experience, which differs from age approximately by a constant). Some studies show that overeducation decreases with age (Aleksynska & Tritah, 2013; Jensen, Gartner, & Rässler, 2010; Robst, 2008; Sutherland, 2012), while others report that age is irrelevant (Blázquez & Budría, 2012; Chevalier & Lindley, 2009; Frei & Sousa-Poza, 2012; Kiersztyn, 2013; Wirz & Atukeren,

[7] See also Quintini (2011) and DG Employment, Social Affairs and Inclusion (2012, p. 371, Footnote 17).

2005). In the model of Beckhusen, Florax, Poot and Waldorf (2013), the age effect has different signs, depending on the level of education of respondents. Overall, these results imply that the young may still have a comparatively higher probability of mismatch after controlling for other relevant factors.

2.1.4.3 Immigrant Background

In general, studies consistently show that immigrants face a higher risk of overeducation and undereducation (Aleksynska & Tritah, 2013). In the literature on immigrant mismatch, attention has been brought to two issues: how the likelihood of mismatch changes with residence duration and how the experiences of the immigrant in the origin country affect their risk of mismatch in the destination country.

In Europe in general, residence duration has no effect on the exposure to overeducation or undereducation (Aleksynska & Tritah, 2013). A similar result has been shown for Spain, where labour force participation and unemployment rates of immigrants converge in five years to those of the main population, but overeducation rate stays the same (Fernández & Ortega, 2008). In Norway, however, overeducation decreases with time since migration, but this is explained by immigrants preferring unemployment to overeducation (Støren & Wiers-Jenssen, 2010). Similarly, in Australia (Tani, 2012) and the US (Beckhusen, Florax, Poot, & Waldorf, 2013), the duration of stay improves the likelihood of match, but in the latter case only immigrants with doctoral and professional, but not master's, education are affected.

The general situation in the origin country, as well as the characteristics of the immigrant's experience there, has a major influence on their experience in the destination country. Immigrating to a country with a common border with the origin country decreases the chances of overeducation (Aleksynska & Tritah, 2013), as this allows immigrants to be intrinsically closer to the local labour force. The same effect comes from having more prior knowledge about the target country (Tani, 2012). Migrating from a country with generally higher-quality education increases the chances of undereducation (Aleksynska & Tritah, 2013). However, being mismatched in the origin country considerably increases the chances of the same type of mismatch in the target country (Piracha, Tani, & Vadean, 2012; Tani, 2012), although the size of this effect diminishes in importance with the labour market experience in the target country, since employers put more weight on recent signals about the individual's productivity (Piracha, Tani, & Vadean, 2012).

2.1.4.4 Job and Labour Market Characteristics

Finding a good match in a larger labour market should be easier. Indeed, in European countries, the risk of overeducation in big cities is substantially smaller than in small cities and, especially, in rural areas (Ramos & Sanromá, 2013). However, in the US, living in a metropolitan area increases the chances of overeducation, as workers unable to find a skilled job in non-metropolitan areas move there and decide to stay (Beckhusen, Florax, Poot, & Waldorf, 2013).

When comparing gender effects in labour markets of different size, the differential overeducation theory (Frank, 1978) is often cited. In essence, it claims that, because of traditional female role in family, the difference between the risk of overeducation women face in rural and urban areas should be higher than what men face. In the European context, this theory was tested in Germany and Spain. In neither case is it unanimously supported. In Germany, married women in rural areas face the highest risk of overeducation when compared to married women in urban areas, single women and men (Büchel & Battu, 2003). Spatial flexibility controls, however, show inconsistent effects for German samples. On the one hand, having a car for personal use and a higher commuting distance to work decrease the overeducation risk by providing access to larger labour markets, where finding a well-matched job is easier; but on the other, a higher travelling time to work increases this risk (Büchel & Battu, 2003; Büchel & van Ham, 2003). In Spain, the hypothesis is supported, unless region-specific controls are added (Ramos & Sanromá, 2013).

Studies also found that overeducation risk decreases with tenure (Büchel & van Ham, 2003; Frei & Sousa-Poza, 2012; Jensen, Gartner, & Rässler, 2010; Karakaya, Plasman, & Rycx, 2007; Wirz & Atukeren, 2005), which allows to conclude that "employers view labour market experience as a substitute for formal education" (Karakaya, Plasman, & Rycx, 2007, p. 513).

Several studies included macro-level factors in mismatch models. Overall, overeducation was found to behave counter-cyclically: the highly educated crowd out the lower educated during economic downturns (Croce & Ghignoni, 2012; Kiersztyn, 2013). Aleksynska and Tritah (2013) report that the overall level of shadow economy does not influence the mismatch likelihood of the locally born, but decreases the chances of overeducation of immigrants, as larger shadow economy eases the job search process for the lower skilled. Croce and Ghignoni (2012) show that higher share of temporary contracts increases overeducation via lowering the selectivity of employers and workers, higher long-term unemployment decreases it by keeping the less-able workers out of labour force, while employment protection legislation does not affect it[8]. Nevertheless, the joint effect of individual-specific supply factors on mismatch likelihood is higher than the joint effect of macro-level demand factors (Ghignoni & Verashchagina, 2014).

2.1.4.5 Characteristics of Individual's Education

To analyse the situation within higher education, the signalling theory (Spence, 1973) has proved to be useful. In Spence's terminology, signals are the attributes of the individuals that are changeable. Because job applicant's productivity is unobservable, individuals change signals to reflect their productivity, and firms then use these signals in employee selection processes.

Traditionally, the level of education was used as a signal. For instance, university graduates are viewed as more productive than secondary school graduates. Taking into account massive expansion of higher education undergoing in Europe in the last decade, one can argue that employers were forced to use not only vertical differentiation (by education level) in candidate screening, but also add horizontal differentiation (by field of study (FOS)). At the same time, the growth in the demand for higher education has been highly uneven across fields, with few young people choosing generally more demanding fields such as engineering or natural science. However, it is perhaps not the high supply of graduates in fields like social science or economics *per se* that decreases their signal strength (or quality). A more important role is played by the fact that the expansion of higher education was partly allowed by the reduction in academic and social selection into universities (Convert, 2005). Hence, it is the lower average quality of graduates choosing to study in less demanding fields that decreases their chances of success in the labour market.

Indeed, research shows that economics, law and arts and humanities lead to higher overeducation risk (Barone & Ortiz, 2010; Betti, D'Agostino, & Neri, 2011; Chevalier, 2003; Cutillo & Di Pietro, 2006; Jauhiainen, 2011; Ortiz & Kucel, 2008; Støren & Wiers-Jenssen, 2010). This, however, does not mean that there are no peculiarities inside countries. In Finland, for instance, the effect from graduating in medicine increases the risk of overeducation of women but decreases it for men (Jauhiainen, 2011). Similarly, in the UK, graduates in medicine and business are most likely to become overeducated, while graduating in law and education protects against it (Chevalier & Lindley, 2009). In Spain, the risk is higher for graduates in teaching than in the

[8] Employment protection legislation increases undereducation, however (Aleksynska & Tritah, 2013).

social sciences, business and law field group, but the risk for graduates in science and mathematics is the same as in the latter field group (Ortiz & Kucel, 2008).

Studies differ on the sign of the effect from educational attainment (in terms of years or level). While some find that the risk decreases with attainment (Barone & Ortiz, 2010; Büchel & van Ham, 2003; Jensen, Gartner, & Rässler, 2010), others find that higher years of education (Fernández & Ortega, 2008; Jauhiainen, 2011) or a university degree (Frei & Sousa-Poza, 2012; Wirz & Atukeren, 2005) increase the risk of overeducation. In particular, Frenette (2004) reports that in Canada, masters face a higher overeducation risk than bachelors. There is a consensus among researchers, however, on a very low risk faced by PhD graduates.

Di Pietro (2002) studies the phenomenon of overeducation at macro-level, taking the incidence of overeducation in a country as a dependent variable. He found that the risk of overeducation on a particular level of education increases with the share of population having that education.

Støren and Wiers-Jenssen (2010) study how the overeducation risk of immigrants and Norwegians with foreign education compares to that of individuals educated in Norway. Generally, they find that foreign education is not perfectly transferrable to the local labour market, as the risk of mismatch is higher than with a local education.

Research that does not exclude current students from the sample shows that working in parallel to studies increases the exposure to overeducation risk (Barone & Ortiz, 2010).

2.2 Social Networks: Role in the Labour–Education Market System

> *Educational choice is an outcome deeply intertwined with prior social choices negotiated through structures of constraints and possibilities*
>
> *(Moniarou-Papaconstantinou, Tsatsaroni, Katsis, & Koulaidis, 2010, p. 323)*

In this text, a ***social network*** means the same as a ***social graph***; the former term is simply more prevalent in the literature than the latter. By social networks, I do *not* mean Facebook, Twitter or any other electronic resource for maintaining a network of contacts and exchanging information therein. Rather, I look on social networks in their broadest meaning – as an interconnected set of individuals, where connections may manifest themselves as electronic exchange of information, face-to-face conversations or simply knowing each other.

Graph type: undirected
Vertices: $V = \{a, b, c, d, e\}$
Edges: $E = \{(a, b), (a, c), (a, d), (d, e)\}$
Degrees: $\deg a = 3; \deg d = 2; \deg b = \deg c = \deg e = 1$

Graph type: directed
Vertices: $V = \{a, b, c, d, e\}$
Edges: $E = \{(a, b), (b, a), (a, c), (a, d), (d, e)\}$
Out-degrees: $\deg_+ a = 3; \deg_+ b = \deg_+ d = 1; \deg_+ c = \deg_+ e = 0$
In-degrees: $\deg_- a = \deg_- b = \deg_- c = \deg_- d = \deg_- e = 1$

Figure 2-1 – Example of Undirected and Directed Graphs and Their Main Statistics

As there will be several references to the terminology specific to graph theory, the following subsection gives a short introduction into that. Afterwards, I review the literature on building and changing a social network. Subsections 2.2.3 and 2.2.4 analyse the literature on social network effects in the labour market and in the education market, respectively. The latter subsection also discusses other important factors in making decisions on education.

2.2.1 A Short Note on the Terminology of Graph Theory

A ***graph*** is an ordered pair $G = (V, E)$, where V is a set of vertices and E is a collection of edges. In an ***undirected graph***, an ***edge*** is an unordered pair of vertices $(v, w): v \in V \wedge w \in V$. By definition, in this case, edges are not directed; thus, an edge (v, w) is the same edge as (w, v). In a ***directed graph***, an ***edge*** (sometimes called an ***arc***) is an ordered pair of vertices, so that the arc (v, w) has the reverse direction than the arc (w, v).

Typically, and in all cases in this text, E is a set, meaning that two vertices v and w can either be disconnected (i.e., not connected with any edge) or connected with a single edge (v, w). This applies to both undirected and directed graphs: in the former case, we are talking about a single *unordered* pair (v, w), while in the latter, we are talking about a single *ordered* pair (v, w).

The ***degree*** of a vertex equals the number of edges it is ***incident*** to or, equivalently in our case, the number of vertices it is connected to: $\deg v \equiv |\{w: (v, w) \in E\}|$. In a directed graph, the corresponding terms are ***in-degree***, the number of incoming edges, and ***out-degree***, the number of outgoing edges. Figure 2-1 provides illustrative examples.

2.2.2 Creating, Maintaining and Dissolving Social Ties

Two decisions are to be made with respect to creating, maintaining and dissolving social ties. The first of them is “Whom should I add to my social circle?” with a related question “Whom should I drop from my social circle in the first place?" Theoretically, the answer to both questions is captured by ***homophily*** – "the principle that a contact between similar people occurs at a higher rate than among dissimilar people" (McPherson, Smith-Lovin, & Cook, 2001, p. 416).[9] Because of homophily, personal networks tend to be quite homogeneous on a large number of parameters characterising the individual. Moreover, as summarised by McPherson, Smith-Lovin and Cook (2001), homophily is found in many types of personal networks ranging from marriage to friendship to simply knowing someone. They report that such segregation of society is going by socio-demographic dimensions and at the level of values, beliefs and attitudes, which appear in the following order of decreasing importance:

1. Race and race-like ethnicity
2. Sex, age, religion and level of education
3. Occupation, network position, behaviours and intrapersonal values

In much of the social network modelling and simulation literature (at least that pertaining to labour market), agents representing persons are homogeneous. Only in the model of Abdou and Gilbert (2009), agents belong to distinct groups and are biased in the formation of their network to members of their group. Nevertheless, the idea of maintaining social links mainly with similar agents appears even in models where there are no differences in the *internal* characteristics of the agents (in terms of ability or group membership). In that case, agents discriminate against each other based on economic status, an *external* characteristic. For instance, in Gemkow and Neugart (2011), agents create friendship ties with the employed with a higher probability and dissolve them with a lower probability than with the unemployed. In Bramoullé and Saint-Paul (2010), on the contrary, tie dissolution is a purely random process, while tie creation between agents with the same labour market status is more probable than between an employed and an unemployed.

The second important decision to be made is "How many social ties should I have?" Theoretically, as noted by Granovetter (2005, p. 34), "people have cognitive, emotional, spatial and temporal limits on how many social ties they can sustain." Thus, maintaining a particular number of friends has an inherent cost for a person. Some models where social network evolved endogenously, however, did not impose such a cost, rather defining a Markov process allowing for accumulation of any number of friends (Bramoullé & Saint-Paul, 2010; Tassier & Menczer,

---

[9] A similar principle is ***assortative mating*** – "preference for sexual partners with similar phenotypes, as when humans select mates who tend to resemble themselves in wealth, social class, intelligence or attractiveness, or when other organisms choose mates that resemble themselves on specific criteria" (Colman, 2009).

2001). Other authors imposed a constraint on the maximum number of friends an agent could have as an exogenous constant equal for all agents[10] (Krauth, 2004) or as a normally distributed random characteristic of the agent (Abdou & Gilbert, 2009).

Gemkov and Neugart (2011) go beyond setting the ceiling on the number of friends and then allowing ties to create and dissolve in a purely stochastic way. They introduce agents that adaptively learn their optimal number of friends. Technically, they use ***experience-weighted attraction***, whereby agents choose among strategies, each strategy representing some number of friends. Every strategy has an attraction, which is updated in every model period either according to the payoff it provided if it was chosen in the previous period or according to a fraction of the payoff it would have provided had it been chosen. These payoffs do take into account specific costs of maintaining the corresponding number of friends. Attractions are then depreciated and normalised, taking into account a measure of experience. Finally, they are mapped into the probability space of actual strategies.

While setting up the model without any cap placed on the number of friends contradicts both theory and what is empirically observed, one can hardly believe that a good model of heterogeneous agents could assume that such a cap is the same for everyone. Rather, one should take into account how the degrees of vertices are distributed in real social networks.

Many complex networks, of which social networks are an example, tend to possess the ***scale-free property***, meaning that degrees in these networks are distributed according to the ***power law***, whereby the probability that vertex v has degree k decreases with k:

$$\Pr(\deg v = k) = ck^{-\beta},$$

where c is a normalisation constant and, typically, $2 < \beta < 3$ (Barabási & Bonabeau, 2003). Kumar, Novak and Tomkins (2010) analyse Flickr and Yahoo! 360, online social networks mainly used for sharing photos with friends. They find that in both cases, the degree distribution follows the power law. Kwak, Lee, Park and Moon (2010) find that Twitter, a micro-blogging service, has the number of followers distributed according to the power law until 10^5 followers, after where it departs from this distribution, but this effect has been attributed to a large number of celebrities and large companies using Twitter. In contrast, Cyworld, a popular social network service in South Korea, exhibits the power law only starting from degrees around 10^3, while lower degrees are distributed exponentially (Ahn, Han, Kwak, Moon, & Jeong, 2007). This large number of ties present in virtual social networks does not, however, mean that users actually

[10] Tassier and Menczer (2008) experimented with evolving social networks where the number of friends *per se* was fixed at some constant-for-all level, which is a special case of fixing the maximum number of friends.

communicate with all their "virtual friends." Indeed, it was found that on Facebook, 70 per cent of interactions are made with only 20 per cent of the user's friends; still, both the nominal social network declared on Facebook and the network users actually interact with show the power law distribution of vertex degrees (Wilson, Boe, Sala, Puttaswamy, & Zhao, 2009).

One of the algorithms that can be used to generate a scale-free network is ***Duplication model*** $\mathcal{D}(\rho)$ described by Chung and Lu (2006, Ch. 4). This model depends on parameter ρ and proceeds in two-step iterations. At the first step, ***vertex duplication***, a new vertex is added to the graph and connected randomly with an existing vertex. At the second step, ***edge duplication***, the algorithm goes over each neighbour of the existing vertex and connects it to the new vertex with probability ρ. Thus, ρ is the probability that the new agent creates a connection with a friend of its new friend. It was shown that the model generates scale-free networks with the exponent β that is a solution to the following equation (Chung & Lu, 2006, p. 78):

$$1 + \rho = \rho\beta + \rho^{\beta-1}. \tag{2-1}$$

Thus, in case of using Duplication model, one should first determine the target value of β (or its target interval) and then use (2-1) to determine the value of ρ to be used in the model.

However, there is also evidence that real social networks do not follow the power law strictly and degree distribution could be more accurately described by a combination of a scale-free and a random network (Jackson, 2008). Toivonen, et al. (2009) shows that the mutual friendship network on a Finnish web service is better approximated by a lognormal distribution. Networks of e-mail addressees, in turn, are better represented by an exponential distribution (Adamic & Adar, 2005; Guimerà, Danon, Díez-Guilera, Giralt, & Arenas, 2003).

2.2.3 Referral Hiring and Information Diffusion

The importance of social networks in the labour market was repeatedly shown both empirically and theoretically. Social networks affect the actions of both firms and individuals at different stages of the employment process.

Research shows that in different countries, 30 to 60 per cent of companies hire by employee referral. In the US, 60 per cent of firms use personal contact networks to find job candidates, where in most cases, this meant employee referrals (Bewley, 1999). In Latvia, networking is the most popular recruitment method for enterprises (depending on language used in enterprises, 30 to 50 per cent of them hire by referral), but the intensity of systematic use of social networks decreases with firm size (Hazans, 2011b). Actually, in all Eastern European countries, a usual way of finding and hiring for vacancies is through informal channels (relatives, friends, acquaintances), especially in the small and medium enterprise sector (Kuddo, 2009). This phenomenon is known as ***referral hiring***.

Employees also use their social networks in the process of job search. Montgomery (1991) cites several studies reporting that around 50 per cent of employees in the US found their jobs through friends and relatives. My estimates from Estonian Labour Force Survey data are that every year during 2001–2009, 30 per cent of respondents asked relatives and friends as their most important step taken to find a job.[11] Thus, information on vacancies disseminates not only through formal channels, but also through the personal contact network.

Granovetter (2005) mentions two reasons why social networks are so much used in the hiring process. Firstly, they help mitigate the problem of ***bilateral asymmetric information***, when both prospective employers and employees do not know the other side's quality. In these settings, they search for more information about one another from personal sources they trust. Secondly, the cost of searching for a new employee in existing social networks, which are maintained mainly for non-economic reasons, is far lower than using the formal channels. One could argue that existing employees may inflate the real qualifications of the friend they recommend, but this would contradict their long-term interest in the company. Therefore, using referral hiring is a theoretically clean way of reducing costs.

Related to this discussion is Granovetter's (2005) ***strength of weak ties hypothesis***, which states that acquaintances outside the clique of close friends are a source of more novel information that is unavailable from friends. In particular, the "weak ties" help individuals find a new job in case of unemployment easier, because they become aware of the information that circulates outside their typical social circle of "strong ties." This conjecture found support in several theoretical studies (Calvó-Armengol & Jackson, 2004; Krauth, 2004).

2.2.4 Choices Related to Higher Education

In this subsection, I will consider literature on how choices related to higher education are made. There are at least three such choices: (1) whether to continue studies at tertiary level, (2) what HEI to choose and (3) what field to study in. The first and third choices attracted major attention in the academic community. They are also particularly relevant, given the expansion of higher education itself and its unevenness across fields of study, as was described in Section 1.2.

[11] Another 30% mentioned watching job ads, 15% directly contacting employers and 15% found it most helpful to seek a new job through the state employment office.

However, if a secondary school graduate is to make a decision on their future studies, they should make all these three choices.

In studying the tertiary-education related decisions of individuals, we, in effect, study the motivations behind investing in education. Once we understand what motivates individuals to start making these choices, we will get much closer to the precise factors that shape their actual choices. Are individuals motivated by monetary returns on education? Then they will look at the levels, HEIs and fields that lead them to lucrative positions in terms of salary. Are they motivated by prestige? Then they will enter highly ranked HEIs and study in the fields associated with high status. Are they motivated by power? Then they will seek the education that enhances their power.

From all studies on motivation, reviewing which is far beyond the scope of this text, one can conclude that human beings are highly complex. They almost never make choices based on some simple single factor, even if they claim so. In particular, individuals are likely to engage in thinking about higher education, taking into account three groups of factors: (1) personality, (2) social networks and (3) monetary costs and benefits. The literature supports it and gives a vast number of examples of factors in each of these groups. I will review them below. But to reiterate on the above-mentioned, factors from all three groups will always be taken into account, whatever the person actually claims the reasons are behind his or her actions.

2.2.4.1 Personality, Ability and Other Individual Aspects

Considerable research has been done on the ***personality–job fit theory***. The underlying idea is quite simple. Individuals have different personalities, there are different types of jobs, and individuals will be more satisfied and less likely to quit if they work in jobs whose types are congruent with their personalities (Robbins, Judge, & Campbell, 2010).

The most well-known and tested theory is Holland's (1959) six types of personalities and environments (e.g., jobs or fields of study); see Table 2-2 for details. These types can be arranged in a hexagon (see Figure 2-2), where two given types correspond to each other better if they appear closer in the hexagon. For instance, realistic persons feel most comfortable in realistic environments (same vertex) and least comfortable in social environments (two vertices away).

Empirical research corroborates that students achieve better results in fields that are congruent with their personalities (Hackett & Lent, 1992; Holland J. L., 1997; Smart, Feldman, & Ethington, 2000). Nevertheless, students in a field of study (FOS) from a given Holland type gain the same amount of knowledge from studies, irrespective of their own Holland type; any

Table 2-2 – Description of Holland's Personality Types and Related Fields of Study

Personality Type	Description	Example Fields of Study
Realistic	• Prefer activities involving the manipulation of materials, tools or machines • Avoid educational and interpersonal activities • Practical, conservative, persistent, conforming, shy • Lack social and educational competencies • Value material rewards	• Electrical Engineering • Mechanical Engineering • Military Science • Marine Science
Investigative	• Prefer exploration, understanding and prediction • Avoid persuasive, social and repetitive activities • Analytical, critical, independent, precise, rational • Lack persuasive and leadership competencies • Value knowledge acquisition and scholarly achievements in science and technology	• Biology • Science (Astronomy, Chemistry, Maths, Physics, Geography) • Economics • Finance
Artistic	• Prefer ambiguous, free and non-systematised activities to create art forms of products • Avoid routine and conformity to rules • Expressive, original, intuitive, independent, emotional, sensitive • Lack clerical and business system competencies • Value aesthetic qualities and creative expression of ideas, emotions or sentiments	• Fine Arts (Art, Drama, Music) • Languages • Architecture
Social	• Prefer to inform, teach, cure or enlighten others • Avoid explicit, systematic activities involving materials, tools or machines • Cooperative, empathetic, generous, helpful, sociable, possessing leadership competency • Not mechanically inclined • Value education and social service	• Humanities (History, Philosophy, Religion) • Nursing • Psychology • Political Science
Enterprising	• Prefer persuading and directing others to attain organisational goals or economic gain • Avoid scientific, intellectual and abstruse activities • Aggressive, ambitious, energetic, have leadership ability, self-confident • Lack scientific ability • Value material accomplishment and social status	• Business Management • Journalism • Computer Science
Conventional	• Prefer explicit, ordered and systematic manipulation of data • Avoid ambiguous and unstructured activities • Careful, conforming, orderly, have clerical and numerical ability, efficient, inflexible • Lack artistic competencies • Value material accomplishment and power	• Accounting • Data Processing

Source: adapted from Pike (2006), Robbins, Judge and Campbell (2010, p. 97) and Smart, Feldman and Ethington (2000, pp. 38-39, 59-60)

differences in the level of knowledge after graduation, thus, should be attributable to the level before starting studies, rather than to the ability to absorb new knowledge (Feldman, Ethington, & Smart, 2001). Moreover, students base their FOS choice not only on their Holland type; they also tend to expect strengthening their knowledge in a way rather different from their personality type (Pike, 2006).

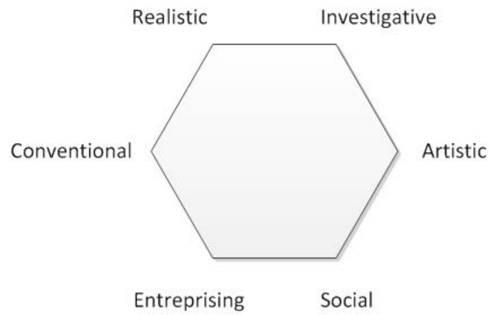

Figure 2-2 – Holland's Personality Types

Source: Pike (2006)

Currently, Holland's six types are part of Strong Interest Inventory under the name General Occupational Themes (GOTs). However, because it divides the whole universe of possible interests and attractions in only six types, its accuracy in classifying individuals into fields of study or occupations is not as strong as one would like it to be. On the contrary, Basic Interest Scales (BISs) (Campbell, Borgen, Eastes, Johansson, & Peterson, 1968) enhance this classification (Ralston, Borgen, Rottinghaus, & Donnay, 2004) due to considering much narrower types (in total, around 30). In general, it is concluded, Holland types "are very useful for guiding the big picture in career assessment," but BISs provide "potent complementary information" for more specific analysis (Ralston, Borgen, Rottinghaus, & Donnay, 2004, p. 215).

There are also studies using many other classifications of personality types (e.g., see Rubinstein, 2005, for applying the Big Five model; Figueira, et al., 2010, for applying the TEMPS-A model), reviewing which goes beyond the scope of this text. The main point is that there is a vast amount of empirical proofs that personality is heavily related to the choice of FOS.

Studies also show that ability in key skills related to a particular field of study affects the probability of choosing that field due to its effect on the probability of success in that field (Tolsma, Need, & De Jong, 2010; Van de Werfhorst, Sullivan, & Cheung, 2003). For instance, students good at reading at the age of 11 tend to study social sciences, arts and humanities, while those good at maths at the same age tend to study medicine, law, natural science and engineering at university (Van de Werfhorst, Sullivan, & Cheung, 2003).

Sex also plays an important role in choosing educational path. For reasons of culture, tradition or any other[12], the important role of "female" (e.g., arts, humanities and education) and "male" (e.g., science and technology) fields of study is very well documented in various countries

[12] While risking to be called a sexist, I would still claim that some narrow professions are more suited for one concrete sex rather than equally suited to both sexes.

(Boudarbat, 2008; Cai, 2003; Dar & Getz, 2007; Goyette & Mullen, 2006; Pike, 2006; Van de Werfhorst, Sullivan, & Cheung, 2003).

2.2.4.2 Social Environment: Family, Social Class and Social Networks

One of the most important factors affecting individuals' choices on education is their social environment. It may be called family (or parental) effects, social (or socio–economic) class or social network effects, but the essence remains the same: individuals build their social networks according to the homophily principle and then use these social networks in decision-making.

The effects from social networks have been discussed in two perspectives. One of them is ***rational action theory*** (Boudon, 1974; Breen & Goldthorpe, 1997). While, as a typical theory assuming rational behaviour of agents, it is built around individuals making cost–benefit analyses before making their choices, it does not downgrade the effects from social origin. For instance, Breen and Goldthorpe (1997) build three mechanisms of social origin effects in their model:

- **Relative risk aversion:** Families seek to ensure that their children end up in the social class at least as high as from which they originate. In other words, they seek to avoid downward social mobility.
- **Primary effects:** Ability (which is related to social background) determines the likelihood of success at different educational tracks. Thus, children have to have a certain minimum ability to qualify for continuing education.
- **Availability of economic resources:** Education is not costless, and certain minimum resources of families (the availability of which differs by social class) are needed for children to continue education.

Another theoretical perspective is ***habitus theory*** or ***social reproduction theory*** (Bourdieu, 1984; Bourdieu & Passeron, 1990). According to it, cultural and economic capital of a social class shapes the choices of individuals belonging to that class due to different cognitive and normative predispositions (referred to as ***habitus***) deeply ingrained in members of a particular social class. There are two different elites – economic and cultural – which differ in what capital is available to them and what matters to them. The economic elite uses economic wealth and related lifestyle and values to position itself in the social space and increases the economic capital available to its members. The cultural elite seeks to enhance its cultural capital by living a different lifestyle and using different values.

Whatever the explanation, the conclusion is that there should be a considerable intergenerational transmission of education level and field of study. Empirical literature finds strong evidence of that.

Several studies found direct parental effects on FOS choice, from both the education level and the FOS of parents. In Canada for an average child, tertiary education of father increases the chances of choosing to study education, arts and humanities, while that of mother decreases the probability of choosing business and commerce (Boudarbat, 2008). Dividing children by sex, male offsprings of tertiary-educated fathers are more likely to study medicine and female offsprings – arts & humanities and agriculture & biology. Tertiary education of mothers induces male children to choose agriculture and biology and female children to choose medicine, while impeding the latter to choose education (Boudarbat & Montmarquette, 2009). In the Netherlands, it was shown that children are likely to choose the same FOS as their father (Van de Werfhorst, De Graaf, & Kraaykamp, 2001).

Other studies supported the significance of social class or socio–economic status (SES). For the Netherlands, the relative risk aversion mechanism of the rational action theory was empirically confirmed (Tolsma, Need, & De Jong, 2010). Moreover, the habitus theory is also supported. In particular, it was found that cultural and economic elites are inter-generationally reproduced by the FOS of children: children in the economic elite are more likely to choose the financially lucrative fields of economics and law, children in the cultural elite – cultural fields of study, while low-SES children look more on labour market prospects of their fields, rather than on their intellectual or aesthetic outcomes (Van de Werfhorst, De Graaf, & Kraaykamp, 2001). Longitudinal data from the UK also support the habitus theory, showing that children of the professional social class are more likely to choose medicine and law, while a higher social capital increases chances to study in arts and humanities, thus, reproducing the family's stock of cultural capital (Van de Werfhorst, Sullivan, & Cheung, 2003). US data show that individuals with higher SES favour arts and science fields (humanities, science and social science) over vocational fields (business, education, engineering and other occupation-oriented fields), corroborating the accumulation of social capital in terms of access to exclusive social networks in families of higher SES (Goyette & Mullen, 2006).

Dar and Getz (2007) find that in Israel, SES plays different roles for individuals of low and high ability in choosing the type of HEI and educational track (the combination of the type of HEI and FOS). While higher SES helps *low*-ability students to enter a more prestigious institution, in track placement SES differentiates more the choices of *high*-ability students: applicants characterised by both high SES and high ability choose scientific fields, while high-ability students of low SES choose fields that are more practical. As a result, it is concluded, “factors of *habitus* associated with the family and the social environment, rather than institutional selection,” (p. 57) lead to the self-exclusion of highly able students of low SES from dominant elites.

Vossensteyn (2005) finds that SES has a substantial effect on educational choices through affecting the perceptions of financial incentives. Thus, university entrants of low SES perceive higher education as a risky investment. Coupled with lower future income expectations, it is no surprise they are willing to borrow a smaller amount than high-SES entrants for investing into higher education. Being constrained in income, they are more attracted by grants and scholarships, are more involved in working while studying and choose fields with shorter durations and those perceived as less difficult.

Moniarou-Papaconstantinou, Tsatsaroni, Katsis and Koulaidis (2010) take a broader view and claim that individual choices of higher education are shaped by complex interactions among social factors (social class and available capital), socially-shaped educational experiences and continuous interactions with social environment. They find that motivations of choosing a particular FOS (in their case, library and information science) depend heavily on the levels of these three factors. Individuals with lower socio–cultural background and restricted education experience are extrinsically motivated, while those with middle socio–cultural background are intrinsically motivated, which confirms the results mentioned in previous paragraphs. Moreover, they find, people high in socio–cultural background have a socially-acquired ability to see what may be promising for their future careers.

Taking a dynamic perspective, Reimer and Pollak (2010) do not find students from more privileged social backgrounds *increasingly* deciding to choose either more prestigious vertical (education levels) or horizontal (fields of study) educational alternatives: social background effects are robust but stable.

The research cited in this subsection studied observable effects from general social environment. Interestingly, when asked directly about using their social networks in their decisions on education, individuals do not always admit that they base decisions on their network, rather than making them themselves. As an extreme finding, Kalvans and Salietis (2009) found that opinions of friends and experts have no effect on student choices in Latvia. The self-reported influence of non-family members of social network is rather weak: 13.3% in Romania, compared to 31.2% of family members (Pomazan, Mihalaşcu, Petcu, & Gîrtu, 2010), 28% in Portugal, compared to 80% of academic reputation (Simões & Soares, 2010) and the weakest factor in the Netherlands (Van Deuren & Santema, 2012). At the same time, members of social networks, especially former students, are used as a valuable source of information about a study programme (Simões & Soares, 2010; Van Deuren & Santema, 2012). Nevertheless, these findings should be used with considerable care, as any self-reported results.

2.2.4.3 Monetary Costs and Benefits

If one had to name a single theme that runs like a thread through all chapters of a typical book on economics, one would name money. Rational human beings should make the cost–benefit analysis before making any decision and make only those decisions that appear profitable, it is postulated. Does it sound real in the process of choosing an educational path?

Analysts of US data reply with a resounding “yes”: students’ perceptions of labour market outcomes of their education have a large impact on FOS choice (Hu, 1996; Montmarquette, Cannings, & Mahseredjian, 2002). As one might expect, the probability to choose a particular FOS rises in response to increasing future monetary returns (Freeman & Hirsch, 2008). Moreover, students believe that their parents are more likely to approve their studies in fields with not only a higher social status, but also with larger labour-market returns (Zafar, 2012).

Studies in other countries also support the finding that students react on monetary labour market signals (Boudarbat, 2008; Boudarbat & Montmarquette, 2009; Kalvans & Salietis, 2009; Pomazan, Mihalaşcu, Petcu, & Gîrtu, 2010). However, there are several cautionary notes about the limitations of the effects from earnings on applicants’ choices. Firstly, a *substantial* increase in lifetime earnings is necessary to make individuals choose a FOS they were not inclined to choose initially (Boudarbat & Montmarquette, 2009). Secondly, in assessing monetary costs and benefits, students take a short-term perspective: study loans and grants are more important factors in their decision-making than expected future income (Vossensteyn, 2005). Thirdly, if an individual was already employed before studies, he or she is less sensitive to differences in earnings across fields (Boudarbat, 2008).

Naturally, if earnings are important, then the probability of actually getting these earnings – finding a job after graduation – should also play an important role, and again, there is positive empirical evidence on that (Kalvans & Salietis, 2009; Pomazan, Mihalaşcu, Petcu, & Gîrtu, 2010; Zafar, 2012).

Boudarbat (2008) finds that students of business & commerce and sciences are most sensitive to earnings, while students of social sciences are least sensitive. Freeman and Hirsch (2008) report positive effect from earnings on choosing engineering and technology, mathematics, physics, medicine, English and law; and negative effect on choosing food production. In particular, they find no effects from earnings on choosing computers and electronics as field of study. Goyette and Mullen (2006) report students choosing vocational fields over arts and science (arts and humanities, social science, and science) if money or steady employment is important for them. Thus, studies are consistent in showing that students *for whom money is important* tend to choose fields with more stable and solid earnings opportunities, what Goyette and Mullen (2006) call degrees "with high use value."

2.3 Mechanisms for Finding and Quitting Job in Existing Agent-Based Models of the Labour Market

Whether or not modelling the referral hiring mechanism, the key events included in any labour-market model are finding and losing a job. In many cases, both events are probabilistic, but many advanced models also include a vacancy mechanism, whereby firms publish vacancies and individuals apply for them. This section reviews the mechanisms used in the literature to model the choice among vacancies by individuals, the choice among applicants by firms and the decision to quit by individuals. As this analysis will be used in Chapter 5, here I concentrate on the literature that uses the agent-based modelling approach (see Appendix D for more information about it).

2.3.1 Choice among Vacancies

The choice among vacancies has two sub-choices: which vacancies to apply for and which offer (application for a vacancy accepted by the firm) to accept.

The simplest case to model both choices is to allow firms to pick from the unemployed who passively wait (Abdou & Gilbert, 2009; Takács, Squazzoni, & Bravo, 2012). This is equivalent to an agent applying to all vacancies but receiving the decision to be accepted from no more than one firm, so that the agent does not have to choose from firms after applying.

There are models where agents apply for all vacancies (Gemkow & Neugart, 2011) and the variations of it: all vacancies in their neighbourhood (Martin & Neugart, 2009), all vacancies firms probabilistically inform them about (Tassier & Menczer, 2008), all vacancies they find through social networks (Lewkovicz, Thiriot, & Caillou, 2011; Tassier & Menczer, 2008; Thiriot, Lewkovicz, Caillou, & Kant, 2011) if the proposed wage exceeds reservation wage (Dawid & Gemkow, 2014) or all vacancies in the industries matching their investments in human capital (Neugart, 2008). A variant of that is also applying to a random set of firms (Dawid & Gemkow, 2014), because the number of applications is not bound from above.

On the other side of the spectrum are models where agents select only one vacancy to apply for or one firm where to search for vacancies. Agents may apply for a single vacancy with the maximum utility exceeding the reservation utility (Ballot, 2002; Ballot, Kant, & Goudet, 2013), choose to search for vacancies in a single firm chosen probabilistically (Fagiolo, Dosi, & Gabriele, 2004) or firstly contact the last employer and choose a random firm only if the last employer does not have vacancies corresponding to the type of the agent (Silva, Valente, & Teixeira, 2012). One could also include in this group studies where agents do not search for vacancies, but are informed about them by firms always (Ng & Kang, 2013) or probabilistically (Tassier & Menczer, 2001) or by their friends (Tassier & Menczer, 2001). In both models, agents apply for the vacancy with the highest wage, but in Tassier and Menczer (2001), employed agents apply only if it is higher than the wage they currently earn.

Somewhere in the middle are models that put a cap on the number of vacancies an agent could apply for. Typically, these are randomly selected vacancies (Richiardi, 2006) with wages exceeding the reservation wage (Dawid, et al., 2008; Dawid, Gemkow, Harting, & Neugart, 2009; Dawid, Gemkow, Harting, & Neugart, 2012; Deissenberg, van der Hoog, & Dawid, 2008). Although not modelling vacancies as such, agents in Tesfatsion (2001) send work offers to a number of firms (limited from above) with highest non-negative utility, which puts this model in this middle category.

In case of receiving several offers, agents typically decide based on the proposed wage only: they choose the offer with the highest wage (Dawid & Gemkow, 2014; Deissenberg, van der Hoog, & Dawid, 2008), sometimes taking into account commuting costs (Dawid, et al., 2008; Dawid, Gemkow, Harting, & Neugart, 2009; Dawid, Gemkow, Harting, & Neugart, 2012), if the proposed wage exceeds the reservation wage (Fagiolo, Dosi, & Gabriele, 2004; Richiardi, 2006; Silva, Valente, & Teixeira, 2012). In Neugart (2008), however, agents accept the first job offer they receive.

2.3.2 Choice among Applicants

Along this dimension, studies can be divided into three groups.

In studies from the first group, firms choose the applicant on the first-applied-first-accepted basis (Ng & Kang, 2013; Tassier & Menczer, 2001) or through purely random choice (Martin & Neugart, 2009; Neugart, 2008; Tassier & Menczer, 2008), in some models not accepting the applicants they have just laid off (Lewkovicz, Thiriot, & Caillou, 2011; Thiriot, Lewkovicz, Caillou, & Kant, 2011). Some studies do not explicitly state whether the first or a random applicant is chosen (Fagiolo, Dosi, & Gabriele, 2004; Silva, Valente, & Teixeira, 2012), but it is otherwise clear that one of these mechanisms should be used: the firms there just propose a wage for a vacancy and accept whoever agrees with that wage offer.

In studies from the second group, firms classify applicants into few distinct groups, preferring some groups to others, and then choose agents inside the non-empty group they prefer the most. Models using referral hiring typically fit in this category. In Abdou and Gilbert (2009), firms probabilistically choose whether to use formal channels or hire through referral. In the former case, they choose a random applicant, while when hiring through referrals, the firm has a higher probability of choosing the agent of the same type as it has itself. In Gemkow and Neugart (2011), firms first choose randomly from the applicants having friends among its employees, and in Dawid and Gemkow (2014), instead of random choice agents with the highest general skills are chosen. Firms in Takács, Squazzoni and Bravo (2012) use a five-step procedure, moving to the next step if the previous one failed to fill the vacancy: (1) re-hire the previous employee with the highest quality if his quality meets standards and exceeds group reputation scores; (2) randomly hire an employed friend *of the firm*; (3) hire the previous employee of a business partner with the highest quality if his quality meets standards and exceeds group reputation scores; (4) hire a friend of recently employed workers randomly or in the order of the quality of referents; and (5) randomly hire an unemployed from the group with higher reputation.

Studies from the third group model an explicit ranking mechanism firms use to sort the list of applicants and then choose the applicant with the highest rank. Studies then differ with respect to which ranking mechanism they use. In some, firms use general skill level (Dawid, Gemkow, Harting, & Neugart, 2012; Dawid, Harting, & Neugart, 2013; Deissenberg, van der Hoog, & Dawid, 2008), and if agents have the same general skill level they are sorted by specific skill level (Dawid, et al., 2008; Dawid, Gemkow, Harting, & Neugart, 2009). In others, firms choose the applicant with the highest expected profit above the minimum tolerated profitability of the vacancy (Ballot, 2002; Ballot, Kant, & Goudet, 2013). The firms in Richiardi (2006) employ the most productive applicant. Finally, while vacancies are not explicitly modelled in Tesfatsion (2001), firms keep the waiting list of applicants with the highest non-negative utility, the length of the list being limited from above, and update the list after every wave of applicants; firms employ everyone from the list if they receive no more applications, thus, employing the applicants with the highest utility.

2.3.3 Decision to Quit

In most studies, decision to quit is not modelled and the separation of working relationships occurs only through the firing mechanism. Other studies may be divided into those modelling it purely probabilistically and those defining a more elaborate decision function of the agent.

In the former group of studies, the probability is set globally (Dawid, Harting, & Neugart, 2013) or depends on agent's homophily level and workplace composition in terms of agent types (Abdou & Gilbert, 2009). The model in Tesfatsion (2001) should also be put in this group: while it assumes that agents quit because their employer did not provide them the benefits they expected to receive, this behaviour of firms is governed by playing a prisoner's dilemma game.

In the latter group, the agents that quit are high-quality workers who are not promoted for a long time or have the utility of quitting that exceeds the utility of continuing work (Ballot, 2002; Ballot, Kant, & Goudet, 2013).

2.4 Summary

The literature review showed that:

- Many theories assume that overeducation is a real phenomenon and present different explanations for it
- Overeducation has strong negative effects: lower wages, lower wage growth, lack of commitment to workplaces without enough career opportunities, likeliness to engage in job search and lower job satisfaction
- Many overeducation measures exist, each with own benefits and drawbacks
- Overeducation is affected by (1) ability, academic performance and personality, (2) sex and age, (3) immigrant background, (4) job and labour market characteristics and (5) the characteristics of individual's education
- Social networks are built using the principle of homophily and tend to possess the scale-free property
- Social networks are extensively used in recruiting and searching for jobs
- Individuals decide on their education based on its fit with their personality, ability, social environment and monetary factors
- Existing agent-based models of the labour market, even if include referral hiring or job search through social networks, impose a simplistic choice mechanism among vacancies on their agents. According to it, agents either accept a random vacancy or

choose the best vacancy only based on the proposed wage (but not on non-wage factors)

It also showed a need for:

- A broad study on the causes of overeducation, taking into account nearly non-studied effects from personality
- Labour-market simulation models with
    - Heterogeneous vacancies
    - More realistic vacancy choice mechanism
    - Heterogeneous agents with complex social network building mechanisms

3 FACTORS AFFECTING OVEREDUCATION

This chapter studies what factors influence whether an individual becomes overeducated. The analysis is typically done at an aggregate European or country-group level, but some country-level particularities are also noted, as necessary.

The chapter has the following structure. Through it, I use a single data source, European Social Survey (ESS) data. Section 3.1 introduces the data. Section 3.2 defines the overeducation variable that will be used throughout this and the next chapters, as well as discusses its benefits and drawbacks. ESS data include information on the personality of respondents. Section 3.3 shows that this information is valuable in overeducation models. Section 3.4 presents the general model of overeducation. For some years, ESS data includes information on the field of study at tertiary level. Section 3.5 uses this information in studying the relationships between field of study and the risk of overeducation. Finally, Section 3.6 discusses how the exposure to overeducation differs by industry. The last section concludes.

3.1 European Social Survey Data

In all further sections of this chapter and some sections of Chapters 4 and 5, I use European Social Survey (ESS) data, rounds 1 through 5 (Norwegian Social Science Data Services, 2002; 2004; 2006; 2008; 2010). This is a biennial survey covering over 30 countries. Country coverage differs by round, with only 16 countries surveyed in all five rounds (see Table 3-1).

One of the particularities of ESS data, as compared to, for instance, Eurostat's Labour Force Survey, is relatively small samples for each of the participating countries (between 579 and 3031 observations, depending on country and round). Hence, especially if studying subgroups of populations, it is necessary to pool country-level data together to have reliable sample sizes.

In this case, it is of utmost importance to ensure that educational attainment data, which will be used in the definition of overeducation (see Section 3.2), match across countries. Fortunately, the producers of ESS data made significant improvements in the comparability of the data in their latest updates to all five rounds. Appendix A describes the comparability problems and actions I performed to further enhance the quality of the variable measuring education level.

Table 3-1 – Country Coverage in ESS Data, by Round

Country	Round 1 (2002–2003)	Round 2 (2004–2005)	Round 3 (2006–2007)	Round 4 (2008–2009)	Round 5 (2010–2011)
Austria	X	X	X	X	
Belgium	X	X	X	X	X
Bulgaria			X	X	X
Croatia				X	X
Cyprus			X	X	X
Czech Rep.	X	X		X	X
Denmark	X	X	X	X	X
Estonia		X	X	X	X
Finland	X	X	X	X	X
France	X	X	X	X	X
Germany	X	X	X	X	X
Greece	X	X		X	X
Hungary	X	X	X	X	X
Iceland		X			
Ireland	X	X	X	X	X
Israel	X			X	X
Italy	X	X			
Latvia			X	X	
Lithuania				X	X
Luxembourg	X	X			
Netherlands	X	X	X	X	X
Norway	X	X	X	X	X
Poland	X	X	X	X	X
Portugal	X	X	X	X	X
Romania			X	X	
Russia			X	X	X
Slovakia		X	X	X	X
Slovenia	X	X	X	X	X
Spain	X	X	X	X	X
Sweden	X	X	X	X	X
Switzerland	X	X	X	X	X
Turkey		X		X	
UK	X	X	X	X	X
Ukraine		X	X	X	X

3.2 ISCO-Based Measure of Overeducation: Definition, Benefits and Drawbacks

In this text, overeducation is defined by the ISCO-based measure. Based on first-digit ISCO levels, I divide occupations into five broad groups (see Table 3-2). Leaving out the military group[13], I then assign a relevant education level to each broad group. Individuals working in a given broad occupation group and having the assigned level of education are considered well matched. Those having a higher level of education are then overeducated. Thus, for instance, a

[13] Firstly, those employed in the armed forces are, by definition, skilled for that work. Secondly, the number of observations in this group is always low. Thirdly, military people are generally out of interest for labour market analysis.

Table 3-2 – Correspondence between ISCO Major Groups and Education Levels

ISCO Major Group	Broad Occupation Group	Relevant Education Level
0: Armed forces	Military	
1: Legislators, senior officials, managers 2: Professionals 3: Technicians and associate professionals	High-skilled non-manual	Tertiary (ISCED 5–6)
4: Clerks 5: Service workers, shop, market sales workers	Low-skilled non-manual	Secondary (ISCED 3–4)
6: Skilled agricultural and fishery workers 7: Craft and related trades workers 8: Plant and machine operators and assemblers	Skilled manual	
9: Elementary occupations	Unskilled	Primary (ISCED 1–2)

university graduate working as a clerk (a low-skilled non-manual occupation) is considered overeducated.

It is important to note that education levels are defined using ISCED groups (see the last column of Table 3-2). Thus, secondary level also includes post-secondary non-tertiary education (ISCED 4), while tertiary level also includes short-cycle tertiary and vocational tertiary programmes (ISCED 5B).

The largest benefit of the ISCO-based measure is that the definition of who is overeducated does not change over time and, hence, the results are comparable over time (of course, if the definition of education levels is stable over time).

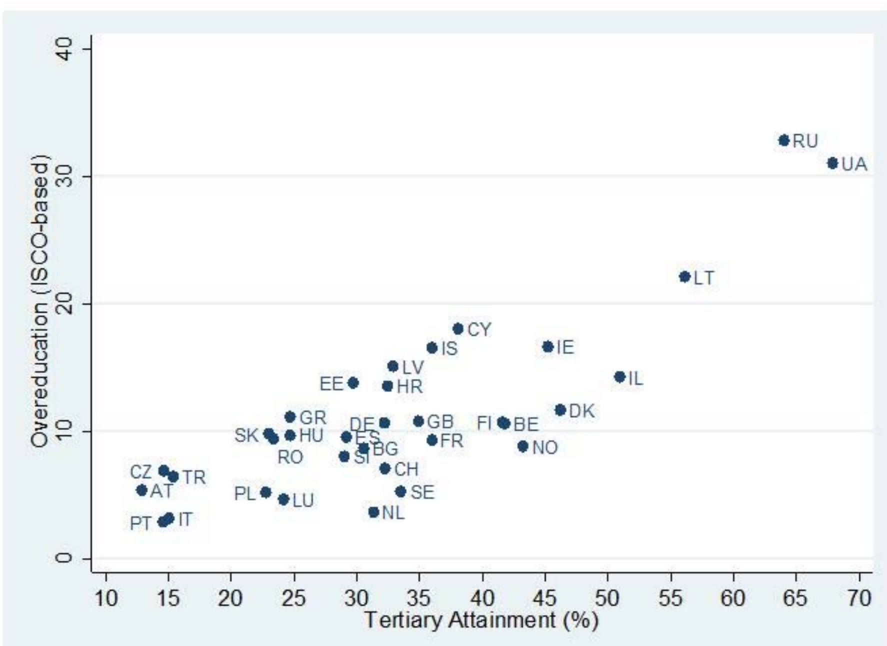

Figure 3-1 – Relationship between Tertiary Attainment and ISCO-based Overeducation

Source: compiled based on ESS data

Data on overeducation incidence and tertiary attainment first computed for each country and for each ESS round, then averaged over ESS rounds. Tertiary attainment defined as the share of respondents with tertiary education from respondents with known level of education.

A downside of it is that it does not take into account the actual distribution of educational attainment. Therefore, in high-attainment countries, the proportion of the overeducated might be higher (see Figure 3-1). Another disadvantage of this measure is that, by definition, it does not allow to measure overeducation in the top three major ISCO groups, which gives an inherent downward bias in the overall proportion of overeducation in comparison with alternative methods. Finally, because the ISCO-based method reflects educational attainment levels, the mismatch incidences for men and women are different, as the latter are usually concentrated in a narrower range of occupations.

3.3 Unobserved Heterogeneity in Overeducation Models: Personality vs. Ability

In a perfect labour market, overeducation should not be observed, as workers should always be employed in the jobs matching their level of education or, more generally, skills. Nevertheless, statistics show that overeducation does exist (see Section 1.2). It might, however, be that we observe overeducation simply because we cannot accurately measure the quality of match between an individual and a job. In particular, there might be some unobserved individual- or job-specific factors that, together with the observed factors, make the worker a perfect candidate for his or her job. Not controlling for such unobserved heterogeneity in mismatch-related models, thus, introduces a bias on the results of these models.

I already reviewed how different studies model unobserved heterogeneity in Section 2.1.2. To summarise, it is (1) controlled for by choosing an appropriate statistical model if panel data are available, or (2) approximated by approximating ability or splitting the sample into more homogeneous subsamples, or (3) controlled for by adding the environment where the individual was raised to the model.

Another block of potentially important factors residing in unobserved heterogeneity is individual's personality. For instance, it was shown that personality, as represented by the Five Factor Model (McCrae & John, 1992) (also known as the Big Five) affects job satisfaction, performance, intention to quit and turnover (Zimmerman, 2008).

Nevertheless, to the best of my knowledge, there is only one study that includes personality aspects in overeducation models (Blázquez & Budría, 2012). One of the reasons might be that variables reflecting relevant individual behaviour, beliefs and values are not available in survey datasets. However, another important reason is that personality, being a complex concept not readily pluggable into the mathematical framework of economics, is simply overlooked by economists. Indeed, economic theories, such as human capital theory or signalling theory, typically focus on productivity, for which ability is the key internal (i.e., unobserved) driver.

In this section, I will compare the performance of selected personality aspects and ability (as approximated by the difference between expected and actual income) on explaining the overeducation status of the individual.

The structure of this section is as follows. Section 3.3.1 describes the data and methods. Then Section 3.3.2 presents the results. Section 3.3.3 concludes.

3.3.1 Data and Methods

I use ESS round 5. Only rounds 2 and 5 of ESS contain the tenure variable, which is an important predictor of overeducation (see Section 2.1.4.4). I choose to focus on round 5 only so as not to introduce additional issues regarding homogeneity.

From the whole set of 27 countries, I take 23 (Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Ireland, Israel, Lithuania, the Netherlands, Norway, Poland, Slovakia, Slovenia, Spain, Sweden, Switzerland and the UK). Estonia and Portugal were removed because they do not contain some of the key variables employed in the models used here; Russia and Ukraine were removed because they are quite different from other countries.

In general, I follow the idea of Chevalier (2003), who proxies individuals' abilities with the residuals in the regression of wages and then feeds these residuals into the regression of overeducation. Because the structure of ESS data is quite different, alternative methods have to be used.

In Chevalier (2003), the former model used data on individual income. However, ESS does not have such individual-level data. Instead, its respondents state the total income of their household from all sources. In addition, there is a question about the share of this total income provided by the respondent. Because there are relatively few observations where the majority of household income was due to the respondent, I take a broader view and consider two samples: (1) respondents provide at least half of the total household income and (2) they provide most or all income of the household. While it is straightforward that residuals from the wage regression on sample (2) could be used to approximate individual's ability, the use of sample (1) deserves more explanation. In using it, I follow the logic of the homophily principle (McPherson, Smith-Lovin, & Cook, 2001), according to which people form ties (e.g., friendship or marriage) with those similar to them by a subset of characteristics, including race, ethnicity, religion, as well as education and social class. In particular, I assume that the level of respondent's ability could be approximated with the abilities of those living with him, i.e., the residuals from the regression on household income.

Another particularity of ESS data on income is that it is given in deciles of national income. For instance, while there are no data on the absolute income of a given household, the dataset shows that this income belongs to the third decile of income distribution in a given country. While this means that regressions of log-wages, as run by Chevalier (2003), cannot be employed here, this national-decile-based reporting of income has a great benefit of being comparable over countries. In particular, this allows me to pool the data on all countries in the dataset without worrying about cross-country differences in income levels.

The methodology I employ in this study is, thus, as follows. Firstly, I run ordered logit regression (see Section C.5) of the decile of household income:

$$\Pr(decile_i = d) = \Pr(k_{d-1} < \mathbf{x}_i\boldsymbol{\beta}_\mathbf{1} + \mathbf{y}_i\boldsymbol{\gamma} + \varepsilon_i \leq k_d). \tag{3-1}$$

Standard errors are estimated by sandwich estimator, assuming zero inter-country correlations but non-zero intra-country correlations of observations (see Section C.7). Vector $\mathbf{x}$ consists of variables that will also be used in the subsequent regression of overeducation (age, immigrant dummy, student dummy, disability dummy, informal employment dummy (defined as working without a contract, see Hazans (2011a)), tenure and country). Vector $\mathbf{y}$ consists of variables appearing only in the income model (age squared, supervisor dummy, firm size and industry (NACE rev.2)).

Based on (3-1), the probabilities of each respondent's household income falling in each decile are predicted. The decile with the highest probability is then identified as the predicted income decile. Denoting the actual income decile by d_i and the predicted income decile by $\hat{d}_i$, ability is measured as:

$$a_i \equiv d_i - \hat{d}_i.$$

Thus, if the respondent's household actual income decile is higher than the predicted income decile, he/she is attributed positive ability. If it is lower, ability is negative. If the prediction is correct, ability is zero. Absolute values of ability are not important; these are relative values of this variable that will matter.

Then I put this ability variable in the logit regression (see Section C.2) of overeducation status:

$$\text{logit}[\Pr(o_i)] = \mathbf{x}_i\boldsymbol{\beta}_\mathbf{2} + \mathbf{z}_i\boldsymbol{\delta} + a_i\xi + \eta_i. \tag{3-2}$$

Standard errors are estimated using the same sandwich estimator as for model (3-1). Vector $\mathbf{z}$ contains variables describing respondent's personality. ESS contains several variables measuring the extent of different personality traits in respondents. Each variable measures respondent's opinion about how much a person described in the question resembles him/her. Based on their significance in overeducation models, I selected the variables shown in Table 3-3. For each of these variables, vector $\mathbf{z}$ includes a dummy reflecting whether the respondent is "very much like" the description.

All models are run on currently employed individuals aged below 65 and having tenure below 42 years (to mitigate coefficient bias due to small fractions of the sample with large values of age and/or tenure). Individuals who did not complete primary education were removed from the sample. Separate models are run for males and females.

I compare the explanatory power of four versions of model (3-2):

1. Base model (without personality and ability variables)
2. Base model with personality variables (BP model)
3. Base model with ability variable (BA model)

Table 3-3 – Personality Variables in ESS Data

Short Name	Original Description
Important to be creative	Thinking up new ideas and being creative is important to him. He likes to do things in his own original way.
Important to be rich	It is important to him to be rich. He wants to have a lot of money and expensive things.
Important to be treated equally	He thinks it is important that every person in the world should be treated equally. He believes everyone should have equal opportunities in life.
Important to show abilities	It is important to him to show his abilities. He wants people to admire what he does.
Important to try new things	He likes surprises and is always looking for new things to do. He thinks it is important to do lots of different things in life.
Important to follow rules	He believes that people should do what they are told. He thinks people should follow rules at all times, even when no one is watching.
Important to make decisions freely	It is important to him to make his own decisions about what he does. He likes to be free and not depend on others.
Important to help people	It is very important to him to help the people around him. He wants to care for their well-being.
Important to seek adventures	He looks for adventures and likes to take risks. He wants to have an exciting life.
Important to get respect	It is important to him to get respect from others. He wants people to do what he says.
Important to be loyal to friends	It is important to him to be loyal to his friends. He wants to devote himself to people close to him.

4. Full model (base model with personality and ability variables)

The explanatory power of a model is measured by:

1. Log-likelihood
2. McFadden's pseudo R^2, which takes into account only log-likelihood, i.e., general quality of fit
3. Akaike (AIC) and Bayesian (BIC) information criteria, which correct for the number of parameters
4. Sensitivity (the share of overeducated respondents correctly classified by the model), and specificity (the share of non-overeducated respondents correctly classified by the model), which gauge the classification capabilities of the model

The cut-off value for computing sensitivity and specificity was set to the average overeducation level in the sample on which the model was run. The predicted probability of overeducation for each respondent was computed after the model, and the respondent was predicted to be overeducated if this probability was greater than the cut-off value.

3.3.2 Results

The results of model (3-2) are shown in Table 3-4 for respondents providing at least half of household income and in Table 3-5 for those providing most of household income.

Regardless of whether their income is at least half or most of household income, males have twice fewer personality factors influencing their chances of overeducation than females have. Moreover, the significant factors rarely overlap across gender. Even when they do overlap, they not necessarily work in the same direction (as in the case of the importance of equal treatment when respondent's income is most of household income).

The general quality of fit, measured by the change in log-likelihood and pseudo R^2, is better in BP models than in BA models for females, regardless of their income's share in household income, and for males bringing most of household income.

Information criteria, however, mainly favour BA models over BP models. The only exception is for females bringing at least half of household income, where AIC favours the BP model over the BA model.

Table 3-4 – Odds Ratios after Logit, Respondent's Income Is At Least Half of Household Income

	Males				Females			
	Base	BP	BA	Full	Base	BP	BA	Full
Age	1.012*	1.011*	1.011*	1.010†	1.008	1.009	1.009	1.010
Immigrant	1.453***	1.435***	1.508***	1.489***	1.943***	1.972***	1.992***	2.023***
Student	1.552	1.586	1.501	1.535	0.626	0.644	0.624	0.630
Disabled	1.124	1.113	1.166	1.153	1.446***	1.490***	1.521***	1.558***
Informal employment	1.246	1.225	1.302†	1.283	1.423*	1.548**	1.535**	1.663**
Tenure	0.973***	0.973***	0.973***	0.973***	0.970***	0.969***	0.968***	0.967***
Imp. to be creative		0.833†		0.844†		0.620***		0.626***
Imp. to be rich								
Imp. to be treated equally		0.843†		0.841†				
Imp. to show abilities						0.801†		0.816
Imp. to try new things						1.258†		1.280*
Imp. to follow rules						0.664†		0.670†
Imp. to make decisions freely								
Imp. to help people		1.263*		1.263*				
Imp. to seek adventures						1.552†		1.550*
Imp. to get respect								
Imp. to be loyal to friends						0.742***		0.740***
Ability			0.934**	0.934**			0.909***	0.912***
Constant	0.081***	0.086***	0.081***	0.085***	0.044***	0.052***	0.042***	0.049***
Quality of Fit								
N	5778				4255			
Pseudo R^2	0.0275	0.0297	0.0307	0.0328	0.0563	0.0691	0.0641	0.0763
Log-likelihood	−1755.88	−1751.84	−1750.11	−1746.22	−1334.96	−1316.87	−1323.93	−1306.73
AIC	3523.76	3521.67	3514.21	3512.44	2683.91	2659.74	2663.85	2641.46
BIC	3563.73	3581.63	3560.85	3579.06	2728.40	2742.37	2714.70	2730.45
Sensitivity	61.4%	60.3%	63.0%	63.0%	65.7%	66.1%	67.7%	67.3%
Specificity	55.9%	56.4%	56.8%	57.2%	57.6%	60.2%	59.1%	60.5%

*** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$ † $p < 0.15$

Odds ratios for country dummies not reported.

Table 3-5 – Odds Ratios after Logit, Respondent's Income Is a Very Large Fraction of Household Income

	Males				Females			
	Base	BP	BA	Full	Base	BP	BA	Full
Age	1.016*	1.017**	1.015*	1.016**	1.011	1.013	1.008	1.010
Immigrant	1.383	1.389	1.409	1.415	1.890***	1.843***	1.926***	1.880***
Student	1.495	1.563	1.406	1.467	0.696	0.701	0.639	0.597
Disabled	0.926	0.943	0.961	0.977	1.433†	1.502*	1.469*	1.553*
Informal employment	1.387	1.417	1.431	1.466	1.839***	2.069***	1.933**	2.178***
Tenure	0.969***	0.967***	0.967***	0.966***	0.973***	0.969***	0.974***	0.971***
Imp. to be creative						0.445***		0.477***
Imp. to be rich						2.140**		2.312**
Imp. to be treated equally		0.779*		0.781*		1.376†		1.363†
Imp. to show abilities								
Imp. to try new things		1.394*		1.396*				
Imp. to follow rules						0.610†		0.607†
Imp. to make decisions freely		0.749*		0.756†				
Imp. to help people						0.672**		0.653**
Imp. to seek adventures								
Imp. to get respect						1.978***		1.952***
Imp. to be loyal to friends						0.676**		0.681**
Ability			0.936*	0.937*			0.808***	0.811***
Constant	0.076***	0.081***	0.076***	0.080***	0.032***	0.038***	0.037***	0.044***
Quality of Fit								
N	2687				1943			
Pseudo R^2	0.0282	0.0328	0.0317	0.0361	0.0763	0.1066	0.1006	0.1296
Log-likelihood	–809.26	–805.44	–806.37	–802.68	–575.29	–556.38	–560.14	–542.07
AIC	1632.53	1630.88	1628.73	1627.36	1164.59	1140.76	1136.29	1114.13
BIC	1673.80	1689.84	1675.90	1692.22	1203.59	1218.77	1180.86	1197.71
Sensitivity	61.8%	62.5%	61.1%	60.7%	72.4%	68.7%	72.8%	72.8%
Specificity	57.0%	57.8%	58.3%	57.8%	55.5%	62.8%	60.4%	62.8%

*** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$ † $p < 0.15$

Odds ratios for country dummies not reported.

Table 3-6 – Quality of Fit of Logit Regressions in Country Groups, Respondent's Income Is At Least Half of Household Income

	Males				Females			
	Base	BP	BA	Full	Base	BP	BA	Full
Eastern Europe	N = 1501				N = 1200			
Pseudo R^2	0.0548	0.0747	0.0551	0.0756	0.0453	0.0817	0.0644	0.0994
Log-likelihood	–378.31	–370.36	–378.19	–369.99	–397.33	–382.17	–389.36	–374.79
AIC	768.63	754.71	770.39	753.97	806.66	778.34	792.72	763.58
BIC	800.51	791.91	807.59	791.17	837.20	813.97	828.35	799.21
Sensitivity	60.2%	58.4%	61.9%	61.1%	54.0%	62.6%	64.0%	61.9%
Specificity	60.5%	64.0%	59.8%	63.8%	61.6%	61.3%	61.2%	62.6%
Western Europe	N = 2163				N = 1565			
Pseudo R^2	0.0268	0.0303	0.0429	0.0461	0.0675	0.0912	0.0859	0.1067
Log-likelihood	–754.30	–751.65	–741.85	–739.40	–433.24	–422.25	–424.69	–415.06
AIC	1520.60	1515.31	1495.71	1490.81	878.49	856.49	861.39	842.12
BIC	1554.67	1549.38	1529.78	1524.89	910.62	888.63	893.52	874.25
Sensitivity	62.6%	59.3%	64.0%	64.0%	75.2%	68.5%	72.5%	72.5%
Specificity	57.0%	58.5%	60.7%	61.4%	53.7%	60.7%	58.3%	61.9%
Northern Europe	N = 1313				N = 878			
Pseudo R^2	0.0388	0.0480	0.0764	0.0838	0.0410	0.0497	0.0712	0.0787
Log-likelihood	–310.50	–307.52	–298.34	–295.97	–236.70	–234.54	–229.23	–227.40
AIC	629.01	623.04	604.69	599.94	479.40	475.09	464.46	460.80
BIC	649.73	643.76	625.41	620.66	493.74	489.42	478.79	475.13
Sensitivity	56.6%	59.0%	68.0%	67.2%	57.4%	59.6%	64.9%	66.0%
Specificity	63.4%	64.1%	66.0%	66.2%	60.6%	61.5%	64.9%	64.9%
Southern Europe	N = 801				N = 612			
Pseudo R^2	0.0049	0.0062	0.0098	0.0112	0.0544	0.0742	0.0570	0.0765
Log-likelihood	–296.97	–296.59	–295.52	–295.10	–259.03	–253.60	–258.30	–252.96
AIC	599.95	599.19	597.04	596.20	526.05	515.21	524.60	513.93
BIC	614.00	613.24	611.10	610.25	543.72	532.87	542.27	531.94
Sensitivity	40.7%	49.6%	46.9%	51.3%	58.4%	62.8%	58.4%	66.4%
Specificity	68.8%	58.1%	56.5%	58.4%	63.9%	59.9%	60.9%	60.7%

Significant personality variables for males in Eastern Europe are: important to show abilities (–), help people (+), seek adventures (–) and get respect (–); in Western Europe: to be rich (+) and treated equally (–); in Northern Europe: to be creative (–), try new things (+), follow rules (+) and help people (+); in Southern Europe: to be treated equally (–). For females in Eastern Europe: to be creative (–), be treated equally (+), follow rules (–), seek adventures (+) and be loyal to friends (–); in Western Europe: to show abilities (–), try new things (+), make decisions freely (–), seek adventures (+) and be loyal to friends (–); in Northern Europe: to be creative (–); in Southern Europe: to be creative (–), be treated equally (+), follow rules (–) and be loyal to friends (–).

Table 3-7 – Quality of Fit of Logit Regressions in Country Groups, Respondent's Income Is a Very Large Fraction of Household Income

	Males				Females			
	Base	BP	BA	Full	Base	BP	BA	Full
Eastern Europe	N = 584				N = 481			
Pseudo R^2	0.1158	0.1406	0.1355	0.1550	0.0575	0.1121	0.0852	0.1438
Log-likelihood	–131.25	–127.57	–128.32	–125.42	–145.15	–136.75	–140.89	–131.87
AIC	274.49	267.14	270.65	264.84	306.31	289.50	297.78	279.74
BIC	300.71	299.72	301.23	295.43	339.71	322.91	331.19	313.15
Sensitivity	52.2%	54.3%	65.2%	63.0%	56.1%	63.2%	64.9%	66.7%
Specificity	74.5%	72.7%	71.4%	70.4%	63.9%	64.2%	64.2%	65.3%
Western Europe	N = 1114				N = 786			
Pseudo R^2	0.0283	0.0428	0.0536	0.0677	0.0982	0.1311	0.1114	0.1412
Log-likelihood	–380.05	–374.37	–370.16	–364.63	–204.89	–197.41	–201.89	–195.12
AIC	774.09	762.73	754.32	743.26	421.78	406.83	415.77	402.24
BIC	809.20	797.84	789.43	778.37	449.78	434.83	443.78	430.25
Sensitivity	62.5%	63.4%	68.8%	67.0%	81.8%	72.7%	78.4%	75.0%
Specificity	56.0%	59.2%	58.5%	61.7%	50.0%	59.0%	53.4%	61.3%
Northern Europe	N = 521				N = 370			
Pseudo R^2	0.0672		0.1163		0.0768	0.1421	0.1059	0.1706
Log-likelihood	–132.96		–125.96		–98.87	–91.88	–95.75	–88.82
AIC	273.92		259.93		205.73	191.75	199.50	185.64
BIC	290.94		276.95		221.39	207.41	215.15	201.29
Sensitivity	60.0%		67.3%		58.5%	61.0%	68.3%	70.7%
Specificity	63.5%		69.1%		58.7%	69.3%	66.0%	69.6%
Southern Europe	N = 468				N = 306			
Pseudo R^2	0.0104	0.0177	0.0104	0.0178	0.0762	0.1682	0.0967	0.1955
Log-likelihood	–147.35	–146.27	–147.35	–146.26	–119.68	–107.76	–117.03	–104.22
AIC	300.69	298.54	300.69	298.51	247.35	223.52	242.06	216.44
BIC	313.14	310.99	313.14	310.96	262.25	238.42	256.96	231.33
Sensitivity	56.5%	48.4%	56.5%	48.4%	66.7%	70.2%	70.2%	77.2%
Specificity	50.2%	50.7%	50.2%	51.5%	65.9%	66.7%	63.1%	72.7%

Significant personality variables for males in Eastern Europe are: important to show abilities (–); in Western Europe: to be treated equally (–), to show abilities (+) and help people (–); in Southern Europe: to be creative (+), be treated equally (–), show abilities (–), try new things (+) and be loyal to friends (–). For females in Eastern Europe: to show abilities (–), follow rules (–) and seek adventures (+); in Western Europe: to be creative (–), show abilities (–), try new things (+), make decisions freely (–) and help people (–); in Northern Europe: to be creative (–), be rich (+), be treated equally (+), try new things (+), make decisions freely (+), help people (–) and seek adventures (–); in Southern Europe: to be creative (–), be rich (+), show abilities (+), follow rules (–), get respect (+) and be loyal to friends (–). No personality variables are significant for males in Northern Europe. Ordered logit estimated by BFGS algorithm for females in Southern Europe, as Stata's modified Newton–Raphson algorithm did not converge.

This means that while BP models allow for a better fit than BA models, the increase is not enough for the inclusion of so many variables. However, this should not be a surprise. The ability variable was generated as a residual, which, by definition, alone represents all other relevant characteristics not accounted for in the regression of income. On the other hand, each personality variable measures a distinct trait, which, however complex it might be, covers only a part of the whole universe of personality traits influencing the possibility of mismatch. Thus, the increase in the overall quality of fit is more relevant than the decrease in the quality of fit reported by information criteria. Nevertheless, in cases where information criteria favour the BP model, personality factors are so powerful that they are able to beat the ability factor, even taking into account that there are multiple variables representing the former. Unanimously better classification accuracy can be determined only for males bringing at least half of household income. In that case, the BA model leads to both higher sensitivity and specificity than the BP model. In the other three cases, the model performing better in predicting one of the two outcomes performs worse on predicting the other. For males bringing most household income, superior sensitivity is given by the BP model, but the BA model allows for better specificity. For females, regardless of their income's share in household income, the BP model performs better in classifying the non-overeducated, but the BA model on classifying the overeducated.

To find out whether the results will differ when analysing smaller country groups, I divide the countries geographically into Eastern Europe (Bulgaria, Croatia, Czech Republic, Hungary, Lithuania, Poland, Slovakia and Slovenia), Western Europe (Belgium, France, Germany, Ireland, the Netherlands, Switzerland and the UK), Northern Europe (Denmark, Finland, Norway and Sweden), and Southern Europe (Cyprus, Greece, Israel and Spain). Tables 3-6 and 3-7 show the quality-of-fit statistics.

On the country-group level, the best model by the overall quality of fit is also the best by the information criteria. In all four country groups, data for females are better fit by BP models everywhere, except for Northern European females bringing at least half of household income. On the contrary, data for males are better fit with BA models, except for Eastern Europe (regardless of the share in household income) and Southern Europe (when bringing most of household income).

With a single exception of Southern European respondents bringing at least half of household income, BA models are better able to classify the overeducated correctly. Regarding specificity, BP models perform better for respondents bringing most of household income (except for Northern European males, where no personality factor was found significant). Among those contributing to at least half of household income, BP models perform better for Eastern Europeans, Western European females and Southern European males, while BA models have higher correct classification ratios for Northern Europeans, Western European males and Southern European females.

To understand whether it pays off to include both personality and ability factors in models, I will now compare the performance of Full models to that of BP and BA models. Both on the pan-European and country-group levels, Full models lead to higher overall quality of fit than either BP or BA models. In country groups, Full models were also selected as the best by information criteria, while on the pooled sample, AIC favours Full models and BIC – BA models.

Unfortunately, in only half of the cases there are improvements in *both* sensitivity and specificity in Full models over BP and BA models. In most of the other cases, one of these classification indicators (mostly, specificity) is improved in Full models. There are, however, cases when one of these indicators is *worse* in Full models than either in BP or BA models.

3.3.3 Conclusions

This section showed that personality is an important factor affecting the chances of overeducation. Including personality factors allows to better explain the mismatch risk for females, while for males, ability frequently performs better as an explanatory variable. Personality-only models are mostly better in correctly classifying the non-overeducated, while ability-only models – in correctly classifying the overeducated. Researchers should check whether including personality factors pays off in their particular models and samples, but the results of this study show that personality might be an important predictor of individual's state in the labour market and, thus, should be considered when building the model.

Further research on this topic should continue in three non-exclusive directions. Firstly, the effects of personality traits on other measures of overeducation (for instance, those based on mean or median years of education in a given occupation group) should be assessed. Secondly, the performance of personality vs. ability should be assessed on narrower samples (for separate countries, for separate age groups, etc.), which should allow for better sample homogeneity. Thirdly, alternative measures for personality traits (e.g., from the Five Factor Model) and/or for ability could be used.

3.4 General Model of Overeducation Determinants

This section continues the study of overeducation determinants. It was already shown that personality variables are important predictors and it pays off to include them in models. In this section, I include these variables in a larger model covering all countries available in ESS rounds 1 through 5.

3.4.1 Data and Methods

ESS data are in the form of repeated cross-sections: Every round, a cross-section of individuals is surveyed in the participating countries. It can be argued that, because, among other things, labour market policies and education systems affect inhabitants of a country in a similar way, one cannot assume that intra-country observations – in the same round or in different rounds – are uncorrelated. Observations representing different countries, on the other hand, can be assumed to have zero correlation.

To account for this correlation, I use two-level mixed-effects logistic model (see Appendix C for methodological details), where individuals comprise the first level and countries the second. Countries, thus, form clusters of observations. As noted below, random coefficients are added at the country-level, as needed.

Generally, the model looks as

$$\Pr(y_{ik} = 1|\mathbf{u}_k) = \Lambda(\mathbf{x}_{ik}\boldsymbol{\beta} + \mathbf{p}_{ik}\boldsymbol{\gamma} + \mathbf{m}_{tk}\boldsymbol{\delta} + \mathbf{z}_{ik}\mathbf{u}_k).$$

In this equation, there are M clusters (i.e., countries), indexed by k. The dependent binary variable y_{ik} represents the state of overeducation. There are three covariates for the fixed effects (corresponding to the results of the ordinary logistic regression):

- $\mathbf{x}_{ik}$ contains all basic variables
- $\mathbf{p}_{ik}$ contains personality variables
- $\mathbf{m}_{tk}$ contains macro-level variables (note that the first index is t, not i, indicating that the macro-variables are specific to ESS rounds, not to individuals)

Section 3.4.2 describes the explanatory variables in detail. The $1 \times q$ vector $\mathbf{z}_{ik}$ stores the covariates for the random effects, representing both random intercepts and random coefficients, as needed. The meaning of $\mathbf{u}_k$ and $\Lambda(\cdot)$ is the same as in Section C.4.

The three variable vectors are added sequentially in the model, which gives three models:

1. Base model (only vector $\mathbf{x}$ without personality and macro-level variables)
2. Base model with personality variables (BP model, vectors $\mathbf{x}$ and $\mathbf{p}$)

3. Base model with personality and macro-level variables (BPM model, all vectors)

These models are run on the whole sample, as well as separately for men and women. In these three models, country-level random coefficients and sex effects (in the whole sample model only) are added. An additional model is run with a more advanced configuration of random effects:

4. Base model with personality and macro-level variables and country-level random effects of personality variables (BPM-PRE)

This latter model is also run both on the whole sample and by sex.

3.4.2 Explanatory Variables

3.4.2.1 Basic Variables

The basic variables fall into three categories: associated with the respondent, their family and their labour market experience.

I expect that overeducation risk rises with age. To account for possible nonlinearities, I add age-squared to the regression. The effect from sex is expected to be in line with literature findings and Section 1.2 (higher for females than for males). The number of children should increase the exposure to overeducation, as having to care for children could make parents choose any available job and not invest in careful search for a better job.

Living together with a partner should also influence overeducation risk. Partner's employment status is differentiated into unemployed, employed in a non-supervising position and employed in a supervision position. I expect that living with an unemployed partner increases the risk of overeducation (the reason being similar to the case of having children). Living with an employed partner, however, should leave more time to the respondent for search for a qualified job; hence, a decreasing effect on overeducation risk is expected. Finally, the exposure to overeducation is expected to be still lower if the partner has subordinates.

Theoretically, having access to a bigger labour market should increase the probability of a correct match. Hence, I expect that, compared to living in rural areas, where the choice among available jobs is quite limited, living in a small city decreases the probability of both mismatch risks, and living in a big city decreases them further.

The same logic would apply at the firm level: a larger firm (in terms of the number of employees) requires less of generalist and more of specialised workforce than a smaller firm. Hence, less mismatch should be expected in larger organisations.

Immigrant background[14] is expected to increase overeducation risk, as immigrants would tend to be placed in jobs requiring lower-level skills, disregarding their actual qualifications. These effects should be especially pronounced for immigrants from developing countries, but not from highly developed countries like the US or Canada. Regarding non-immigrants, I distinguish among natives, minorities and second-generation immigrants, who are further divided into those having one or both parents-immigrants. I expect the effects from non-immigrant status to be smaller than from immigrant status because of being better integrated into society, although effect directions should be similar.

Other "potentially negative" factors – in the sense that they should increase overeducation – are studying in parallel to work, disability, ever facing a period of 3-month or 12-month unemployment and informal employment defined here as working without a contract (Hazans, 2011a). If an individual ever faced a medium-term or a long-term unemployment during their life, I expect that they would spend less time looking for a job and, hence, be more willing to accept any job offer. Of course, it depends on *when* the person was in these circumstances: being unemployed for 12 months two years ago, obviously, would induce a more profound change on one's behaviour in the labour market than being a long-term unemployed 20 years ago. Unfortunately, the dataset does not include the information on the last occurrence of such unemployment period.

The model also includes dummies for higher education of respondent's mother, father and partner to proxy respondent's social class. I expect that all three indicators would decrease overeducation, as individuals from higher social class would be more oriented to holding the best positions. I also expect that the effect from any parent supervising other employees in their job should have the same direction as in the case of partner, as explained above.

Finally, time-fixed effects are added. As already seen in Section 1.2, overeducation has been increasing over time. I expect it to be reflected in the results for time variables.

Two important variables are not included in models. Data on tenure are not available in rounds 1, 3 and 4. Ability, as defined in Section 3.3, cannot be included, as the measurement of household income changed in round 4. In the first three rounds, there were the same fixed intervals for all countries, while in rounds 4 and 5, ESS questionnaire authors switched to national household income deciles. Hence, ability would be defined differently in the first three rounds than in the other two, which is inappropriate for a model based on the pooled dataset.

[14] The idea about measuring immigrant background relative to the fully natives using factor variables, as opposed to dummy variables, was already present in Hazans (2011a).

3.4.2.2 Personality Variables

The eleven personality variables used in Section 3.3 and the variable measuring the importance of being successful were combined into three composite measures for reasons of parsimony and enhanced interpretability (see Table 3-8). Each composite measure is a summated scale of the four respective binary variables marking respondents “very much like” the description of the respective original personality variables. The summated scale is created by taking the average value of the four respective binary variables. Thus, each of them becomes a continuous variable with range $[0, 1]$.

The exact grouping of the twelve variables into three composite measures was derived from factor analysis. Principal-component factor, iterated principal factor and maximum-likelihood factor methods all gave the same groupings after rotation, whether orthogonal or oblique. Kaiser–Meyer–Olkin measure of sampling adequacy was between 0.80 and 0.90 for the whole sample, males and females; it was in the same interval also for each of the twelve variables individually. Cronbach’s αs were between 0.60 and 0.65. Both measures indicate that the grouping is acceptable (Hair, Black, Babin, Anderson, & Tatham, 2006).

The model includes these three composite personality variables (these will be referred to simply as personality variables).

3.4.2.3 Macro-Level Variables

Three country-level variables are included in the models:

- **The share of tertiary graduates among the employed.** It indicates the supply of tertiary graduates and the strength of competition among them. I expect it to have a

Table 3-8 – Grouping of Initial Personality Variables into Composite Personality Variables

Initial Personality Variable	Composite Personality Variables
Important to be treated equally Important to follow rules Important to help people Important to be loyal to friends	Social orientation
Important to be rich Important to show abilities Important to get respect Important to be successful [a]	Achievement orientation
Important to be creative Important to try new things Important to make decisions freely Important to seek adventures	Openness to experience

[a] This variable was not included into the models of Section 3.3. Its description is “Being successful is important to him. He hopes people will recognise his achievements.”

positive effect on overeducation risk, as higher share of graduates makes them harder to find jobs appropriate for this level of education.

- **Unemployment rate.** I expect it to have a positive effect on overeducation risk, as during higher unemployment individuals might be more willing to accept any job offer, even if they are too educated for it.
- **The share of employees in ISCO major groups 1–3 among the employed.** It complements the first indicator and shows the opportunities available for tertiary graduates. Higher values of this variable indicate more matching jobs for them and, hence, weaker competition for these jobs. Hence, I expect it to have a negative effect on overeducation risk.

3.4.3 Results

Table 3-9 shows the results. Note that, because of using mixed effects logistic model, all results are country-specific, as opposed to population-averaged[15].

3.4.3.1 Effects from Individual Characteristics

As expected, females are significantly more open to overeducation risk than males, corroborating the findings of the existing literature (see Section 2.1.4) and descriptive statistics (see Section 1.2). The difference in the risk faced is nearly 20 per cent, which is substantial.

However, the model does not show any significant effect from age (except from barely significant U-shaped nonlinearity for females). In order to study this issue deeper, I run additional models with more elaborate structure of age variables (see Figure 3-2). One can observe that compared to the very beginning of the working life (age 15–24), overeducation risk firstly rises during age 25–34. During the next ten years, it levels off, and by the age of 45–65, the risk is substantially (10 per cent for males and 20–30 per cent for females) smaller than in the start of career. In the end of a typical working life, the risk returns to its starting level. Interestingly, employed respondents aged over 75 have approximately twice as low risk of mismatch as their 15–24 year old counterparts.

Immigration status has a substantial effect on the risk of overeducation. Overall, immigrants from non-advanced economies outside Europe face a 1.8–2.1 times higher risk than the fully natives, where CEE immigrants are hurt more than Former Soviet Union (FSU) immigrants, and the latter, in turn, more than immigrants from Latin America, Africa and Asia (LAA). However,

[15] See Section C.4 for discussion.

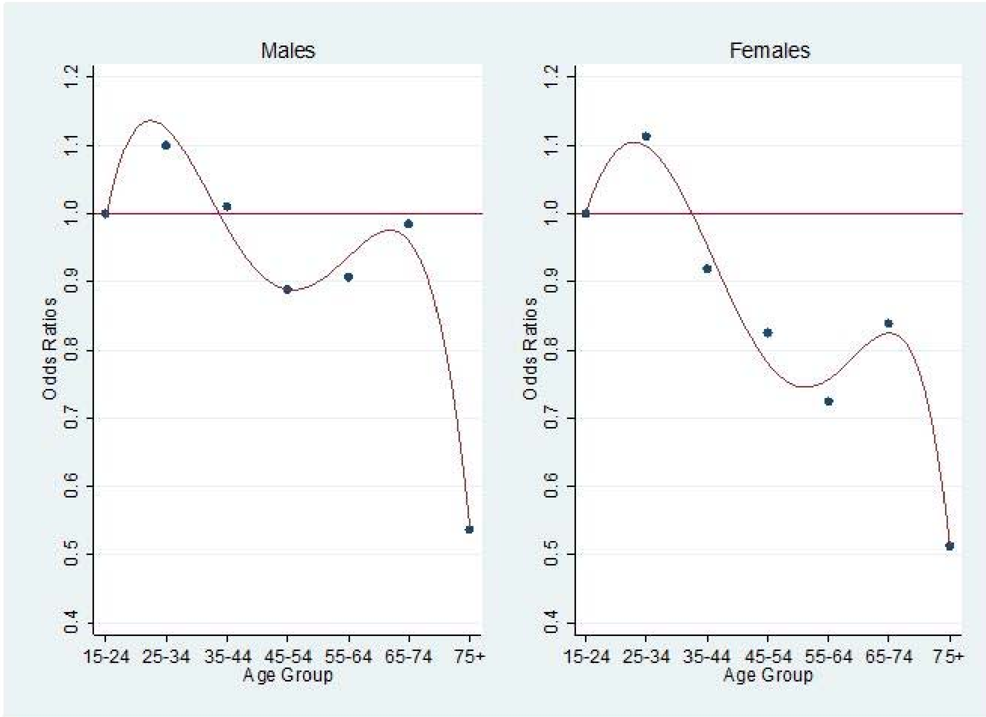

Figure 3-2 - Odds Ratios of Age Groups after Mixed Effects Logit Regression of Overeducation

The graph does not take into account the statistical significance of odds ratios.

there are substantial differences across sex. While for male immigrants, the risk premium is in the region of 40–85 per cent (being the largest for LAA and CEE immigrants), for females, it reaches 90–150 per cent (being the largest for FSU and CEE immigrants). For immigrants from non-CEE European countries, there is no significant risk premium for males, but there is a risk premium of around 27 per cent for females. Immigrants from advanced countries do not have any significant risk premium, although odds ratios are quite high for males.

Locally born females with one parent-immigrant face a premium on overeducation risk of 14–15 per cent, compared to the fully natives. On the contrary, minorities, male second-generation immigrants and locally born females with both parents-immigrants do not appear to have a significantly different risk exposure than the fully natives.

Personality variables strongly influence the risk of overeducation. Social orientation increases the exposure to this risk, while achievement orientation and openness to experience dampen the risk. Personality effects are more pronounced for males than for females. For the latter, openness to experience has a very strong effect, while achievement orientation is not significant and social orientation is barely significant. However, when looking at the contribution of personality variables to the overall explanatory power of models, one can observe that AIC changes more profoundly for females than for males. On the other hand, BPM-PRE model results show that there is substantial heterogeneity in the effects of personality factors across countries.

3.4.3.2 Effects from Family Characteristics

Contrary to expectations, having a child decreases the exposure to overeducation risk, as compared to not having one. This effect holds for both males and females, but the decrease is more pronounced for the latter. With three and more children, there is a barely significant increase in the risk for females.

Parental education affects same-sex descendants, but the direction of effects differs. Higher education of the father increases a male respondent's exposure to overeducation, but that of the mother decreases a female respondent's exposure to mismatch risk. This is a surprising finding, which shows that higher social status does not necessarily act as a buffer against the risk of overeducation, at least for men.

The fact that at least one of the respondent's parents has subordinates decreases the exposure to overeducation for men, but does not affect it for women.

Respondents living with a partner face a 10–20 per cent lower exposure to overeducation risk than those living without one. However, the relative structure of the effects from partner's employment status does not always follow my expectations. Only for women, the smallest reduction comes from an unemployed partner, a slightly higher from an employed partner and the highest reduction from a partner supervising other employees. For men, however, the smallest reduction is when their partner is employed, while the least overeducated are those whose partner is unemployed.

3.4.3.3 Effects from Labour Market Experience

Students who work in parallel to studies face a 20 per cent higher risk of overeducation than non-students. There are no sex differences in the effect of this variable.

Length of unemployment spell plays an important role in influencing mismatch risk. Whether it lasted for 3 months or 12 months, experiencing unemployment makes an individual more exposed to a job with lower requirements than their education. However, a 3-month unemployment experience appears to have a more profound effect than a 12-month experience. If the former increases the risk on average by 32 per cent, the latter does it only by 15 per cent. Moreover, there are significant sex differences. Women are much more affected by 3-month unemployment (the risk is 40 per cent higher for them, but only 20 per cent higher for men). On

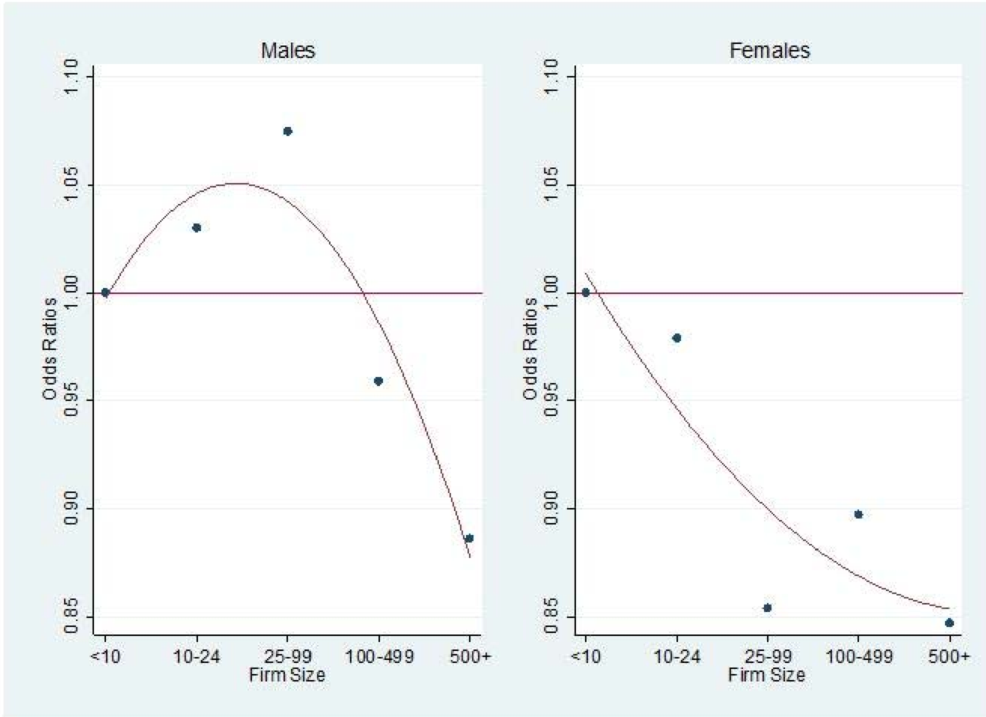

Figure 3-3 – Odds Ratios of Firm Size after Mixed Effects Logit Regression of Overeducation

The graph does not take into account the statistical significance of odds ratios.

the contrary, men are slightly more affected by 12-month unemployment (17 per cent vs. 11 per cent risk increase for women).

As expected, informal employment pressures individuals to get a lower-qualified job. Women appear to be more affected by working without a contract than men (32 per cent vs. 13 per cent increase in risk), but the overall effect of this variable is quite modest, as compared to expectations.

Overall, overeducation risk decreases with the size of the labour market to which the respondent is exposed, in terms of both domicile and firm. However, for domicile, the risk is smaller only in big cities (but not in small cities), as compared to rural areas. Moreover, I did not find a difference in the exposure of females based on their domicile. The risk also drops when firm size increases, reflecting the fact that larger organisations tend to be more accurate in taking into account individuals' education when assigning them to jobs. However, there are strong nonlinearities in this case (see Figure 3-3).

Table 3-9 – Odds Ratios after Mixed Effects Logistic Regression of Overeducation

	Total				Male				Female			
	Base	BP	BPM	BPM-PRE	Base	BP	BPM	BPM-PRE	Base	BP	BPM	BPM-PRE
General Demographics												
Age	1.000	0.998	0.998	0.998	1.001	1.000	0.999	0.999	1.004	1.002	1.001	1.001
$Age^2/100$	0.992	0.993	0.993	0.993	0.993	0.995	0.995	0.995	0.984†	0.986†	0.986†	0.986†
Female	1.189***	1.186***	1.189***	1.187***								
Number of Children (rel. to No children)												
1	0.932**	0.930**	0.933**	0.934**	0.938	0.937	0.940	0.940	0.924**	0.921**	0.923**	0.923**
2	0.965	0.961	0.965	0.967	0.917†	0.916*	0.921†	0.921†	1.009	1.000	1.005	1.006
3+	1.026	1.017	1.023	1.024	0.945	0.941	0.946	0.947	1.125†	1.110	1.115	1.116
Partner Employment Status (rel. to No partner)												
Unemployed	0.800***	0.795***	0.789***	0.791***	0.749***	0.749***	0.743***	0.743***	0.884**	0.871**	0.867***	0.868***
Employed	0.858***	0.851***	0.852***	0.852***	0.822***	0.820***	0.819***	0.818***	0.867***	0.854***	0.856***	0.857***
Supervising others	0.791***	0.788***	0.792***	0.794***	0.783***	0.783***	0.789***	0.788***	0.794***	0.786***	0.790***	0.792***
Domicile (rel. to Rural)												
Big city	0.955*	0.961†	0.963†	0.964	0.939*	0.944†	0.943†	0.944†	0.977	0.984	0.987	0.987
Small city	0.981	0.983	0.983	0.983	0.959	0.962	0.961	0.962	1.005	1.009	1.010	1.010
Firm Size (rel. to <10 employees)												
10 – 24	1.005	0.998	0.999	0.999	1.031	1.027	1.030	1.030	0.988	0.981	0.979	0.979
25 – 99	0.953*	0.946**	0.947*	0.947*	1.077*	1.069*	1.075*	1.075*	0.861***	0.855***	0.853***	0.854***
100 – 499	0.929**	0.919***	0.922**	0.922**	0.962	0.953	0.959	0.959	0.906**	0.897**	0.897**	0.897**
500+	0.870***	0.862***	0.864***	0.864***	0.887**	0.880**	0.886**	0.886**	0.856***	0.849***	0.847***	0.847***
Immigrant Background (rel. to Native)												
Minority	1.032	1.035	1.036	1.038	1.036	1.037	1.034	1.035	1.037	1.041	1.034	1.036
Parent-immigrant	1.055	1.059	1.058	1.058	0.964	0.965	0.968	0.969	1.141**	1.149**	1.141**	1.142**
Both parents immigrants	0.959	0.963	0.965	0.965	0.947	0.949	0.956	0.957	0.977	0.983	0.977	0.976
CEE[a] immigrant	2.108***	2.101***	2.079***	2.075***	1.780***	1.779***	1.762***	1.762***	2.455***	2.447***	2.422***	2.421***
FSU[b] immigrant	1.961***	1.956***	1.982***	1.984***	1.466**	1.464**	1.479**	1.483**	2.481***	2.462***	2.482***	2.478***
LAA[c] immigrant	1.847***	1.859***	1.859***	1.853***	1.813***	1.822***	1.820***	1.816***	1.901***	1.918***	1.910***	1.912***
Other European immigrant	1.135*	1.143*	1.143*	1.141*	0.989	0.994	0.998	0.997	1.266**	1.279***	1.269**	1.268**
Other[d] immigrant	1.167	1.171	1.157	1.149	1.282	1.284	1.271	1.270	1.058	1.067	1.044	1.038

Table 3-9 (cont.)

	Total				Male				Female			
	Base	**BP**	**BPM**	**BPM-PRE**	**Base**	**BP**	**BPM**	**BPM-PRE**	**Base**	**BP**	**BPM**	**BPM-PRE**
Potentially Negative Factors												
Student	1.227***	1.235***	1.233***	1.232***	1.221**	1.224**	1.222**	1.222**	1.219***	1.233***	1.233***	1.235***
Disabled	1.092***	1.090***	1.094***	1.093***	1.055	1.055	1.057	1.058	1.123***	1.121***	1.123***	1.121***
Was unempl. for 3 months	1.324***	1.321***	1.317***	1.317***	1.206***	1.202***	1.195***	1.195***	1.443***	1.444***	1.442***	1.442***
Was unemployed for 1 year	1.149***	1.151***	1.156***	1.154***	1.168***	1.170***	1.177***	1.176***	1.108**	1.110**	1.113**	1.112**
Informal employment	1.224***	1.221***	1.226***	1.229***	1.132**	1.131**	1.132**	1.134**	1.324***	1.324***	1.329***	1.330***
Parental & Partner Effects												
Higher education, mother	0.920**	0.923**	0.910***	0.911***	0.973	0.974	0.962	0.962	0.882***	0.887**	0.867***	0.867***
Higher education, father	1.025	1.029	1.023	1.025	1.079*	1.084*	1.079*	1.080*	0.976	0.980	0.971	0.972
Higher education, partner	1.164***	1.168***	1.153***	1.152***	1.238***	1.240***	1.218***	1.218***	1.104**	1.109***	1.095**	1.093**
Parent supervises others	0.968	0.976	0.979	0.978	0.938*	0.942*	0.947*	0.947*	0.995	1.005	1.010	1.009
ESS Round (rel. Round 1 (2002))												
Round 2 (2004)	1.154***	1.151***	1.013	1.012	1.172***	1.169***	1.022	1.022	1.140**	1.138**	1.021	1.019
Round 3 (2006)	1.131***	1.128***	1.001	1.000	1.062	1.059	0.959	0.959	1.204***	1.200***	1.045	1.044
Round 4 (2008)	1.225***	1.230***	1.009	1.007	1.184***	1.190***	0.979	0.979	1.270***	1.276***	1.051	1.048
Round 5 (2010)	1.309***	1.314***	1.013	1.012	1.259***	1.262***	0.953	0.952	1.362***	1.369***	1.098†	1.095†
Personality Factors												
Social orientation		1.170**	1.183**	1.179**		1.195*	1.208*	1.208*		1.143†	1.154†	1.148
Achievement orientation		0.824***	0.820***	0.825**		0.753***	0.741***	0.739***		0.904	0.903	0.914
Openness to experience		0.748***	0.746***	0.741***		0.848*	0.849*	0.852*		0.668***	0.667***	0.654***
Macro-Level Factors												
Tertiary graduates, share			88.166***	89.735***			86.872***	87.928***			85.777***	87.982***
Unemployment rate			13.765***	13.984***			115.996***	114.920***			0.888	0.920
ISCO 1-3, share			0.037***	0.037***			0.027***	0.027***			0.048***	0.049***
Constant	0.097***	0.106***	0.098***	0.097***	0.097***	0.104***	0.099***	0.099***	0.106***	0.118***	0.110***	0.109***

Table 3-9 (cont.)

	Total				Male				Female			
	Base	**BP**	**BPM**	**BPM-PRE**	**Base**	**BP**	**BPM**	**BPM-PRE**	**Base**	**BP**	**BPM**	**BPM-PRE**
Random Effects, Country-Level [e]												
Female, std. dev.	0.276	0.277	0.272	0.273								
	(0.042)	(0.042)	(0.041)	(0.041)								
Constant, std. dev.	0.623	0.626	0.329	0.331	0.623	0.628	0.351	0.351	0.603	0.605	0.315	0.314
	(0.078)	(0.078)	(0.045)	(0.046)	(0.079)	(0.079)	(0.049)	(0.049)	(0.078)	(0.078)	(0.046)	(0.046)
Personality factors, std. dev.				0.230				0.151				0.213
				(0.059)				(0.115)				(0.093)
Personality factors, correlation				–0.308				–0.336				–0.184
				(0.148)				(0.358)				(0.392)
Regression Diagnostics												
LR Test[f], p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
N	110642				57240				53402			
AIC	69914	69867	69640	69604	34008	33994	33848	33847	35903	35869	35744	35735
BIC	70289	70270	70072	70056	34339	34352	34233	34250	36232	36224	36126	36135

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$; † $p < 0.15$

Base, BP and BPM models were approximated by adaptive Gaussian quadrature with 7 integration points. BPM-PRE models were estimated by Laplace approximation.

[a] Central and Eastern Europe

[b] Former Soviet Union

[c] Latin America, Africa and Asia

[d] Immigrants from the US, Canada, Japan, Korea, Australia or New Zealand

[e] Standard errors in brackets

[f] Likelihood ratio test vs. logistic regression

3.4.3.4 Effects from Macro-Level Variables

As expected, the share of tertiary graduates increases the risk of mismatch and the share of ISCO 1–3 occupations decreases it. Both variables perform consistently in the whole sample and separately for males and females. This reflects the importance of the strength of competition in the labour market on the successful employment of graduates in matched jobs.

Unemployment has an expected pronounced positive effect in the whole sample and, especially, for males. Interestingly, it does not affect the risk of overeducation for females.

Importantly, the inclusion of macro-level variables made time-fixed effects, which were showing a strongly increasing trend before, insignificant. They, thus, explain the dynamics of overeducation over time quite well.

3.4.4 Conclusions

There are several important takeaways from this section. Firstly, women are at a significant disadvantage. Not only have they a higher risk exposure than men, all other characteristics equal, but also female immigrants are much more hurt than male immigrants. Secondly, immigrants have an around 100 per cent risk premium over the fully natives; immigrants from CEE and Former Soviet Union countries are at particular disadvantage. Both findings require additional attention from governments to the issue of making the labour market more just and improving the integration of immigrants into society.

Thirdly, the three general personality variables have important effects, whereby achievement orientation and openness to experience act as a buffer against overeducation, while social orientation acts as a catalyst. Fourthly, the size of the labour market open to an individual improves the possibilities for them to find a matching job. Fifthly, certain labour market experience, such as studies in parallel, lengthy unemployment spells and informal employment, all increase one’s exposure to overeducation. Sixthly, a higher social class does not necessarily decrease the risk of mismatch.

Finally, temporal dynamics of overall mismatch was well explained by the macro-level variables measuring supply and demand – the share of tertiary graduates and the share of ISCO 1–3 occupations. In times of high unemployment, individuals tend to accept any job, whatever its quality of match with their education.

3.5 Field of Study Effects for Tertiary Graduates

Given the expansion of higher education and an obvious bias in student demand for particular fields of study, described in Section 1.2, it is of high importance to study how field of study affects the risk of overeducation of its graduates. This section describes these findings.

3.5.1 Data and Methods

For this study, I use data from ESS rounds 2–4, as the field of study variable is available only in these rounds. The sample is further constrained to HEI graduates aged 20–40. The age limitation is necessary to mitigate age and cohort effects that could increase heterogeneity within fields of study.

To facilitate the analysis, I merge the original fields of study of ESS into seven broader fields, as shown in Table 3-10. Graduates of general fields and personal care and public order related fields were removed from the analysis, as they seem quite vague at tertiary level and the number of tertiary-level graduates of these fields is very limited.

I compare the performance of three models:

- Logistic model with standard errors clustered by country (see Sections C.2 and C.7)
- Fixed effects logistic model with country-level intercepts and standard errors clustered by country (see Section C.3)
- Mixed effects logistic model with three groups of country-level random effects: intercept, sex coefficient and field of study coefficients (see Section C.4)

The variables in these models mostly are the same as in the models of Section 3.4. The only difference is the addition of fields of study and running models on relatively young tertiary

Table 3-10 – Merging the Original Fields of Study from ESS into Broad Fields

ESS Field of Study	Broad Field of Study
Arts – Fine or Applied	Arts & Humanities
Humanities – Languages, Classics, History, Theology, etc.	Arts & Humanities
Technical & Engineering, incl. Architecture and Planning, Industry, Craft, Building Trades, etc.	Engineering
Transport & Telecommunications	Engineering
Agriculture & Forestry [a]	Science
Science, Mathematics, Computing, etc.	Science
Teacher Training or Education	Education
Medical, Health Services, Nursing, etc.	Health
Economics, Commerce, Business Administration, Accountancy, etc.	Economics, Business & Law
Law & Legal Services	Economics, Business & Law
Social & Behavioural Studies, Public Administration, Media, Culture, Sport & Leisure Studies, etc.	Social Sciences

[a] I am not the first one to unite Agriculture with Science; this was also done, e.g., by Kim and Kim (2003).

graduates in rounds 2–4, while in the mentioned section, models were run on all respondents, irrespective of their level of education (as soon as they completed at least primary level) and age and ESS round. Thus, differences in results are mainly due to this sample restriction, rather than from adding fields of study.

Data from different countries are pooled. These models are run on the pooled dataset of 32 European countries. In addition, these are run separately on each of four country groups:

- Northern Europe (Denmark, Finland, Norway and Sweden)
- Western Europe (Austria, Belgium, France, Germany, Ireland, Luxembourg, the Netherlands, Switzerland and the UK)
- Eastern Europe (Bulgaria, Croatia, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Romania, Russia, Slovakia, Slovenia and Ukraine)
- Southern Europe (Cyprus, Greece, Israel, Portugal, Spain and Turkey)

3.5.2 Results

Table 3-11 reports the unconditional incidence of overeducation in the countries and country groups considered. Southern Europe has the highest overeducation risk indicators in all fields of study, compared to other country groups. In engineering, science, education and social sciences, Eastern Europe also has quite high average overeducation incidence rates. On the contrary, Eastern Europe has the lowest incidence rate in health and the second lowest in arts and humanities (close to Northern Europe). The overeducation risk faced by graduates of economics, business and law (in this section also referred to as “economists” for short) is very close in Northern, Eastern and Western Europe.

Importantly, in Eastern Europe, the overeducation incidence of economists is 7.4 percentage points lower than that of engineers. This is a characteristic unique for this country group: in Southern Europe, both fields show essentially the same incidence rates, while in the other two country groups, the incidence for economists is 2–3 percentage points higher.

All three econometric models on the European sample show very close results for the overeducation risk from fields of study (see Table 3-12). Compared to the graduates in economics, business and law, a lower risk of overeducation is faced by graduates in arts and humanities and, especially, education and health. Graduates in engineering face higher mismatch risk than economists. Finally, economists face statistically the same overeducation risk as natural and social scientists.

This finding is unexpected and appears to be at odds with the descriptive statistics of Table 3-11. Hence, I reran models on each separate country group. Indeed, Table 3-13 shows that the risk-increasing effect of the engineering field comes exclusively from Eastern Europe, where engineers are more than 75 per cent more prone to become overeducated than economists. This goes in line with the findings of other researchers on the difficulties faced by engineers in the labour markets of CEE countries (Baranowska-Rataj & Unt, 2012). In no other field and in no

Table 3-11 – Unconditional Distribution of the Incidence of Overeducation over Fields of Study, Tertiary Graduates Aged 20–40

	Arts & Humanities	Engineering	Science	Education	Health	Economics, Business & Law	Social Sciences
EUROPE	**18.6%**	**24.8%**	**21.1%**	**10.9%**	**9.5%**	**22.7%**	**24.7%**
Northern Europe	**13.5%**	**16.6%**	**17.2%**	**7.2%**	**9.9%**	**19.9%**	**16.0%**
Denmark	18.9%	31.7%	18.9%	2.6%	3.6%	19.4%	29.0%
Finland	10.0%	13.7%	16.7%	9.7%	15.5%	26.2%	5.9%
Norway	17.4%	13.1%	23.2%	14.9%	11.8%	19.3%	24.2%
Sweden	7.7%	7.8%	10.0%	1.6%	8.5%	14.7%	4.8%
Western Europe	**19.3%**	**18.1%**	**18.1%**	**8.8%**	**10.8%**	**20.2%**	**18.4%**
Austria	20.5%	4.5%	30.0%	10.5%	10.0%	31.7%	19.0%
Belgium	25.0%	7.8%	6.6%	8.0%	14.9%	20.1%	22.2%
France	19.4%	18.8%	9.7%	7.7%	7.8%	25.9%	17.1%
Germany	18.5%	29.7%	24.5%	7.9%	10.3%	12.9%	10.9%
Ireland	32.8%	47.3%	40.0%	28.8%	20.8%	36.4%	48.0%
Luxembourg	7.1%	0.0%	13.3%	0.0%		6.3%	0.0%
Netherlands	15.8%	8.3%	4.2%	0.0%	5.0%	6.0%	6.7%
Switzerland	6.8%	19.6%	14.9%	0.0%	4.7%	8.6%	6.7%
UK	28.0%	26.5%	19.8%	15.9%	13.0%	33.6%	35.1%
Eastern Europe	**14.9%**	**29.2%**	**23.4%**	**11.7%**	**6.3%**	**21.8%**	**27.2%**
Bulgaria	10.5%	3.7%		0.0%	5.6%	12.2%	
Croatia	9.1%	30.0%		26.7%		25.0%	16.7%
Czech Republic	8.3%	6.7%	11.1%	0.0%	0.0%	16.7%	
Estonia	16.7%	13.8%	20.7%	3.6%	0.0%	10.6%	17.1%
Hungary	16.0%	34.3%	39.1%	7.1%	15.8%	27.3%	35.7%
Latvia	32.3%	39.4%	13.3%	9.8%	5.0%	22.0%	26.1%
Lithuania	20.0%	40.0%	36.4%	24.0%	0.0%	30.8%	33.3%
Poland	8.8%	25.5%	10.8%	20.4%	0.0%	23.4%	38.7%
Romania	0.0%	15.2%	6.3%	11.1%	7.7%	25.3%	28.6%
Russia	19.6%	62.3%	41.7%	26.4%	9.1%	31.3%	12.5%
Slovakia	23.1%	28.0%	9.1%	4.0%	16.7%	11.1%	36.4%
Slovenia	14.3%	26.7%	17.2%	13.8%	11.1%	30.6%	22.7%
Ukraine	15.6%	54.5%	51.3%	5.4%	4.0%	17.5%	31.6%
Southern Europe	**29.0%**	**30.6%**	**24.2%**	**15.1%**	**13.8%**	**30.2%**	**35.5%**
Cyprus	16.1%	28.0%	29.2%	7.9%	10.3%	38.1%	45.2%
Greece	42.3%	57.1%	41.0%	20.9%	15.4%	44.7%	55.0%
Israel	45.0%	43.2%	20.0%	16.7%	16.1%	21.2%	30.3%
Portugal	8.0%	10.0%	6.3%	11.4%	20.6%	18.6%	30.3%
Spain	40.3%	22.0%	23.9%	26.8%	12.9%	37.9%	27.0%
Turkey	22.2%	23.3%	25.0%	6.7%	7.7%	20.7%	25.0%

Numbers at the country group level are simple averages of the countries in that group.

Table 3-12 – Odds Ratios after Different Logistic Regressions of Overeducation for Europe

	Mixed Effects	Standard Logistic	Fixed Effects
Field of Study, rel. to Economics, Business & Law			
Arts & Humanities	0.721**	0.789**	0.734***
Engineering	1.198†	1.349†	1.381*
Science	0.873	0.922	0.857
Education	0.330***	0.364***	0.344***
Health	0.320***	0.353***	0.329***
Social Sciences	0.997	1.024	0.994
General Demographics			
Age	0.797***	0.796***	0.789***
$Age^2/100$	1.327***	1.339***	1.344***
Female	1.109	1.094	1.110
Number of Children (rel. to No Children)			
1	1.019	0.964	1.017
2	1.078	1.052	1.057
3+	0.907	0.907	0.889
Partner Employment Status (rel. to No Partner)			
Unemployed	1.236*	1.227	1.309*
Employed	1.374***	1.333***	1.430***
Supervising others	1.190*	1.200*	1.235**
Domicile (rel. to Rural)			
Big city	0.705***	0.704***	0.709***
Small city	0.843**	0.858*	0.837**
Firm Size (rel. to <10 employees)			
10 – 24	0.773***	0.766***	0.789***
25 – 99	0.572***	0.564***	0.572***
100 – 499	0.538***	0.559***	0.543***
500+	0.519***	0.570***	0.528***
Immigrant Background (rel. to Native)			
Minority	1.309**	1.100	1.315***
Parent-immigrant	0.933	0.875	0.937
Both parents immigrants	1.239	1.119	1.265
Central & Eastern Europe immigrant	3.859***	3.648***	3.714***
Former Soviet Union immigrant	3.121***	2.499**	2.805**
LAA[a] immigrant	2.799***	2.874***	2.763***
Other European immigrant	1.111	1.091	1.117
Other[b] immigrant	1.355	1.626†	1.342
Potentially Negative Factors			
Student	1.716***	1.714***	1.689***
Disabled	1.246**	1.182*	1.245***
Was unemployed for 3 months	1.386***	1.378***	1.391***
Was unemployed for 1 year	1.449***	1.360***	1.441***
Informal employment	1.576***	1.575***	1.576***

country group was there a higher overeducation risk than for graduates in economics, business and law.

The benefits of graduates in education and health appear to be ubiquitous, which cannot be said of other fields of study. For instance, natural scientists face lower overeducation risk than economists in Southern and Western Europe, but not in Northern and Eastern Europe, while the situation with arts and humanities is exactly the opposite. Social scientists and economists do not have statistical differences in their exposure to mismatch in Northern, Eastern and Southern

Table 3-12 (cont.)

	Mixed Effects	Standard Logistic	Fixed Effects
Parental & Partner Effects			
Higher education, mother	0.808***	0.789***	0.826***
Higher education, father	0.628***	0.625***	0.619***
Higher education, partner	0.568***	0.583***	0.559***
Parent supervises others	0.947	0.974	0.938
ESS Round (rel. Round 2 (2004))			
Round 3 (2006)	1.050	1.059	1.026
Round 4 (2008)	1.126†	1.172*	1.087
Personality Factors			
Social orientation	1.350*	1.429**	1.358†
Achievement orientation	0.787*	0.776*	0.775*
Openness to experience	0.596***	0.617***	0.603***
Macro-Level Factors			
Tertiary graduates, share	67.404***	29.660***	613.960***
Unemployment rate	1.293	0.343	1.434
ISCO 1–3, share	0.005***	0.007***	0.001***
Constant	68.467***	79.869***	
Random Effects, Country-Level [c]			
Female, std. dev.	0.310		
	(0.076)		
Constant, std. dev.	0.288		
	(0.067)		
Fields of study, std. dev.	0.309		
	(0.042)		
Regression Diagnostics			
N		10771	
LR test vs. logistic model, p-value	0.0000		
Log-likelihood	–4724	–4814	–4630
AIC	9549	9690	9351
BIC	9913	9916	9687
Area under ROC curve [d]	0.7814	0.7510	0.6663
	[0.77; 0.79]	[0.74; 0.76]	[0.65; 0.68]

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$; † $p < 0.15$
The mixed effects model was estimated by Laplace approximation.
[a] Latin America, Africa and Asia
[b] Immigrants from the US, Canada, Japan, Korea, Australia or New Zealand
[c] Standard errors in brackets
[d] 95% confidence intervals in brackets

Europe, but the former are better positioned in Western Europe (though the significance of this effect is marginal).

While this section aimed at analysing the effects coming from fields of study, it is interesting to compare the results on tertiary graduates aged 20–40 with the results on all respondents described in Section 3.4.3.

Both models show similar effects from personality factors, macro-level factors (except for unemployment rate, which is not significant in the current models), domicile (although the risk is lower in small cities than in rural areas in the current models), firm size, immigrant background (although minorities are at disadvantage in the current models) and potentially negative factors.

While the current models clearly report a U-shaped relationship between age and mismatch risk, which was not shown in the general model of Section 3.4.3, it is generally in line with Figure 3-2.

Contrary to the general model, the current models show no effect from sex and living with children, a risk-increasing effect from living with a partner (with the highest effect from a partner working in a non-supervising position), risk-decreasing effects from higher education of both parents and partner, and no effect from the supervising position of a parent.

Comparing the performance of the three models, standard logistic model has worse performance than the mixed effects model (as one might have expected). However, regression fit measures of mixed effects and fixed effects models give contrary results. Measures based on the likelihood function – log-likelihood itself and the information criteria – speak in favour of the fixed effects model. Area under ROC curve strongly favours the mixed effects model; note that 95 per cent confidence intervals of both models' areas under ROC curve lie considerably far away from each other. The same picture is observed in the results on country groups (not reported).

How should this be interpreted? Recall from Appendix C that log-likelihood measures operate in terms of the probability of observing the results of the model (given its parameters). ROC analysis, on the other hand, assesses the classification accuracy of the model, taking into account the quality of classifying both positive and negative cases.

Returning to the case at hand, the mixed effects model gives results having a lower probability of observing them, but with considerably higher classification accuracy. Clearly, the latter measure should be used in measuring the quality of the model. Hence, the mixed effects model appears to be better than the fixed effects model.

Predictions of the probability of overeducation averaged over fields of study reported in Table 3-14 are not homogeneous inside country groups. Take Northern Europe as an example.

Table 3-13 – Odds Ratios of Field-of-Study Effects after Mixed Effects Logit Regressions of Overeducation for Country Groups

	North	East	South	West	Europe
Field of Study, rel. to Economics, Business & Law					
Arts & Humanities	0.496*	0.657*	0.755	0.839	0.721**
Engineering	0.820	1.768***	0.769	1.087	1.198†
Science	0.695	1.128	0.646*	0.758*	0.873
Education	0.214***	0.408***	0.306***	0.314***	0.330***
Health	0.320***	0.177***	0.267***	0.394***	0.320***
Social Sciences	0.652	1.335	1.110	0.766†	0.997

*** p < 0.01; ** p < 0.05; * p < 0.10; † p < 0.15

Models estimated with Laplace approximation. Models for Western and Eastern Europe contain country-level random effects of gender (as in the case of the whole Europe); for Northern and Southern Europe, these were not significant and were dropped.

While Finland and Norway have the same overeducation risk of engineers, in Denmark this risk is much higher, while in Sweden much lower. For graduates in education, however, it is Norway where the risk is twice as high as in the other three Northern countries. Compared to other country groups, Nordic countries are much more homogeneous, as was expected.

Table 3-14 – Predicted Probabilities of Overeducation over Fields of Study after Mixed Effects Logistic Model, Tertiary Graduates Aged 20–40

	Arts & Humanities	Engineering	Science	Education	Health	Economics, Business & Law	Social Sciences
EUROPE	**18.9%**	**25.7%**	**20.8%**	**11.3%**	**9.2%**	**22.8%**	**24.5%**
Northern Europe	**13.5%**	**16.3%**	**17.5%**	**6.8%**	**9.8%**	**19.8%**	**16.0%**
Denmark	15.0%	26.0%	19.0%	6.0%	8.0%	20.0%	20.0%
Finland	12.0%	15.0%	18.0%	7.0%	13.0%	24.0%	13.0%
Norway	19.0%	15.0%	21.0%	10.0%	11.0%	21.0%	22.0%
Sweden	8.0%	9.0%	12.0%	4.0%	7.0%	14.0%	9.0%
Western Europe	**18.8%**	**20.1%**	**16.8%**	**9.1%**	**11.0%**	**19.9%**	**18.8%**
Austria	23.0%	18.0%	22.0%	9.0%	12.0%	25.0%	21.0%
Belgium	19.0%	14.0%	11.0%	9.0%	10.0%	20.0%	18.0%
France	19.0%	19.0%	15.0%	7.0%	9.0%	22.0%	18.0%
Germany	23.0%	25.0%	18.0%	8.0%	12.0%	17.0%	17.0%
Ireland	40.0%	44.0%	37.0%	23.0%	22.0%	38.0%	42.0%
Luxembourg	5.0%	7.0%	8.0%	3.0%		8.0%	6.0%
Netherlands	5.0%	9.0%	6.0%	3.0%	3.0%	8.0%	5.0%
Switzerland	10.0%	16.0%	11.0%	5.0%	5.0%	11.0%	10.0%
UK	25.0%	29.0%	23.0%	15.0%	15.0%	30.0%	32.0%
Eastern Europe	**15.4%**	**30.8%**	**24.3%**	**12.0%**	**5.7%**	**21.7%**	**27.5%**
Bulgaria	8.0%	13.0%		6.0%	3.0%	11.0%	
Croatia	17.0%	29.0%		19.0%		24.0%	22.0%
Czech Republic	7.0%	13.0%	12.0%	4.0%	2.0%	14.0%	
Estonia	10.0%	17.0%	17.0%	6.0%	3.0%	12.0%	17.0%
Hungary	17.0%	36.0%	31.0%	12.0%	9.0%	29.0%	31.0%
Latvia	23.0%	36.0%	22.0%	12.0%	6.0%	23.0%	28.0%
Lithuania	23.0%	40.0%	30.0%	17.0%	8.0%	30.0%	37.0%
Poland	13.0%	27.0%	17.0%	16.0%	4.0%	23.0%	30.0%
Romania	13.0%	20.0%	13.0%	11.0%	5.0%	22.0%	25.0%
Russia	21.0%	60.0%	42.0%	23.0%	8.0%	32.0%	29.0%
Slovakia	15.0%	29.0%	18.0%	10.0%	7.0%	14.0%	25.0%
Slovenia	16.0%	28.0%	21.0%	12.0%	7.0%	28.0%	28.0%
Ukraine	17.0%	52.0%	44.0%	8.0%	6.0%	20.0%	30.0%
Southern Europe	**30.2%**	**29.5%**	**22.7%**	**15.8%**	**13.7%**	**31.7%**	**33.5%**
Cyprus	29.0%	28.0%	20.0%	13.0%	14.0%	36.0%	39.0%
Greece	39.0%	46.0%	40.0%	23.0%	21.0%	44.0%	49.0%
Israel	40.0%	38.0%	25.0%	19.0%	17.0%	25.0%	28.0%
Portugal	19.0%	17.0%	14.0%	10.0%	8.0%	22.0%	24.0%
Spain	33.0%	28.0%	23.0%	20.0%	14.0%	36.0%	37.0%
Turkey	21.0%	20.0%	14.0%	10.0%	8.0%	27.0%	24.0%

In each country group, the results for country are taken from the model run on that country group (rather than on the whole Europe). Numbers at the country group level are simple averages of the countries in that group. Models estimated with Laplace approximation. Models for Western and Eastern Europe contain country-level random effects of sex (as in the case of the whole Europe); for Northern and Southern Europe, these were not significant and were dropped. Highlighted country–field pairs have predicted overeducation risk of at least 0.25.

The best (predicted) match in most or all fields of study was found in Luxembourg, the Netherlands, Switzerland, Sweden, Bulgaria and Czech Republic.

For such an important field of study as engineering, most Eastern and Southern European countries have a huge (at least 25 per cent) exposure to overeducation; however, high overeducation risk for engineers was also found in Ireland and the UK.

3.5.3 Conclusions

The main result of this section is that only Western Europe shows signs of overproduction of economists, as four out of six other fields of study have significantly lower exposure to overeducation. Education and health are the most lucrative fields of study in terms of having the lowest risk of overeducation in all country groups.

Eastern Europe is unique in having tremendously poor position of graduates in engineering. Note that this result was found for engineers aged 20–40, i.e., quite recent graduates of engineering programmes in HEIs. This means that governments of Eastern European countries – especially, Russia and Ukraine – should make a major revision in their LEMS policies towards these graduates. Without providing sizeable incentives for businesses to open positions for engineers (where these engineers would *like* to work – providing such positions for minimum wage is not a solution), no incentive for studying engineering at tertiary level should be provided.

3.6 Industry Effects

As shown, field of study has important effects on overeducation. However, not only field of study determines what competition one has in the labour market. Effects from the industry where one ends up in might also be important. One can defend a master's thesis in one field, but work in a completely different industry. This section studies how different overeducation risk is across industries.

3.6.1 Data and Methods

For that purpose, I use data from ESS round 5, where industry is classified using NACE rev.2. I take 12 industries for analysis: manufacturing; construction; trade; transportation and storage; accommodation and food; information and communication technology (ICT); finance and insurance; professional, scientific and technical activities; administrative and support activities; public administration and defence; education; and human health and social work. These industries were chosen because they have a sufficient number of observations in most countries.

Twenty-seven countries available in the data[16] are pooled. However, models do not include Estonia, because the tenure variable, which is an important predictor of overeducation, is undefined in that country. The sample contains employed individuals with at least primary education level aged below 65 and having tenure below 42 years.[17] Country–industry pairs with less than eight observations are removed from the analysis. In female-specific models, construction and ICT industries are removed from the analysis, as the number of observations per country in these industries is very low in nearly all countries.

I run mixed-effects logistic regressions (see Section C.4) with the dependent variable being respondent's overeducation status. Firstly, models are run on the whole sample with the aim of finding the relative risk of mismatch in the industries under consideration, controlling for the factors to be explained shortly. Secondly, the whole sample is split by sex to check for sex-specific effects. Finally, separate models are run for each industry with the aim of analysing possible differences of the effects from explanatory variables across industries.

Explanatory variables include:

- Basic demographics (age and its square if significant, sex, living with a partner)
- Living in a big city to control for the size of the labour market
- Immigrant background to find out which immigrants are more disadvantaged
- Potentially negative factors (dummies for immigrant, disability and ever experiencing at least 3-month unemployment)
- Tenure to control for labour market experience
- Personality traits
- Social class (a dummy reflecting that at least one parent is tertiary-educated)
- Country- and industry-specific macro-level variables (share of tertiary graduates, unemployment rate and share of ISCO 1–3 occupations)

In non-industry-specific models, three-level mixed effects structure is imposed on the model, where the first level is country, the second level is industry and the third is individual. In industry-specific models, naturally, only country and individual levels are kept in the model. Country-level random coefficient on sex is introduced if significant.

---

[16] Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Israel, Lithuania, the Netherlands, Norway, Poland, Portugal, Russia, Slovakia, Slovenia, Spain, Sweden, Switzerland, the UK and Ukraine.

[17] Few respondents have higher age or tenure. Leaving them in the estimation sample would lead to potentially biased coefficients on these two variables. Assume, for instance, that overeducation decreases with tenure. Assume further that there is a single respondent at the tenure of 50 years, who happens to be overeducated. Then overeducation at that tenure level is 100%, which would drive the estimated tenure coefficient to zero.

3.6.2 Results

3.6.2.1 Average Overeducation Rates in Industries

Consider first simple averages of overeducation incidence over countries by industry (Figure 3-4). Three groups of industries emerge. The first group has average overeducation rate exceeding 15 per cent. It consists of administrative (24 per cent) and accommodation (20 per cent) industries. The second group has medium mismatch levels of around 13–14 per cent. It includes trade, transportation, manufacturing, public administration and finance. The last group is characterised by low overeducation of 10 per cent and below. Construction, professional and scientific activities, ICT, health and education form this group.

However, one of the major factors influencing the rate of overeducation is the share of occupations belonging to ISCO major groups 4–9 (below-tertiary occupations) in a given industry. Let me analyse this issue deeper.

There are in general three patterns of relationships between below-tertiary occupation share and overeducation rate (see Figures 3-5 through 3-7 for country examples). Firstly, there are countries with two industry clusters: (1) with low below-tertiary profession share (below 30 per cent) and (2) with high below-tertiary profession share (above 50 per cent). Generally, the former clusters have lower overeducation rates than the latter. Secondly, most countries show a continuous positive relationship between the two indicators (this especially concerns Russia and Ukraine, where, depending on industry, overeducation stretches from below 10 per cent to above 50 per cent). Finally, in several countries, overeducation and below-tertiary profession share appear to be unrelated.

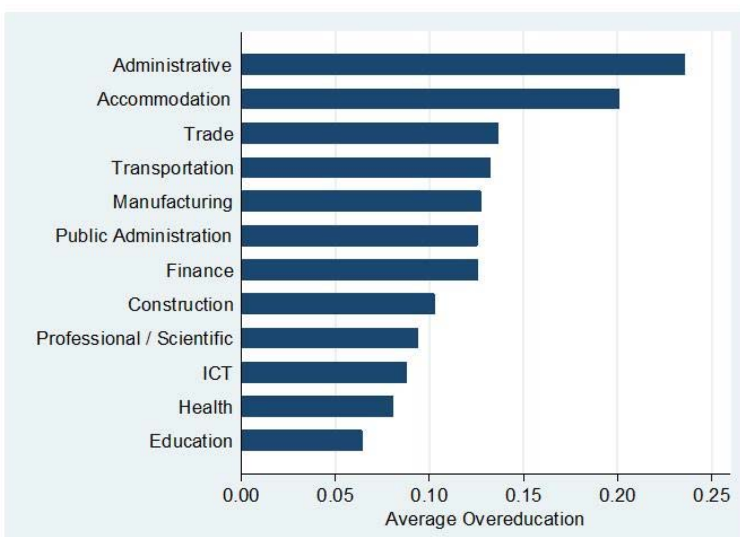

Figure 3-4 – Industry Positions on Overeducation, Averaged over Countries

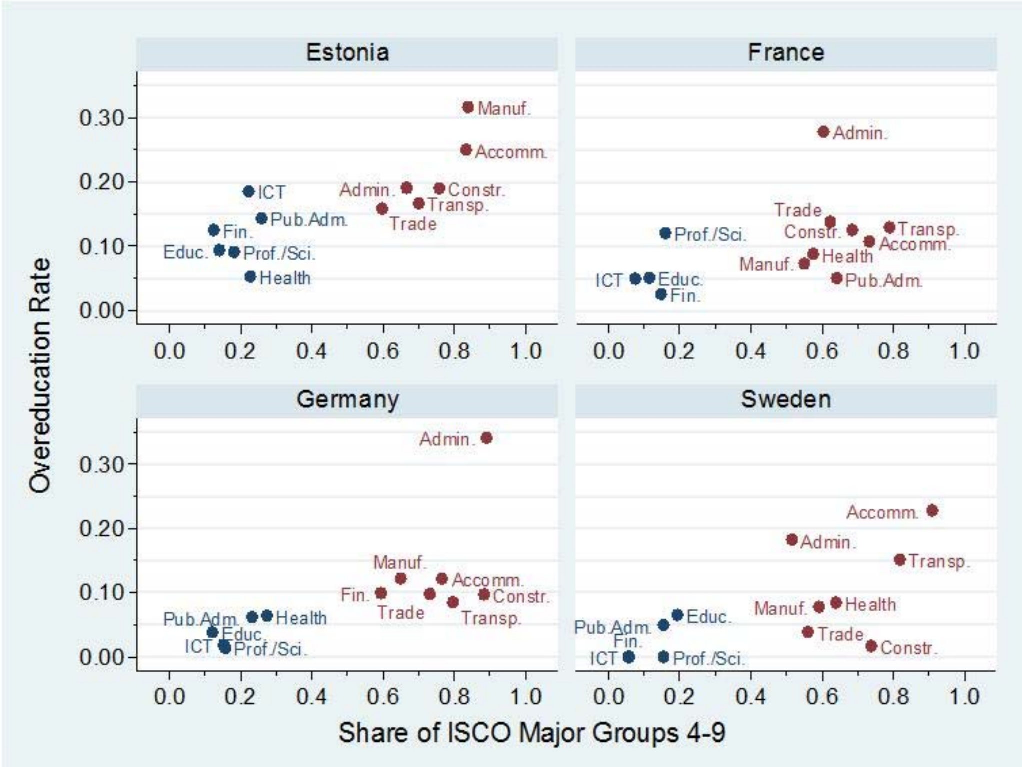

Figure 3-5 – Country-Specific Industry Positions on Overeducation vs. Share of ISCO 4–9: Countries with Two Distinct Industry Groups

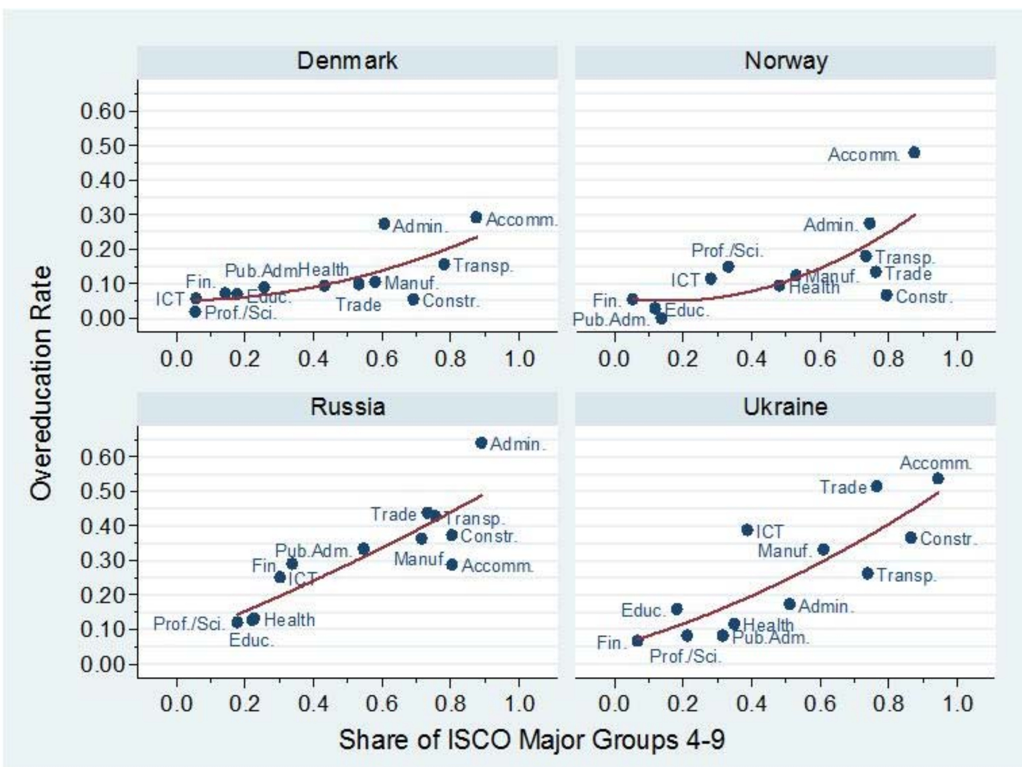

Figure 3-6 – Country-Specific Industry Positions on Overeducation vs. Share of ISCO 4–9: Countries with Continuous Positive Relationship

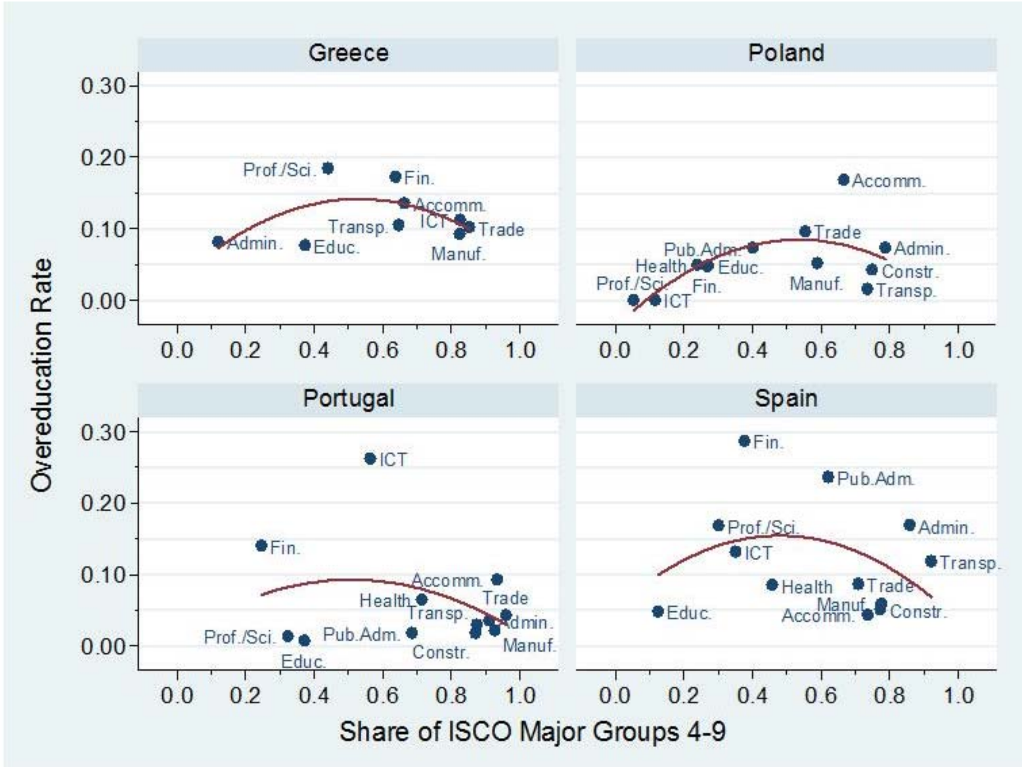

Figure 3-7 – Country-Specific Industry Positions on Overeducation vs. Share of ISCO 4–9: Countries with No Relationship

3.6.2.2 General Models

Table 3-15 presents models run over the whole sample. I will focus on the effects of industry, as other effects have already been discussed in previous sections.

Figure 3-8 compares the positions of industries in the overall model relative to manufacturing before and after controlling for macro variables. While previously, the inclusion of these variables influenced only the explanatory power of time fixed effects, it has important consequences for the relative position of industries in this model (of course, partially because macro-variables are defined here at industry level). On the upper end of the distribution of odds ratios, it keeps administrative services and accommodation at high-risk level, but places finance as the most vulnerable industry to mismatch and adds public administration to the risky industries. The lower end, however, is changed completely. Professional and scientific occupations, ICT, health and education, thus, show comparatively lower overeducation incidence purely due to macro-level factors. Controlling for them, construction becomes the only industry employing a significantly lower share of overeducated.

Figure 3-9 shows considerable differences in industry effects by sex. The only effect that is significant for both males and females is that from administrative services. Otherwise, the

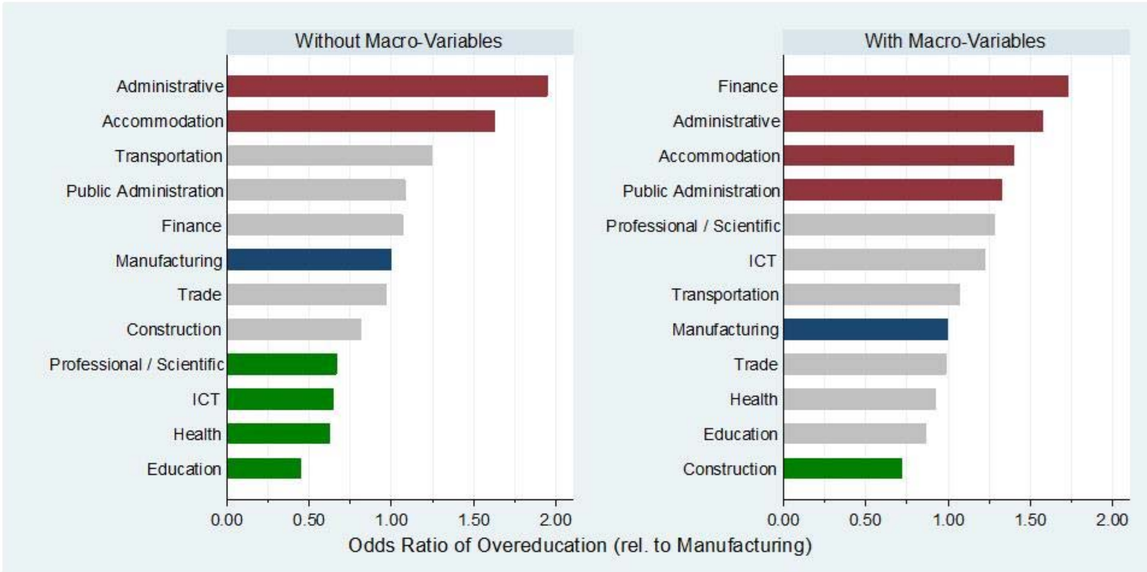

Figure 3-8 – Odds Ratios of Overeducation for Different Industries Relative to Manufacturing, Without and With Macro Variables

The blue bar represents the reference industry. Grey bars represent industries that are not significantly different from the reference industry. Red (green) bars represent industries whose odds ratios of overeducation are significantly higher (lower) than for the reference industry.

significance and direction of industry effects appear to be related to the traditional bias towards employing a particular sex in the industry. For instance, women have a comparatively higher overeducation risk in finance and professional and scientific services, while for men, the effect from the former is not significant and that from the latter is significantly smaller than in manufacturing. For men, significant risk-increasing effects come from accommodation and public administration, both of which are not significant for women.

3.6.2.3 Industry-Specific Models

The results of these models are presented in Table 3-16.

The results for the female dummy somewhat reflect the combination of the first two panels of Figure 3-9. This especially concerns the very large odds ratio in the model for the professional and scientific services industry, where the figure showed significant effects of different directions for males and females, or the risk-decreasing odds ratio in the model for public administration, where the effect direction in the figure was also the opposite for both sexes.

Notably, the 12 industries analysed divided precisely in half based on whether women are at disadvantage relative to men. The traditions of primarily employing individuals of a certain sex in a particular industry referred to when discussing the results of the general model, however,

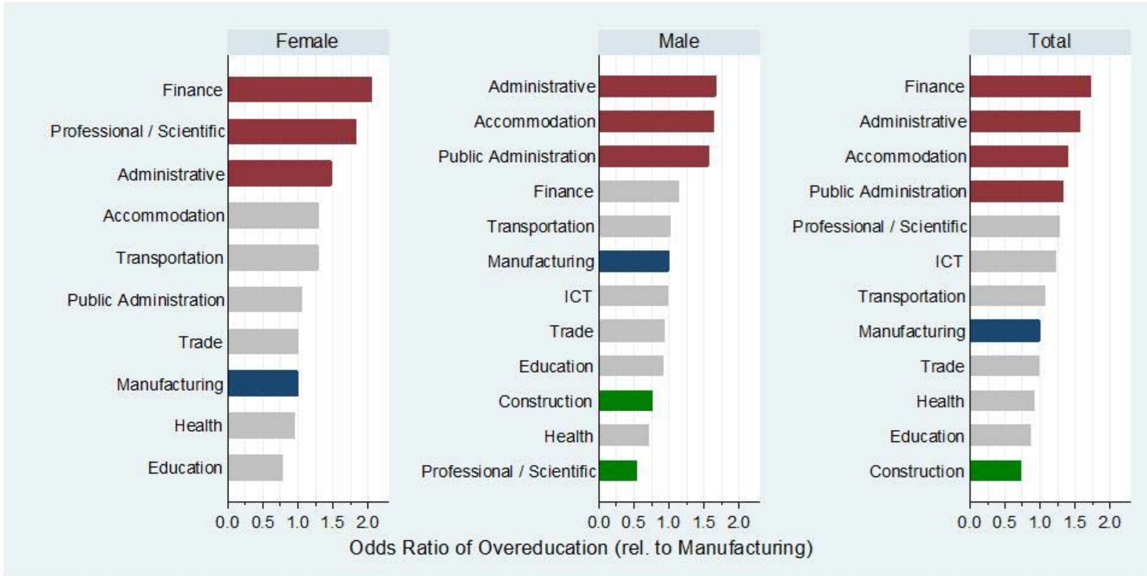

Figure 3-9 – Odds Ratios of Overeducation for Different Industries Relative to Manufacturing, by Sex

In all cases, results are from the models controlling for macro-level factors. The blue bar represents the reference industry. Grey bars represent industries that are not significantly different from the reference industry. Red (green) bars represent industries whose odds ratios of overeducation are significantly higher (lower) than for the reference industry.

appear to play a smaller role here. For instance, the odds ratio is not significant in construction, which is more associated with men, or accommodation and education, which are more associated with women. Given that the risk premium for females is in the range of 30 to nearly 290 per cent where significant and in five out of six industries where their exposure is not statistically significantly higher it is also not significantly lower than for males, it is not surprising that in the general model, the risk premium for females averages to 30 per cent.

Of the other explanatory variables, only tenure has a significant and stable direction of the effect in the majority of industries, although it is not significant in construction or ICT and risk-increasing in finance. Finance appears to be a particular industry in several respects. Firstly, sex differences there are quite high, as already noted. Secondly, it is the only industry where overeducation risk decreases with age. Thirdly, together with health industry, it exposes those living in big cities to higher overeducation risk.

Professional and scientific industry is the only one where social orientation has a significant effect. Personality traits, in general, appear to be weakly related to industry-specific overeducation risk. Openness to experience, for instance, is able to decrease the exposure to it only in manufacturing, construction, professional and scientific services, and health services.

Table 3-15 – Odds Ratios after Mixed Effects Logit of Overeducation, All Industries

	Total	Male	Female
Industry (rel. to Manufacturing)			
Construction	0.725**	0.775†	
Trade (wholesale & retail)	0.989	0.947	1.012
Transportation & Storage	1.071	1.021	1.289
Accommodation & Food	1.400**	1.651**	1.305
Information & Communication	1.227	0.986	
Finance & Insurance	1.732***	1.149	2.050***
Professional, Scientific & Technical	1.285	0.538*	1.841**
Administrative & Support	1.579***	1.671***	1.483**
Public Administration & Defence	1.332*	1.565**	1.064
Education	0.867	0.915	0.793
Human Health & Social Work	0.928	0.719	0.953
Basic Demographics			
Age	1.009***	1.011***	1.009**
Female	1.309***		
Lives with partner	0.880**	0.809**	0.892†
Lives in a big city	1.074	1.073	1.064
Potentially Negative Factors			
Immigrant	1.501***	1.385***	1.624***
Disabled	1.159**	1.200†	1.139
Ever unemployed for at least 3 months	1.126**	1.031	1.210**
Tenure	0.970***	0.977***	0.963***
Personality Traits			
Social orientation	1.142	1.132	1.179
Achievement orientation	1.152	0.940	1.376*
Openness to experience	0.602***	0.668†	0.536***
Parent is tertiary-educated	1.040	1.225**	0.952
Macro-Level Variables			
Tertiary graduates, share	10.559***	14.764***	12.144***
Unemployment rate	0.772	1.530	0.652
ISCO 1–3, share	0.030***	0.031***	0.029***
Constant	0.169***	0.128***	0.223***
Random Effects (country-level)			
Female (std.dev.)	0.124		
	(0.086)		
Constant (std.dev.)	0.331	0.319	0.294
	(0.068)	(0.087)	(0.077)
Random Effects (industry-level)			
Constant (std.dev.)	0.225	0.286	0.301
	(0.051)	(0.077)	(0.067)
Diagnostics			
Likelihood ratio test rel. to logit, p-value	0.0000	0.0000	0.0000
N	15567	7482	7794
AIC	10631	4861	5587
BIC	10861	5054	5768

*** p < 0.01; ** p < 0.05; * p < 0.10; † p < 0.15

Immigrant status and tertiary education of at least one parent affect the risk in approximately half of the industries, but the two sets of affected industries nearly do not intersect. Disability and three-month unemployment experience increase the exposure only in few industries – public administration in case of the former and education and health in case of the latter. Likewise, living with a partner decreases the exposure only in ICT and education.

Table 3-16 – Odds Ratios after Mixed Effects Logit of Overeducation, by Industry

	Manuf.	Constr.	Trade	Transp.	Accomm.	ICT	Finance	Prof. & Sci.	Admin.	Publ. Adm.	Education	Health
Basic Demographics												
Age	1.009	0.989	1.006	1.005	1.192***	0.983	0.941***	1.011	1.010	1.243***	1.045***	1.028***
$Age^2/100$					0.814***					0.786***		
Female	1.292*	0.947	1.293	1.405*	1.125	1.757†	2.092***	3.870***	1.080	0.745†	0.921	1.553*
Lives with partner	0.860	0.824	0.996	0.898	0.994	0.436**	1.220	0.714	1.014	0.918	0.688*	0.797
Lives in a big city	0.890	0.961	1.030	1.155	0.981	1.155	1.743**	1.048	1.069	1.041	0.946	1.595***
Potentially Negative Factors												
Immigrant	1.469**	1.272	1.354	1.753*	2.036***	1.946	0.663	0.858	1.950**	1.163	1.563	1.945***
Disabled	1.155	1.262	1.074	1.009	1.027	0.961	1.334	0.779	1.423	1.656*	1.030	1.019
Ever unemployed for at least 3 months	1.171	1.009	1.003	0.781	1.418†	0.777	1.093	1.002	1.178	0.811	1.638**	1.483**
Tenure	0.972***	0.982	0.969***	0.980*	0.964**	1.028	1.035†	0.942***	0.949***	0.963***	0.963***	0.939***
Personality Traits												
Social orientation	1.104	0.782	1.479	0.965	0.862	0.418	0.636	7.101***	1.049	2.207	0.926	0.888
Achievement orientation	1.095	1.133	1.364	1.761	1.759	2.060	2.428	1.020	0.533	1.161	0.718	1.194
Openness to experience	0.389**	0.331†	0.561	1.007	0.847	0.798	1.101	0.226**	0.923	0.646	0.828	0.257**
Parent is tertiary-educated	1.125	1.747**	1.535***	1.402†	1.142	0.585	1.274	0.579*	0.909	1.380†	0.648*	0.807
Constant	0.112***	0.197***	0.093***	0.122***	0.007***	0.177**	0.333†	0.045***	0.219***	0.002***	0.019***	0.034***
Random Effects (country-level)												
Female (std.dev.)	0.300		0.524					0.409	0.605			
	(0.185)		(0.204)					(0.321)	(0.264)			
Constant (std.dev.)	0.600	0.739	0.676	0.551	0.647	0.724	0.680	0.755	0.462	0.752	0.457	0.262
	(0.119)	(0.166)	(0.165)	(0.137)	(0.175)	(0.284)	(0.208)	(0.236)	(0.171)	(0.173)	(0.145)	(0.128)
Diagnostics												
Likelihood ratio test rel. to logit, p-value	0.0000	0.0000	0.0000	0.0000	0.0002	0.0211	0.0022	0.0008	0.0006	0.0000	0.0049	0.0725
N	2856	1070	2047	1114	658	538	602	917	755	1200	1837	2013
AIC	2041	724	1549	935	670	312	456	573	753	854	857	1150
BIC	2130	794	1633	1006	737	372	517	645	823	930	934	1228

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$; † $p < 0.15$

3.6.3 Conclusions

In general, overeducation is more prevalent in industries with higher share of ISCO major groups 4–9 (or below-tertiary occupations) or higher share of tertiary educated employees. Countries fall into one of three groups by the former relationship: (1) industries form two clusters: low overeducation–low share of below-tertiary occupations and high overeducation–high share of below-tertiary occupations, (2) industries form a continuous cloud along a positively sloped line and (3) industries appear along a horizontal line, suggesting no relationship between the two variables.

Controlling for these macro-level factors, the industry with the highest overeducation risk is financial services, while the most and the least affected industries differ considerably for females and males.

The exposure to overeducation in different industries relative to manufacturing reflects clear consequences of traditional association of men and women with different industries. However, relative positions of men and women within a particular industry appear to be less related to this tradition.

Industry-specific models show that few factors have strong influence in most industries, with the most consistent performance shown by tenure. Thus, it appears that there is no clear translation from the effects in the general model to industry-specific models.

3.7 Summary

This chapter dealt with different factors influencing the risk of overeducation.

Its most important results are:

- Personality is an important factor affecting overeducation risk
- For females, personality has more explanatory power than income-based ability
- Achievement orientation and openness to experience act as a buffer against overeducation, while social orientation acts as a catalyst
- Females are more exposed to overeducation than males, except for narrow industries like construction and public administration
- Female immigrants are much more hurt than male immigrants, relatively to the fully natives of the same sex
- Immigrants from CEE and Former Soviet Union have the highest risk premium among individuals with different immigrant background

- Increasing size of the labour market – whether measured by domicile or by the employing firm – improves the chances of finding a matching job
- There is no universal effect from social class to the exposure to overeducation risk
- Studies in parallel to work, lengthy unemployment spells and informal employment push individuals towards accepting mismatched jobs
- Temporal dynamics of overeducation and its distribution across industries is well explained by the macro-level variables measuring supply and demand – the share of tertiary graduates and the share of ISCO 1–3 occupations; unemployment explains only temporal dynamics and increases overeducation risk
- Education and health are the only fields whose graduates consistently show the lowest risk of overeducation in all four country groups (Western, Eastern, Northern and Southern Europe)
- Only in Western Europe, there are signs of overproduction of economists in terms of difficulties in finding a matched job, as compared to graduates in other fields
- Eastern Europe is a unique country group having the overeducation risk of engineers 75 per cent higher than that of economists
- Compared to manufacturing industry, overeducation risk is higher in finance, administrative activities, accommodation and public administration
- The relative risk of overeducation in different industries is related to traditional association of males and females with particular industries
- Tenure has a strong effect of decreasing overeducation risk, which is consistent over industries, while other factors do not show such consistency

4 EFFECTS OF OVEREDUCATION

This chapter studies some of the effects of overeducation on the labour market position of an individual. As already discussed in Section 2.1.2, overeducation has many negative effects on different aspects of the current and future jobs. Most of this influence comes through affecting the psychological state of the individual.

In this chapter, I focus on two closely interrelated aspects of the current job determined by the psychological comfort of the individual with what he or she encounters at the workplace: job satisfaction (Section 4.1) and propensity to quit (Section 4.2). While overeducation is likely to play an important role in affecting both aspects, I also study the role of other important factors. In particular, I study the role of primary job-related factors in determining the level of job satisfaction, focusing on tertiary graduates in Europe.

When analysing the propensity to quit, I consider the mediating effects from educational track, aiming at understanding whether the role of vocational education is positive or negative for this particular aspect of the labour market dynamics of an individual. This is a case study of Switzerland, which is famous for its highly developed vocational education system and is frequently suggested as a role model for other countries.

Section 4.3 then looks on the other side of LEMS and studies choices related to getting a doctoral degree. The focus of the section is on the motivation of graduates to continue studies at doctoral level. The effect on the type of motivation from the overeducation status is analysed. The Baltic countries (Estonia, Latvia and Lithuania) are taken as a case study.

4.1 Job Satisfaction Determinants of Tertiary-Educated Employees in European Countries

In this section, I study the determinants of job satisfaction of tertiary-educated employees aged 20–60. Particular attention is given to factors specific to higher education: whether masters are more satisfied than bachelors are and whether the overeducated are less satisfied than the well matched are. This section, thus, contributes to the literature by providing rich empirical cross-sectional evidence on the differences in job satisfaction determinants in recent years.

The section is structured as follows. Firstly, the existing literature on job satisfaction is briefly reviewed. Then data and methods are described. The presentation of results follows. Sections 4.1.4 and 4.1.5 provide discussion and conclusions, respectively.

4.1.1 Literature on the Causes and Effects of Job Satisfaction

Kalleberg (1977, p. 126) defines ***job satisfaction*** (JS) as "an overall affective orientation on the part of individuals toward work roles which they are presently occupying" and views it as the result of an interplay between the values workers attach to job characteristics and the extent to which these values are satisfied.

JS is a very important characteristic of an employee. It is known as one of the main determinants of the intention to quit, and hence it was heavily studied in occupations with high turnover like nursing (Acker, 2004; Chen & Johantgen, 2010; Cortese, Colombo, & Ghislieri, 2010; Harris, Winskowski, & Engdahl, 2007; Parry, 2008). Several studies of JS of the general population were also performed (for Europe, see, e.g., Millán, Hessels, Thurik, & Aguado, 2011; Mora, García-Aracil, & Vila, 2007; Poggi, 2010; Skalli, Theodossiou, & Vasileiou, 2008). Moreover, Carless and Arnup (2011) found that JS increases after a job change, which means that workers take into account expected JS when choosing among several job proposals.

At the theoretical level, the ***job demands–resources model*** (see Bakker & Demerouti, 2007, for a review) views the workplace (in its broad meaning) as a field where job demands meet with available resources. When not enough resources are available to satisfy job demands, job strain occurs. Resources, however, are used not only for satisfying demands, but also as a driver for employee's motivation. Hence, this framework argues about the importance of support activities like help from co-workers or allowing employees to manage their working time for increasing satisfaction with job. At the same time, excessive job demands (e.g., leading to regularly not being able to finish all tasks on time) or insufficient resources (e.g., not being paid appropriately, but also having no opportunity for advancement and performing a considerable amount of repetitive tasks) lead to stressful situations and, thus, decrease job satisfaction.

Kalleberg (1977) distinguishes among six groups of values: intrinsic (associated with the task itself), convenience (comfort-providing facets external to the task), financial, relationships with co-workers (satisfaction of social needs, making friends among employees), career opportunities and resource adequacy (incl. support from co-workers). In his survey, administered in the US in 1972–1973, he found that the most important factors affecting job satisfaction are intrinsic and financial, while career and resource adequacy are of moderate importance.

While he did not find that relationships with co-workers significantly affect JS, this does not mean that co-workers are irrelevant to it. Indeed, in his definition of resources, he extensively mentions the help, authority, information, supervision and competency of co-workers, and these are found to influence JS. Other studies found the level of social support from co-workers to be significant in many occupations (Alexander, Lichtenstein, Oh, & Ullman, 1998; Brough & Frame, 2004; Cortese, Colombo, & Ghislieri, 2010; Ducharme & Martin, 2000; Roxburgh, 1999), supporting its role in being a buffer against high job demands, preventing job strain and affecting motivation and productivity, thus, supporting the job demands–resources model.

Skalli, Theodossiou and Vasileiou (2008) consider the importance of five job-related facets of job satisfaction in 10 European countries using European Community Household Panel data, 1994–2001 (their data do not include any CEE country, but include Scandinavian and Southern European countries). They find that in all countries, the most important determinant is the type of work. The other important factors go approximately (i.e., with some minor country-specific differences) in the following order: earnings, working conditions, job security and working times. Clearly, this is in line with the findings of Kalleberg (1977).

Vila, García-Aracil and Mora (2007) use data from Careers after Higher Education, a European Research Survey for tertiary graduates aged 26–35 from seven countries (again, no CEE country appears in the sample). They find that only five job determinants significantly influence job satisfaction (in decreasing order of effect size): career, opportunity to pursue own ideas, good social environment, use of acquired knowledge and skills, and challenging tasks. Notably, financial compensation variables were not included as explanatory variables.

Studies differ on the extent of homogeneity of the directions of the effects from job satisfaction determinants in different European countries. For instance, Mora, García-Aracil and Vila (2007) find that countries are quite homogeneous, while Díaz-Serrano and Cabral Vieira (2005) report considerable heterogeneity across countries.

4.1.2 Data and Methods

Of all ESS rounds available at the time of writing this text, only two contain the job satisfaction variable: round 3 and round 5. While round 3 was used in job-satisfaction models (Lange, 2012), it contains very little information about respondents' perceptions of their current job. On the contrary, round 5 introduces a block of variables that directly measure respondents' attitudes to various aspects of their current job (see Table 4-1), which I will refer to as ***primary job-related factors***. I consider these variables as important control variables and include them in regressions. Unfortunately, nearly all of them are missing in round 3, so it is impossible to compare their effects across time. As a result, data from round 5 only are used in this study.

Table 4-1 – Job-Related Factors Affecting Job Satisfaction from ESS Round 5

Group	Variable Name	Original Description
Content	Variety in work	There is a lot of variety in my work
	Job requires learning	My job requires that I keep learning new things
Effort	Job requires to work hard	My job requires that I work very hard[a]
	Work overload	I never seem to have enough time to get everything done in my job
Risks	Health at risk at work	My health or safety is at risk because of my work
	Risk moving to a less interesting job	I may have to move to a less interesting[b] job in my organisation in the next 12 months
Compensation	Career opportunities	My opportunities for advancement are good
	Paid appropriately	Considering all my efforts[c] and achievements in my job, I feel I get paid appropriately
	Wage depends on effort	My wage or salary depends on the amount of effort I put in my work
	Employment guarantee	My job is secure[d]
Support	Help from co-workers	I can get support and help from my co-workers when needed
	Can manage own work time	I can decide the time I start and finish work

[a] "Hard" refers to intensity or long hours.
[b] Less interesting to the respondent in their own opinion.
[c] "Effort" in the sense of try more than minimum.
[d] "Secure" in the sense of an actual or implied promise/likelihood of continued employment with that employer.

The current job satisfaction variable in ESS measures the respondent's answer to the question "How satisfied are you in your main job?" on the 0–10 scale. Two classes of econometric methods can be applied to such dependent variable. The first one is ordered logistic regression, but this would complicate the analysis, as large tables would have to be produced for each of the eleven categories to show marginal effects. In principle, one could use stereotype logistic regression (Anderson J. A., 1984) to combine the categories of the dependent variable that are not distinguishable by respondents and, consequently, reduce the number of categories. I performed this check and found that the reduction is not big enough to improve the readability of model output (eleven categories were reduced to seven).

Thus, I proceed with the second option – running linear models, which can be justified by having a large number of dependent-variable categories. To be able to make predictions that are always in the interval $[0, 10]$, I employ two-limit tobit regression (see Section C.6) with the lower limit of 0 and the upper limit of 10.

I delimit the analysis to currently employed tertiary-educated individuals aged 20–60. The inclusion of primary job-related factors further reduces the sample size, as they are defined only for employees (while a significant proportion of respondents in the ESS dataset are self-employed). Consequently, removing from the analysis countries with less than 120 observations, the estimation sample consists of 13 countries.

The estimation strategy is as follows. Firstly, I concentrate on the effects of education level; the aim is to study whether masters are more or less satisfied with their current job than bachelors. For that, I run the model (to be described shortly) separately on male and female respondents. Firstly, separate regressions are run for each country. In each case, primary job-related factors are added sequentially to make sure that the previously added factors have stable effects in terms of sign and significance. Secondly, countries are grouped based on the size of the effect from education level (positive, small/no effect, negative). Finally, these groups are divided into more homogeneous (in terms of the effects of other regressors) subgroups, adding interactions with countries where significant. Tobit regressions are then run on the resulting country groups.

Secondly, differences in job satisfaction determinants between bachelors and masters are investigated. For that, the samples of each country are divided into bachelors and masters, and the same model is run on this level. Countries are then grouped based on the similarities of effects. The model is then run on the level of country groups.

In regressions on country groups, I employ the sandwich estimator appropriate for clustering of observations (see Section C.7); in particular, the estimators I use assume that observations are uncorrelated across countries, but can be correlated within countries.

The model consists of the following variable groups:

- Primary job-related factors
- Other job-related factors (tenure and its square, overeducation dummy, supervising position dummy, public firm dummy)
- Firm size
- Immigrant background
- General demographics (age and its square, disability dummy); where relevant, sex or education level is added as an explanatory variable

For readability and analysis purposes, all primary job-related factors were re-coded into dummies. Originally, they are coded on a 1–4 or 1–5 scale. In case of four categories, categories 1–2 were coded as 0 and categories 3–4 as 1. In re-coding five-level variables, categories 1–2 were coded as 0 and categories 3–5 as 1.

4.1.3 Results

Firstly, consider the effects from the level of education (see Table 4-2).

The most surprising result here is that there is *no* country where females with a master's degree are more satisfied with their current job than bachelors. The group Belgium-Bulgaria-Denmark-Spain-UK has a strongly negative effect from education level, while the group Greece-Israel-Sweden has a moderately negative effect (which is not far from being statistically significant, with the p-value of 0.17). The other five countries (Germany, France, the Netherlands, Norway and Poland) show no difference in job satisfaction between females with different education levels.

The job satisfaction level of males is affected by education level to a greater extent. For respondents from Belgium, Greece, Israel and Poland, the effect is strongly negative (comparable to the similar effect for females), but for the group Germany-Spain-UK, it is 1.5 times greater in absolute terms. The Netherlands and Norway show a slightly positive return on education, while the group Bulgaria-Denmark-Sweden shows a strongly positive effect. France is an outlier with an extremely positive effect from education level[18].

When respondents are divided into bachelors and masters, countries from the same geographical region tend to show similar effects of explanatory variables. Hence, I consider four groups of countries:

- Northern Europe (Denmark, Norway, Sweden)
- Southern Europe (Israel, Greece, Spain)
- The Netherlands and the UK
- The remaining Central Europe (Belgium, Bulgaria, Germany, France, Poland)

The Netherlands and the UK were separated from the Central Europe group to keep the latter sufficiently homogeneous. See Table 4-3 for main results.

Besides education level, another factor that is specific for tertiary graduates is overeducation. One would expect that the overeducated would have a considerably lower job satisfaction than the well-matched tertiary graduates (because their potential is useless at work). Surprisingly, there are no statistically significant effects from overeducation for bachelors (except for the Netherlands-UK group), while for masters, the effects are strongly negative in all country groups

[18] The effects of other variables on the job satisfaction of French men also differ from those observed for Bulgaria, Denmark and Sweden. Hence, I ran a separate regression on them.

Table 4-2 – Marginal Effects after Tobit Regressions of Job Satisfaction: Education Level Effects

Females	**BE-BG-DK-ES-UK**	**GR-IL-SE**	**DE-FR-NL-NO-PL**
Education Level, rel. to Bachelor			
Master	–0.339***	–0.154	0.001
Regression Fit Indicators			
N	653	397	576
McFadden's pseudo R^2	0.1175	0.1153	0.1152

Males	**DE-ES-UK**	**BE-GR-IL-PL**	**NL-NO**	**BG-DK-SE**	**FR**
Education Level, rel. to Bachelor					
Master	–0.520***	–0.342*	0.097***	0.316*	1.849***
Regression Fit Indicators					
N	314	374	269	238	66
McFadden's pseudo R^2	0.1418	0.0974	0.1296	0.1644	0.3025

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$; † $p < 0.15$
Standard errors adjusted to account for intra-country correlations.
Only effects of the level of education reported.

except for Northern Europe. Looking on the absolute size of marginal effects, one can observe that job satisfaction of masters is more sensitive to overeducation than that of bachelors.

Now consider the primary job-related factors.

Factors that in all cases increase job satisfaction are variety in work, job requires learning, career opportunities and appropriate monetary compensation. In most cases, positive effects are also found from help from co-workers[19] and ability to manage own working time[20]. Factors that decrease job satisfaction in most cases are work overload[21], health at risk at work[22] and risk moving to a less interesting job[23].

In all country groups, masters are much more sensitive than bachelors to career opportunities and less sensitive to the risk of moving to a less interesting job and appropriate monetary compensation, the latter especially pronounced in Northern Europe. In all country groups except for Northern Europe, masters are much more sensitive to variety in work, while in Central and Southern Europe, they are much more sensitive to both content-related factors.

Table 4-4 reports the top five most important primary job-related factors that affect job satisfaction for each pair of country group and education level. For masters, variety in work is the first or the second most important factor in all four country groups, while for bachelors, job content factors are in top-three everywhere except for Southern Europe.

[19] Not significant for bachelors in Central Europe and the UK; negative in the Netherlands.
[20] Not significant for masters in the Netherlands & the UK and negative for masters in Southern Europe.
[21] An increasing effect is found for masters in Northern and Southern Europe.
[22] Not significant for masters in Southern Europe.
[23] Positive for bachelors in Central Europe and not significant for masters in the Netherlands & the UK.

Table 4-3 – Marginal Effects after Tobit Regressions of the Job Satisfaction of Tertiary Educated Employees Aged 20–60, by Country Group and Education Level

	Central Europe		Netherlands & UK		Northern Europe		Southern Europe	
	Bachelors	**Masters**	**Bachelors**	**Masters**	**Bachelors**	**Masters**	**Bachelors**	**Masters**
Job-Related Factors: Content								
Variety in work	0.800***	0.843***	0.501**	1.071***	1.595***	1.311***	0.293**	0.914***
Job requires learning	0.264**	0.427***	0.947***	0.256***	0.653***	0.663***	0.335***	0.897***
x Bulgaria		*–0.479****						
x Germany		*–0.469****						
x Spain								*–1.297****
Job-Related Factors: Effort								
Job requires to work hard	0.216	0.362***	0.927***	0.554	–0.181***	–0.239***	0.320**	–1.075***
x Belgium	*–0.336†*							
x Bulgaria	*–0.405****							
x Spain								*1.519****
Work overload	–0.343***	–0.384***	–0.668***	–0.403***	–0.445**	0.207***	–0.302***	0.424**
x Belgium		*1.174****						
x Spain							*0.633****	*–0.912****
x Sweden						*–0.643****		
Job-Related Factors: Risks								
Health at risk at work	–0.552***	–0.813***	–0.927***	–0.324***	–0.388***	–0.618***	–0.575*	0.176
x United Kingdom			*0.963****					
x Spain							*1.120****	
x Denmark						*2.685****		
Risk moving to a less interesting job	1.040***	–0.578*	–0.504***	–0.112	–0.895†	–0.831***	–1.128***	–0.824***
x Greece							*1.691****	*1.859****

*** p < 0.01; ** p < 0.05; * p < 0.10; † p < 0.15

Standard errors adjusted to account for intra-country correlations. Country fixed effects not reported.

Country groupings: Central Europe (Belgium, Bulgaria, Germany, France, Poland); Northern Europe (Denmark, Norway, Sweden); Southern Europe (Spain, Greece, Israel).

Table 4-3 (cont.)

	Central Europe		Netherlands & UK		Northern Europe		Southern Europe	
	Bachelors	Masters	Bachelors	Masters	Bachelors	Masters	Bachelors	Masters
Job-Related Factors: Compensation								
Career opportunities	0.523***	0.687***	1.107***	1.135***	0.339***	0.494***	0.562***	0.669***
Paid appropriately	0.916***	0.902***	0.675***	0.610***	0.613***	0.341***	0.448†	0.459***
x Norway						*–0.335****		
Wage depends on effort	0.173†	–0.435**	–0.745***	0.493***	0.000	–0.229*	–0.341***	0.313**
x Belgium		*0.741****						
x United Kingdom				*–1.160****				
x Greece							*0.588****	
x Israel								*–0.353****
Employment guarantee	0.650**	0.450**	–0.144**	0.104	0.131	0.348***	–0.238*	0.525***
x Israel							*0.936****	
Job-Related Factors: Support								
Help from co-workers	0.116	0.431***	0.704**	–0.451***	0.227†	0.363*	0.646***	0.593**
x Belgium	*–0.302**							
x France	*–0.674***							
x United Kingdom				*0.683****				
x Greece							*–1.326****	
x Norway						*–1.091****		
Can manage own work time	0.518***	0.572***	0.381***	0.144	0.139**	0.113***	0.804***	–0.760**
x Bulgaria	*–1.566****							
x Poland	*–2.111****							
x Belgium		*–0.966****						
x Sweden					*–0.661****			
x Spain							*–0.995****	*1.602****

*** p < 0.01; ** p < 0.05; * p < 0.10; † p < 0.15

Standard errors adjusted to account for intra-country correlations. Country fixed effects not reported.

Table 4-3 (cont.)

	Central Europe		Netherlands & UK		Northern Europe		Southern Europe	
	Bachelors	Masters	Bachelors	Masters	Bachelors	Masters	Bachelors	Masters
Job-Related Factors: Other								
Tenure	–0.012	–0.005	–0.008	0.048***	–0.027	0.007	0.035*	0.053***
Tenure^2/100	0.047	0.050	0.021	–0.064	0.060	0.041	–0.050	–0.282***
Overeducated	0.247	–1.391***	–0.381***	–0.500***	–0.133	–0.423	–0.229	–0.408***
x Bulgaria	*–1.841****							
x Poland	*–3.303****							
x Israel								*0.852****
Supervising position	0.242*	0.190***	–0.184*	0.012	–0.146***	–0.197***	–0.272***	–0.283†
x Bulgaria		*–0.284****						
x Poland		*–0.307****						
x Sweden					*0.307****			
x Israel								*0.558****
Public firm	0.027	0.360*	0.721***	0.791***	0.415***	–0.408***	–0.036	0.085
x France	*1.332****							
x United Kingdom			*–0.823****	*–0.798****				
x Denmark					*–0.701****	*1.102****		
Immigrant Background								
Minority	0.699**	–0.951***	–0.748***	–1.215***	–0.082	–0.194	0.363***	–0.027
One parent immigrant	0.083	–0.506***	–0.701***	–0.216†	–0.119	0.518**	0.645***	–0.942***
Both parents immigrants	–0.052	–0.418	0.436	0.542***	–0.484	–0.598	0.130	–1.013***
CEE or FSU immigrant	1.444***	0.416	–1.039*	–0.602†	0.299	–0.008	0.377	–1.534***
LAA immigrant	–0.260	0.932**	0.049	0.694***	0.059	0.367†	–0.567***	–0.562
Other European immigrant	0.242	0.734	0.118*	–0.156***	–0.339	–0.239*	0.080	–0.550†

*** p < 0.01; ** p < 0.05; * p < 0.10; † p < 0.15

Standard errors adjusted to account for intra-country correlations. Country fixed effects not reported.

FSU stands for "Former Soviet Union"; LAA stands for "Latin America, Africa and Asia."

Table 4-3 (cont.)

	Central Europe		Netherlands & UK		Northern Europe		Southern Europe	
	Bachelors	Masters	Bachelors	Masters	Bachelors	Masters	Bachelors	Masters
General Demographic Characteristics								
Age	–0.023	–0.051	0.071***	0.026	–0.118**	–0.104	–0.087*	–0.018
$Age^2/100$	0.043	0.060	–0.058†	–0.030	0.153**	0.128	0.129**	0.037
Female	0.511**	0.487***	–0.421***	–0.159†	–0.139*	–0.018	0.392***	0.448***
x Germany	*–1.102****							
x Belgium		*–0.797****						
x France		*–0.841****						
x Denmark					*1.162****			
Disabled	0.257*	–0.826***	–0.464***	–0.598***	0.124†	0.335***	0.112	0.667***
x Germany	*–0.795****							
x France	*–1.374***							
x Bulgaria		*1.355****						
x United Kingdom				*0.761****				
x Norway					*–0.363****			
x Spain							*–1.161****	
x Denmark						*–0.370****		
Firm Size, rel. 25–99 employees								
< 10	0.111	–0.009	0.570***	0.524***	0.202	–0.074	–0.076***	0.100
10–24	0.272***	–0.284	0.055	0.716**	0.161***	0.164	0.510***	–0.023
100–499	0.123	–0.489**	–0.032	0.664***	–0.162***	–0.249	0.407†	0.302
> 500	–0.196	–0.209†	–0.311	–0.037	–0.059	–0.036	0.313	0.231
Regression Fit Indicators								
N	432	616	229	215	466	281	405	243
McFadden's pseudo R^2	0.1214	0.1431	0.1244	0.1538	0.1001	0.1202	0.0936	0.1262

*** p < 0.01; ** p < 0.05; * p < 0.10; † p < 0.15

Standard errors adjusted to account for intra-country correlations. Country fixed effects not reported.

Table 4-4 – Ranking of the Five Most Important Job-Related Effects on Job Satisfaction, by Country Group and Education Level

Rank	Central Europe		Netherlands & UK	
	Bachelors	Masters	Bachelors	Masters
1	Risk moving to a less int. job	Paid appropriately	Career opportunities	Career opportunities
2	Paid appropriately	Variety in work	Job requires learning Health at risk at work [a] Job requires to work hard	Variety in work
3	Variety in work	Health at risk at work	Wage depends on effort	Paid appropriately
4	Employment guarantee	Career opportunities	Help from co-workers [a]	Wage depends on effort
5	Health at risk at work	Risk moving to a less int. job Can manage own working time	Paid appropriately Work overload	Help from co-workers

[a] Only for the Netherlands. For the UK, this factor is considerably less important.

Rank	Northern Europe		Southern Europe	
	Bachelors	Masters	Bachelors	Masters
1	Variety in work	Variety in work	Risk moving to a less int. job	Job requires to work hard
2	Risk moving to a less int. job	Risk moving to a less int. job	Can manage own working time	Variety in work Job requires learning
3	Job requires learning	Job requires learning	Help from co-workers	Risk moving to a less int. job
4	Paid appropriately	Health at risk at work	Health risk at work Career opportunities	Can manage own working time
5	Work overload	Career opportunities	Paid appropriately	Career opportunities

Factors combined into one group if their absolute effects differ by not more than 0.020.

On the contrary, career opportunities and appropriate monetary compensation never occupy positions higher than the fourth in Northern and Southern Europe, contrary to what theory would predict. Career opportunities significantly affect job satisfaction everywhere, but only in the Netherlands and the UK, they are the most important factor, while in the other country groups, they are never higher than the fourth position.

Apparently, what moves these two compensation-related factors down the ladder of importance is the inclusion of job risks, which were absent from the models of job satisfaction of the tertiary-educated (see Section 4.1.1), but which in most cases[24] are in the top-three. In general, thus, while both Kalleberg (1977) and Skalli, Theodossiou and Vasileiou (2008) found that the first two most important factor groups are content and compensation, my results show that job-related risks are placed between them, so that the relevant order is (1) content, (2) risks and (3) compensation.

Females (both bachelors and masters) are more satisfied than males in Central and Southern Europe, while less in the other two groups. Moreover, except for Southern Europe, sex effect size decreases with education level, meaning that the higher education level is, the less pronounced the differences in job satisfaction across sex are.

Age effect is found only for bachelors, in all country groups except for Central Europe. The effect is U-shaped in Northern and Southern Europe and inverse U-shaped in the Netherlands & the UK. I also run regressions with interactions added between sex and age and age-squared (not reported). A U-shaped relationship for bachelor males was found everywhere except Central Europe and for master males in Central Europe and the Netherlands & the UK. For females, however, a U-shaped effect is observed only for bachelors in Northern Europe (note also that the absolute effect from sex in Northern Europe is the smallest across all country groups, meaning a better gender equality with respect to job satisfaction). On the contrary, in the Netherlands & the UK (both bachelors and masters) and for bachelors in Southern Europe, an inverse U-shaped effect is observed for females.

Masters are more sensitive to disability and/or serious health problems. However, disability decreases the job satisfaction of masters only in Central Europe and the Netherlands, while it significantly increases it for masters in the UK and the other two country groups.

Surprisingly, holding a supervising position increases job satisfaction only in Central Europe, for both bachelors and masters. In the other cases[25], the effect is negative.

---

[24] Except for the Netherlands and the UK.

[25] Small and not significant for masters in the Netherlands & the UK.

Finally, bachelors in all country groups prefer to work for small companies, while for masters such effects are observed only in Central Europe and the Netherlands & the UK.

4.1.4 Discussion

Investment in further education should pay out, e.g., via higher wages, lower risk of unemployment and higher job satisfaction. This would be a typical conclusion from theory. In practice, though, one can observe that individuals are frequently unable to find a job that they would consider a good match.

The most obvious consequence of such failure is overeducation. While it does not affect the job satisfaction of bachelors, masters become highly dissatisfied if they work on positions inappropriate for their level of education. Nevertheless, even after controlling for overeducation, I showed that masters are frequently less satisfied with their jobs than bachelors.

One of the possible explanations is that masters strive for higher status in the organisation, so that they receive benefits that distinguish them from employees with only bachelor degree. Indeed, while both career opportunities and appropriate wages increase the job satisfaction of bachelors and masters, the latter are more career-oriented and less wage-oriented than the former. Evidence on lower returns of the master's degree means that its holders very often have expectations that are much higher than what they actually face in their job.

Career and wages, though, in many cases are not the most important determinants of job satisfaction. Very frequently, it is job content that has the highest influence on employee's contentment, especially in Northern and Southern European countries. In other words, even a highly paid job with perfect opportunities for career growth could distract employees if it is monotonous or stagnates one's personal progress by not requiring learning anything new. As a further proof of this claim, recall the negative returns to working in a supervising position, observed in all countries except for the Central European. Moreover, employers should decrease risks associated with the job, as in many cases, they are more important than compensation for employees.

The results also support the job demands–resources model in that support activities are important to mitigate stressful situations in the job and, consequently, increase job satisfaction. Nevertheless, resources (including content and compensation groups) are the first thing employers should concentrate on.

Considerable attention has long been paid to gender effects in empirical literature. I find that, with minor exceptions, women are actually more satisfied with their job than men. Moreover, gender gap in job satisfaction decreases with higher levels of education. One should also note that the difference in job satisfaction between men and women is very small in Northern Europe.

4.1.5 Conclusions

Based on the results of this section, recommendations can be made for both employers and employees. Employers should concentrate on providing jobs with attractive content and lower risk. They also should keep in mind that masters are more responsive to career opportunities, while bachelors look more on wages.

Before they choose to continue studies at master's level, employees should realise that their actual gains in the labour market could be lower than their expected gains, whatever claimed by their universities. The more realistic graduates' expectations are, the lower the gap in job satisfaction between bachelors and masters should be.

Further studies on this topic should include the field of study variable to check whether the supply of graduates from a field affects the job satisfaction of a graduate from this field. Unfortunately, ESS round 5 does not include this information.

4.2 Overeducation and the Propensity to Quit: Does Vocational Education Help?

Education and skills mismatch (respectively, overeducation and overskilling) has been recently reviewed by the International Labour Organisation (2013a) and included as a global risk by the World Economic Forum (2014). The latter, in particular, mentioned Swiss education system (along with German) as an example for other countries of how to build vocational education in their education systems. In 2010, however, Switzerland and Germany had only 7th and 8th smallest overeducation rates in Europe, respectively (International Labour Organisation, 2013b). Thus, vocational education system *per se* does not appear to guarantee low overeducation incidence.

Nevertheless, I decided to take Switzerland as a country with well-developed vocational education and study whether the propensity to quit the current employer in six months depends on overeducation and whether those choosing a (semi-)vocational educational track are better off in terms of smaller propensity to quit or smaller effect from overeducation on the propensity to quit.

At this moment, I have to clarify why it makes sense to study propensity to quit at all. It is typically studied in professions with high turnover, such as nursing (Hayes, et al., 2012). Here, I am not binding the analysis to any particular profession. Rather, I am studying the exposure of an individual to a potentially negative experience of unemployment, which will happen if in case of quitting he or she does not have a ready position elsewhere.

In this context, overeducation plays a role of a negative factor in a workplace that may trigger the decision to quit. However, the study will also add to the literature on the debate of whether overeducation is real. If the person is theoretically overeducated but effectively well-matched then he or she should have the same propensity to quit as a non-overeducated. However, if it is not the case and the study shows that the overeducated are more likely to quit (and it will indeed show it), then the position of those who believe in perfect labour markets might be impaired. Then we have a problem that a particular state of a person in the labour market leads to a potentially negative outcome, and we cannot say that it is simply an error of measurement.

The study proceeds as follows. The next subsection provides the descriptive statistics on education and key labour market variables in the data. The following two subsections build (1) a model of the propensity to quit and (2) models of three variables highly significant and sizeable in the former model, respectively. The last subsection concludes.

4.2.1 Education Statistics in the Data

I use a subset of panel data from TREE. It is a longitudinal study of young people in Switzerland, starting from their participation in the PISA survey in 2000 at the moment when they left compulsory school. The sample of approximately 6000 young people was then followed up eight times, annually from 2001 through 2007 and again in 2010. I use data from the last three waves, gathered in 2006, 2007 and 2010. This is the time when respondents transitioned from secondary education to tertiary tracks and to the labour market.

Among the high level of detail related to the education, apprenticeship and employment history, TREE contains rich data on the personality of the respondent. Importantly for the current study, it contains a set of variables describing the inherent tendency of the respondent to quit.

I apply the following restrictions to the sample. I take only those individuals having a job at the time of participating in the survey. Having a fixed-term contract accurately determines the chances that the worker will continue being employed in the current job in 6 months. Hence, I exclude individuals with fixed-term contracts from analysis. This study concentrates on secondary and tertiary education, which puts education levels lower than secondary out of scope. Thus, I exclude cases where respondents remained at lower secondary education level (i.e., at the level where they filled the PISA survey in 2000). I also exclude very few cases where

Table 4-5 – Descriptive Statistics on Education Level and Type

Year	% Cases		% Vocational secondary education					Voc. tert.
Education	Secondary	Tertiary	Secondary	Tertiary			Total	as % of
				Academic	Vocational	Total		tertiary
2006	100%		97%				97%	
2007	91%	9%	97%	52%	100%	63%	94%	22%
2010	55%	45%	96%	47%	100%	56%	78%	16%
Total	**77%**	**23%**	**97%**	**48%**	**100%**	**57%**	**87%**	**17%**

"Voc. tert." abbreviates vocational tertiary education.

respondents already have tertiary education in 2006 and those where respondents pursued vocational tertiary education after general secondary education (in the whole sample).

The size of the sample used in this paper (after leaving out cases with missing values of control variables) is 2336 cases (1464 individuals). Of these, most individuals (39 per cent) are observed in the last wave only, 17 per cent in all three waves, around 10 per cent in each pair of two consecutive waves, while other 24 per cent appear in the data in other patterns. Most individuals (97 per cent) were aged 14–16 in 2000, the rest were 17 or 18 years old at that time. Women form 62 per cent of the sample.

Table 4-5 provides details on the level and type of education observed in the sample in each of the three waves. Only 9 per cent of the sample completed tertiary education by 2007, but it rose to nearly half of the sample by 2010. Of those who remained at secondary level, most had vocational degrees, but the overall share of vocational secondary education in the sample dropped to below 80 per cent by 2010. Around 20 per cent graduated with a vocational tertiary diploma, meaning that most of those who continued to tertiary level chose an academic track. Of the latter, only half had vocational secondary background.

When interpreting these figures, one has to bear in mind that the sample consists only of employed respondents. Thus, the declining share of vocational secondary education in the sample reflects the fact that the individuals graduating from a general secondary school choose to delay their entry in the labour market, contrary to the behaviour of those graduating from a vocational secondary track.

The primary variable of interest in the study is whether the respondent considers it likely that (s)he does not work with the current employer after 6 months after the interview. Table 4-6 shows that of the respondents with secondary education, graduates from a general track are more likely to quit, but also more frequently work on evenings and weekends than graduates from a vocational track. At tertiary level, the likeliness to quit in half a year is 14 per cent regardless of whether the track is tertiary or secondary. At the same time, at both secondary and

Table 4-6 – Sample Averages of Labour Market Variables by Education Level and Type

	Secondary			Tertiary			Academic Tertiary	
	General	Vocational	Total	Academic	Vocational	Total	Gen. Sec.	Voc. Sec.
Likely to quit job in 6 months	25%	13%	13%	14%	14%	14%	14%	14%
Overeducated	0.0%	0.3%	0.3%	8%	44%	14%	8%	8%
Overskilled	30%	19%	20%	14%	14%	14%	15%	12%
Frequently thinks about changing job	25%	22%	22%	20%	12%	18%	19%	21%
Job satisfaction	5.1	5.2	5.2	5.4	5.5	5.4	5.4	5.3
Works on evenings or weekends	46%	30%	31%	29%	32%	29%	31%	26%

The range of the job satisfaction variable is [1,7]. Other variables are dummies. The overskilling variable defined only in 2010. The table does not include the division of vocational tertiary into a subset with general secondary and vocational secondary background, as all vocational tertiary graduates have vocational secondary background (see Table 4-5).

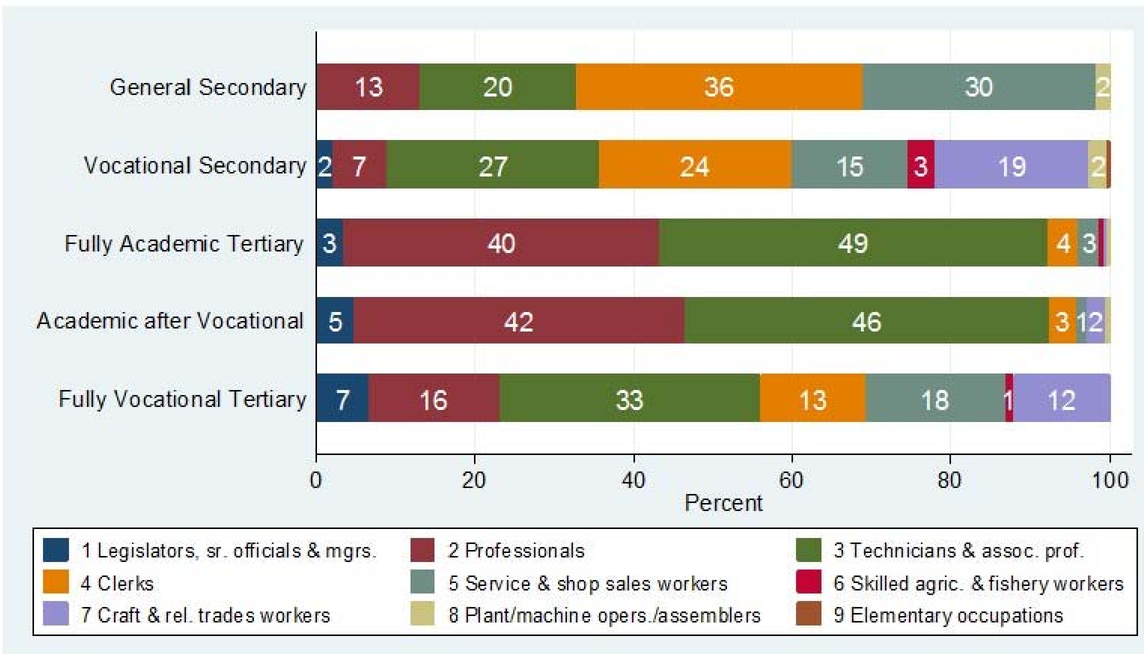

Figure 4-1 – Occupational Structure in the Sample by Educational Track

tertiary levels, vocational track graduates less frequently think about quitting than general or academic track graduates do. This is surprising, as the intent to quit seems to be positively related to the probability to quit at secondary level but unrelated to it at tertiary level, while the empirical literature strongly favours positive correlation. Pairwise correlations between the two variables are highly significant and positive at secondary level (around 0.47) and for academic tertiary graduates (0.52 with general secondary education and 0.67 with vocational secondary education) but small (0.14) and insignificant (p-value of 0.19) for vocational tertiary graduates.

Overeducation is defined here using the normative measure, which is based on the comparison of skills needed in the major occupations groups of the ISCO[26] and provided at different education levels. Secondary graduates are overeducated if they are employed in elementary occupations (ISCO major group 9). Table 4-6 shows that overeducation is rare at this education level and observed only among those graduating from a vocational track. Tertiary graduates are overeducated if they are employed in the major groups 4 through 9 (clerks and below). At that level, the incidence of overeducation is below 10 per cent for academic degree holders but nearly 50 per cent for vocational degree holders. Nevertheless, the share of overskilled (subjectively assessed, available only in 2010) is similar for academic and vocational tertiary graduates. A cross-tabulation of overeducation and overskilling variables suggests that of the formally overeducated academic tertiary graduates, 40 per cent consider themselves as overskilled, while of the formally overeducated vocational tertiary graduates, only 1/6 adhere to this opinion. Nevertheless, the share of those who were both overeducated and overskilled in 2010 was 2.8 per cent of academic tertiary graduates and 6.5 per cent of vocational tertiary graduates.

The occupational structure of the sample is shown in Figure 4-1. ISCO major group 3 is the most popular among tertiary and vocational secondary graduates, while general secondary graduates prefer working in major groups 4 and 5 instead. There is substantial difference in the occupational structure of academic tertiary and vocational tertiary graduates. The choice of occupations by vocational tertiary graduates is quite close to that of vocational secondary graduates, although the share of respondents in the first three major groups increased at the expense of major groups 4 and 7. By contrast, nearly all academic tertiary graduates are employed in the first three major groups, and only a small minority is employed as clerks, service and shop sales workers and craft and related trades workers – major groups each employing between 10 and 20 per cent of vocational tertiary graduates.

Table 4-7 shows more detailed statistics on the occupations where overeducated tertiary graduates work. It is clearly seen from the list of occupations that these are unlikely to require tertiary qualifications. Even a police officer's job should not require vocational tertiary education, given the ILO's description of this occupation.

[26] TREE uses the 1988 version of ISCO.

Table 4-7 – Number of Overeducated Respondents in Each Occupation

Code	Occupation Title	Higher education type		
		Academic	Vocational	Total
4	**Clerks**	**16**	**12**	**28**
41	**Office clerks**	**11**	**12**	**23**
4110	Secretaries & keyboard-operating clerks	8	3	11
4120	Cashiers, tellers & related clerks		2	2
4190	Other office clerks	3	7	10
42	**Customer service clerks**	**5**		**5**
4222	Receptionists & information clerks	5		5
5	**Service workers & shop/market sales workers**	**9**	**16**	**25**
51	**Personal & protective services workers**	**7**	**12**	**19**
5122	Cooks		1	1
5123	Waiters, waitresses & bartenders	2		2
5130	Personal care & related workers	2		2
5149	Other personal services workers	1		1
5161	Fire-fighters	2		2
5162	Police officers		11	11
52	**Models, salespersons & demonstrators**	**2**	**4**	**6**
5220	Shop salespersons & demonstrators	2	4	6
6	**Skilled agricultural & fishery workers**	**1**	**1**	**2**
61	**Market-oriented skilled agricultural & fishery workers**	**1**	**1**	**2**
6129	Market-oriented animal producers & related workers	1		1
6141	Forestry workers & loggers		1	1
7	**Craft & related trades workers**	**6**	**11**	**17**
71	**Extraction & building trades workers**	**1**	**7**	**8**
7122	Bricklayers & stonemasons		1	1
7136	Plumbers & pipe fitters	1	2	3
7137	Building & related electricians		2	2
7141	Painters & related workers		2	2
72	**Metal, machinery & related trades workers**	**4**	**4**	**8**
7222	Tool-makers & related workers		1	1
7231	Motor vehicle mechanics & fitters		1	1
7241	Electrical mechanics & fitters	2	1	3
7242	Electronics fitters	1		1
7244	Telegraph & telephone installers & servicers	1		1
73	**Precision, handicraft, printing & related workers**	**1**		**1**
7311	Precision-instrument makers & repairers	1		1
8	**Plant & machine operators & assemblers**	**2**		**2**
82	**Machine operators & assemblers**	**1**		**1**
8278	Brewers, wine & other beverage machine operators	1		1
83	**Drivers & mobile-plant operators**	**1**		**1**
8322	Car, taxi & van drivers	1		1

4.2.2 Direct Effects on the Propensity to Quit

4.2.2.1 Methods

I use random-effects logit with random intercepts at the level of respondents for estimating the relationships between the propensity to quit (formally, high likeliness of not working with the current employer) in 6 months.

In general, I use six groups of explanatory variables:

- educational track (or type of education): vocational secondary, general secondary, academic tertiary after general secondary, academic tertiary after vocational secondary and vocational tertiary after vocational secondary
- variables defined before the choice of educational track: sex, tertiary education of father[27], non-native by birth (although some of them got Swiss nationality later), year of the wave as time fixed effects and monthly[28] unemployment rate in the greater region where the respondent's lower-secondary school was located as a contextual variable (Swiss Statistics data)
- personality variables (the subset of the personality variables in the data that was significant in the models used here): self-efficacy[29], intrinsic values at work[30], extrinsic values at work[31] (and its square), positive affectivity[32] and negative affectivity[33]
- inherent tendency to change jobs (an averaged summated scale of five questions[34])
- work-related variables: overeducation, working on evenings or weekends, industry, how well vocational training fits the current job and how well the respondent mastered this training (i.e., quality of training), which alternatives the respondent saw to the current job when starting to work there, and logarithm of net monthly personal income from all sources and its square
- job perceptions: frequent thoughts about quitting and job satisfaction

[27] I also tried the dummy for tertiary education of mother, but it was not significant.
[28] Corresponding to the month when the interview with the particular respondent took place.
[29] Average score on the extent of agreeing with the following statements: "I can always manage to solve difficult problems if I try hard enough," "I am confident that I could deal efficiently with unexpected events," "Thanks to my resourcefulness, I know how to handle unforeseen situations" and "I can usually handle whatever comes my way." Each question was measured on a 1–4 scale.
[30] Average score on the importance of the following: "To have a job where I can always learn something new," "To pursue an occupation in which I can fully deploy my competences," "To have a job where I am in touch with other people," "To pursue an occupation in which I can help other people" and "To have a job which gives me a feeling of doing something sensible." Each question was measured on a 1–4 scale.
[31] Average score on the importance of the following: "To earn a lot of money, a good wage," "To have a secure job position (security of employment)," "To have a position with good opportunities to get promoted" and "To have a job which is recognised and respected by others." Each question was measured on a 1–4 scale.
[32] Average score on five questions measuring the extent of feeling active, strong, enthusiastic, determined and interested, respectively, during the last month (before the interview). Each question was measured on a 1–5 scale.
[33] Average score on five questions measuring the extent of feeling irritated, preoccupied, angry, anxious and guilty, respectively, during the last month (before the interview). Each question was measured on a 1–5 scale.
[34] These questions measure to what extent respondents agree that "Occasional change of jobs broadens one's horizon," "Change of personnel can have positive effects on a company, as new people bring new ideas," "To stick to one job for too long leads to inflexibility," "It is unfair towards an employer to quit a job only a short time after having started it" (reversed), and "One should be happy to have a reasonably satisfying job and not change jobs all the time" (reversed).

There are potential limitations in estimating the simultaneous effects of all these variables in a single model, as there may be significant interrelationships between the regressors, biasing their reported effects on the propensity to quit. For instance, the job occupied by the respondent, and hence, job-related variables, actually depends on the educational track chosen by the respondent. In this section, I add blocks of regressors one by one and look on changes in the values of odds ratios, which gives five models:

- *BefEdCh*: includes the educational track and the variables defined before choosing it
- *Pers*: adds personality variables
- *ChngJbs*: adds the inherent tendency to change jobs
- *WrkRel*: adds work-related variables
- *Percpt*: adds job perceptions

4.2.2.2 Results

Table 4-8 shows the results.

Overall, it appears that there is no relationship between educational track and the propensity to quit. The only significant effect, from general secondary education, was observed in the first two models, but it completely lost significance, once the inherent tendency to change jobs – which has a very strong positive effect, as expected – was introduced.

Overeducation, controlling for evening or weekend jobs and other job-related variables, has a very strong positive effect on the propensity to quit: the overeducated are more than three times more likely to quit than the non-overeducated. After adding job perceptions, the size of the overeducation effect decreases only slightly – to 2.7.

One would expect the effect from overeducation to vary by educational track, so I rerun models *WrkRel* and *Percpt*, adding the respective interactions (see Table 4-9). In both models, the strongest link from overeducation to the propensity to quit is for respondents who moved to academic tertiary education after vocational secondary education. These respondents are 4.2 times as likely to quit as non-overeducated respondents are. Then there are differences in the behaviour of fully academic vs. fully vocational graduates. On the one hand, overeducated fully vocational graduates are 3.9 times more likely to quit than the non-overeducated, while the respective odds ratio for overeducated fully academic graduates is only 2.9. On the other hand, after adding perception controls, the odds ratio for fully academic graduates increased to 3.1, while that of fully vocational graduates lost statistical significance.

Table 4-8 – Odds Ratios after Random Effects Logit Regressions of a High Likeliness of Not Working with the Current Employer in 6 Months

	BefEdCh	Pers	ChngJbs	WrkRel	Percpt
Type of education (rel. to General secondary & Academic tertiary)					
Vocational secondary	0.856	0.893	0.941	0.748	0.667
General secondary	2.057*	1.919†	1.782	0.951	1.167
Vocational secondary & Academic tertiary	1.129	1.104	1.036	1.212	1.063
Vocational secondary & Vocational tertiary	1.153	1.255	1.191	0.748	1.239
Female	1.089	0.961	0.951	1.053	0.935
Tertiary-educated father	1.747***	1.676**	1.597**	1.693**	1.651**
Non-native (by birth)	0.642†	0.509**	0.670	0.623†	0.594
Year (rel. to 2010)					
2006	1.260	1.258	1.329	1.115	1.226
2007	2.078***	2.002***	2.050***	1.791***	1.715**
Regional unemployment, monthly	0.891*	0.852**	0.845**	0.789***	0.795***
Personality variables					
Auto-efficacy		1.620***	1.461**	1.558**	1.334
Intrinsic values at work		1.745***	1.455**	1.490**	1.323
Extrinsic values at work		0.023***	0.034***	0.046***	0.035**
Extrinsic values, squared		1.868***	1.773***	1.695***	1.791***
Positive affectivity		0.671***	0.672***	0.651***	1.052
Negative affectivity		1.439***	1.377***	1.287**	0.856
Tendency to change jobs			2.456***	2.583***	1.437**
Overeducated				3.053***	2.720**
Evening/weekend job				1.416**	1.644**
Industry (rel. to Trade, transport & accommodation)					
Agriculture				0.258†	0.343
Manufacturing				0.832	0.684
Construction				1.979**	2.095**
ICT				0.632	0.457†
Finance & insurance				0.860	0.974
Real estate				0.634	0.383
Professional, scientific, technical & admin.				1.061	1.197
Public admin., education & health				0.513***	0.849
Other				0.361**	0.386*
Employed in public sector				0.610*	0.486**
Vocational training: fit to job and quality (rel. to Good fit and quality)					
Poor fit and/or quality				0.738	0.661*
Excellent fit and quality				0.664**	0.824
No vocational training				1.049	1.026
Alternatives to current job (rel. to No alternatives)					
Another job				0.854	0.828
Non-job activity				1.848**	1.891*
Log personal income				9.093**	5.734*
Log personal income, squared				0.650***	0.716*
Frequently thinks about quitting					8.340***
Job satisfaction					0.575***
Constant	0.153***	3.886	0.260	0.027†	2.109
Regression diagnostics					
N			2336		
AIC	1817	1765	1705	1663	1301
BIC	1886	1869	1814	1882	1531
Panel-level variance proportion	0.143**	0.136*	0.104†	0.090	0.092

*** p < 0.01; ** p < 0.05; * p < 0.10; † p < 0.15 Random intercepts added at the level of respondents.

A common drawback of the ISCO-based overeducation measure used in this study is that it measures normative mismatch and ignores the unobserved characteristics of the respondent. As

Table 4-9 – Comparison of Overeducation Effects: Average Effect vs. Effect by Type of Education

	WrkRel	Percpt
Average	3.053***	2.720**
Interacted with Type of education		
General secondary & Academic tertiary	2.859†	3.092†
Vocational secondary & Academic tertiary	4.200**	4.316*
Vocational secondary & Vocational tertiary	3.920*	2.676

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$; † $p < 0.15$ The average effect is the effect from overeducation in *WrkRel* and *Percpt* models as shown in Table 4-8 (reported here again for convenience). Interaction effects estimated by the same models; they were included in those models instead of the single overeducation variable but in addition to the effects from educational track. The reference category of the interactions, as for the average effect, is "not overeducated." The General secondary category does not appear in interactions, because it does not contain any cases where respondents were overeducated (recall from Table 4-6). The Vocational secondary category was removed, because it predicts failure perfectly.

Table 4-10 – Comparison of Type of Mismatch Effects: Overeducation vs. Overskilling in 2010

	WrkRel			Percpt		
	(1)	(2)	(3)	(1)	(2)	(3)
Overeducated	3.582***			2.556**		
Overskilled		1.981***			0.700	
Type of Mismatch (rel. to Not overeducated & Not overskilled)						
Not overeducated & Overskilled			1.888**			0.645
Overeducated & Not overskilled			3.523***			2.429†
Overeducated & Overskilled			5.242**			2.058

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$; † $p < 0.15$ All models here are simple logit models run on the data gathered in 2010, because the overskilling variable is defined only for that year. In all models, $N = 1022$. Except for changing the mismatch variables and the absence of random effects, the models are identical to *WrkRel* and *Percpt* models used in this section.

already noted, the sample contains respondents who are formally overeducated but do not consider themselves overskilled (as well as vice versa). To compare the effects of overeducation and overskilling, and possibly, their interactions, I rerun models *WrkRel* and *Percpt* on the 2010 data[35] (see Table 4-10).

The odds ratio from overeducation is 1.8 times larger than that from overskilling: the overeducated are 3.6 times more likely to quit, while the overskilled are only around twice more likely to quit in the *WrkRel* model. In *Percpt* models, no overskilling effects are significant, as might be expected, given that overskilling is also a job perception and, thus, would be incorporated in job satisfaction and frequently thinking about quitting. The *WrkRel* model also shows that the propensity to quit rises with the severity of mismatch. The odds ratio is around 1.9 for overskilled but not overeducated, 3.5 for overeducated but not overskilled and 5.2 for overeducated and overskilled. Overall, these results suggest that normative mismatch is a stronger predictor of the propensity to quit than subjective mismatch.

[35] Recall that the overskilling variable is available only in 2010.

Other variables in Table 4-8 mostly have expected effects. There is no effect from sex. The propensity to quit is around 60–70 per cent higher for people from higher social class (with tertiary-educated fathers). At the same time, the non-natives are more likely to stay in their current jobs, although this result is near the border of significance.

Individual's personality plays an important role in assessing the likeliness of quitting the job. Auto-efficacy, the importance of intrinsic values and negative affectivity increase the probability to quit, while positive affectivity has the opposite effect.

Higher self-efficacy, most likely, is related to the desire for more challenging tasks that, the individual believes, (s)he would be able to handle, but which the current job cannot provide. This is also related to the effects from respondent's orientation to the extrinsic characteristics of the job, such as salary or career. Individuals with very high expectations related to extrinsic values are more inclined to quitting. However, the effect from this variable is U-shaped, so the same is also true for individuals with very low expectations. The importance of intrinsic values is also positively and linearly related to the probability to quit, so the decision to quit is related not to whether the person has stronger extrinsic values or stronger intrinsic values, but rather to how well the current job fits these values. Of course, the stronger the values are the more difficult it is for the job to fit them, hence, the higher tendency to quit. The effects from positive and negative affectivity also have expected directions. Respondents high on negative affectivity are more likely to see the negative aspects of their jobs and, hence, more likely to quit. Those high on positive affectivity are, on the contrary, less likely to quit due to greater openness to experiencing positive emotions.

Most personality variables, however, lose statistical significance as soon as perception variables enter the model. The only personality variable that remains significant is the importance of extrinsic values.

The tendency to change jobs, as expected, has a strong positive effect. A 25-percentage-point higher value[36] of this variable increases the probability to quit around 2.5 times. After adding perception variables, the size of the effect drops to 1.4, but remains significant.

There are several industries with significant effects. Compared to trade, transport and accommodation, the probability to quit in construction industry is twice higher, but it does not seem to be related to the financial crisis of 2008–09. According to Swiss Statistics (seasonally adjusted data), the construction industry was indeed one of the heaviest hit by that crisis:

[36] The values of this variable are in the range $[1,5]$. Hence, a one-point increase is equivalent to a 25-percentage-point increase.

compared to the last quarter of 2007, the industry lost more than one per cent of its workers and did not recover until the last quarter of 2009, when it got back to the pre-crisis level. Because years 2008 and 2009 were skipped in TREE data, the difficult times for the construction business are left out of scope of the model; hence, it should not bias its results. The probability is lower in agriculture; information and communication technology (ICT); public administration, education and health; and arts, other services, activities of households and extraterritorial organisations and bodies (appearing as "Other" industries in Table 4-8). Being employed in public sector also results in smaller chances to quit; and note that after adding perception variables, the effect from the public administration, education and health industry loses statistical significance, while that of public sector becomes more pronounced.

Besides overeducation, there are other indicators of the negative prospects the respondent faces in the labour market, some of which might have been chosen by the respondents themselves (the self-selection issue). For instance, working in an evening or weekend job might be the choice of the respondent to get a temporary[37] non-full-time job. Hence, the higher probability to quit associated with this variable might reflect the *ex-ante* decision of individuals more than the *ex-post* evaluation of the actual conditions of the job.

Respondents who received vocational training behave differently depending on how well they mastered this training and how well it fits the current job. In case of poor fit and/or quality of training, as well as in case of excellent fit and quality, respondents are more willing to stay in current job, compared to those whose fit and quality are at least good but not both of them are excellent. Respondents with poor fit or quality of training are willing to continue working because it might be difficult for them to find another job (they might simply like the current job, of course). On the contrary, respondents with excellent fit and mastering the skills received during vocational training on the top level are more likely to stay because they tend to be fully satisfied with their job – and this is reflected by the loss of statistical significance of this category once job satisfaction is included in the model.

Surprisingly, the probability to quit of the respondents who saw no other alternative to the current job when choosing to work there is statistically equal to that of the respondents who saw other employment opportunities. It is unrelated to job satisfaction, because the respective odds ratio has nearly not changed in the last model. On the contrary, respondents who considered another non-job activity, such as further education and/or training, are nearly 90 per cent more

[37] Temporary as regarded by the respondent, but not as stated in the contract, as fixed-term jobs were filtered out from the sample.

likely to quit in six months. This again points at respondents considering the current job as a temporary source of income or experience, but not as a permanent job.

Regional unemployment rate is a strong contextual factor. The more difficult it is to find a job in the greater region, the more the person is inclined to keep on working in the current job. Every percentage point increase of unemployment at the greater region level makes the person 10 per cent to 20 per cent (depending on the model) more likely to stay in the job. Note that the effect of general unemployment aggravates with adding work-related variables.

Personal income, which includes income from all sources, not only salary, has a quadratic relationship with the probability to quit. The higher the total income, the increasingly less likely the person is to quit. Note that while respondents for whom extrinsic rewards are very important have a higher probability to quit, those with substantial personal income are less likely to quit than those in the middle of the distribution.

Frequent thinking about quitting and job satisfaction have expected directions: the former increases the probability to quit and the latter decreases it.

4.2.3 Effects on Overeducation, Overskilling and Working in an Evening or Weekend Job

The absence of a direct effect from educational track on the propensity to quit does not mean that educational track is completely irrelevant. Indeed, recall from Table 4-9 that the effect from overeducation on that variable differs depending on the type of education of the respondent. In this section, I specifically study how it affects some of the most important variables in the model of the previous section: overeducation, overskilling and working in an evening or weekend job.

4.2.3.1 Methods

I use a seemingly unrelated regression approach, because it shows that the errors of the individual equations are highly correlated. There will be two models. In the first model, a seemingly unrelated bivariate probit is run on the equation of overeducation (*Overed*) and the equation of working in an evening or weekend job (*EWJob*); individual-level random intercept is added to the latter equation (it was not significant in the overeducation equation). In the second model, a third equation is added – that of overskilling (*Oversk*) – as previously, this model is estimated only based on the year 2010 data (hence, the specification does not include random effects). Both models were fit by seemingly unrelated multivariate probit using the conditional mixed process estimator (Roodman, 2011). Table 4-11 shows the results.

Table 4-11 – Marginal Effects after Multivariate Probit Models of Overeducation, Overskilling and Evening or Weekend Job

	Waves 6–8 (bivariate probit)		Wave 8 (trivariate probit)		
	Overed	**EWJob**	**Overed**	**Oversk**	**EWJob**
Type of education (rel. to General secondary & Academic tertiary)					
Vocational secondary	–0.134***	0.092**	–0.118***	–0.004	0.083**
General secondary		0.357***		0.119	0.375***
Vocational secondary & Academic tertiary	–0.011	0.018	–0.013	–0.041	0.018
Vocational secondary & Vocational tertiary	0.346***	0.128†	0.266***	–0.038	0.084†
Female		0.032***			
Male with child	0.045**		0.078**		
Non–native (by birth)	0.038***		0.077***		
Regional unemployment, monthly		–0.021†			
Personality variables					
Auto-efficacy				0.105***	
Intrinsic values at work		0.120***			0.079***
Positive affectivity				–0.044**	
Negative affectivity	0.011**		0.020**	0.051***	
Industry (rel. to Trade, transport & accommodation)					
Agriculture	0.023	0.264	0.211**	0.056	0.205
Manufacturing	–0.019	–0.499***	0.047*	0.102**	–0.257***
Construction	–0.016	–0.508***	0.081	0.013	–0.295***
ICT		–0.454***	0.095***	0.098***	0.091*
Finance & insurance	–0.052***	–0.509***		0.106*	–0.254***
Real estate	–0.006	–0.413***	–0.001	0.050	–0.321***
Professional, scientific, technical & admin.	–0.047***	–0.493***	0.116	0.010	
Public admin., education & health	–0.048***	–0.272***	0.014	0.035	–0.251***
Other	0.039	–0.135	0.119**	0.070	–0.035
School located in Zurich		–0.121**		0.086*	–0.112**
Random Effects (individual-level)					
Standard deviation		1.928			
(standard error)		(0.201)			
Correlations of errors					
Overed with EWJob	0.443**		0.463***		
Overed with Oversk			0.256**		
Oversk with EWJob				0.092	
N	2346		1047		

*** p < 0.01; ** p < 0.05; * p < 0.10; † p < 0.15 Univariate (unconditional on the success in other equation(s)) marginal effects reported. General secondary education and ICT industry constrained to zero in the overeducation equation of the bivariate probit model, as these variables predict failure perfectly.

4.2.3.2 Results

The most interesting effect in these models comes from fully vocational educational track: respondents from this track have a 25 to 35 per cent (depending on the model) higher probability to become overeducated. At the same time, graduates from academic tertiary programmes, whether with general or vocational secondary background, are not statistically different in any of the equations. Note, however, that educational track does not affect the self-perception of being overskilled for the current job.

Males with a child and non-natives also have a higher probability of becoming overeducated, pointing at certain inequalities in the Swiss labour market. In preliminary specifications of these models, however, there was no effect from sex or from being a female with a child.

Negative affectivity increases the chances of being overeducated or overskilled. In addition, positive affectivity decreases the chances of being overskilled, while auto-efficacy has a strong positive effect on overskilling. The effect from psychological variables is both more ubiquitous (more variables are significant) and stronger (the size of the effect is larger) in the case of the *self-reported* overskilling measure than on the *objective* overeducation measure. In other words, the self-perception of being overskilled depends largely on the psychological peculiarities of the respondent. Confront it with one of the results from the previous subsection that overskilling is a weaker predictor of the *likeliness* (assessed subjectively) of quitting the job in six month than overeducation.

Depending on the set of survey waves under consideration, different industries tend to have different effects on overeducation. Only manufacturing and ICT have the same signs in the equations of overeducation and overskilling; moreover, the sizes of the effects from ICT on both mismatch measures are nearly the same.

The equation explaining the choice of evening or weekend job is simpler. Respondents with secondary education, especially on the general track, are significantly more likely to work in these jobs, which might be explained by the need (or willingness) to earn income while studying. As one would expect, the probability of getting an evening or weekend job is negatively related to regional unemployment: with higher unemployment, these jobs appear to be more vulnerable than day jobs. Women are more likely to choose these jobs, as are those for whom intrinsic values at work are important. Most jobs of this type are found in trade, transport and accommodation industries (as nearly all other industries have negative and strongly significant effects). Finally, weekend or evening jobs are more prevalent in greater regions other than Zurich.

4.2.4 Discussion and Conclusions

It seems to be a typical stand that an increasing role of vocational education will help in keeping overeducation (and its negative effects) low. This study showed that this is not necessarily true. Compared to academic tertiary graduates (with or without vocational secondary background), vocational tertiary graduates are 25 to 40 per cent more open to normative overeducation risk. Educational track does not affect subjective overskilling, but this mismatch measure is nearly twice weaker than normative overeducation in predicting the probability to quit the current job in six months. The most likely to quit, as one would expect, are respondents who both are overeducated and feel overskilled. While vocational tertiary graduates are the most open to overeducation, they have only the second largest propensity to quit once they actually become overeducated after the graduates of the mixed track (academic tertiary after vocational secondary).

There are several important implications for the view of education and labour markets and for related policies. The main implication is that overeducation does increase the propensity to quit. This adds to the volume of the literature on non-monetary negative effects from overeducation, such as increased engagement in on-the-job search, lower job satisfaction and more symptoms of depression, as reviewed in Section 2.1.2.

Secondly, arguments favouring subjective mismatch measures over normative mismatch measures (because the latter may be ignoring unobserved heterogeneity, quickly become obsolete etc.) should be more questioned. In this paper, the subjective measure had smaller effect on the propensity to quit than the normative measure. The latter, thus, is more relevant to the individual's status in the labour market than the former. In particular, this is an additional justification of using normative mismatch measures in datasets where subjective measures are not available.

Thirdly, vocational education (secondary or tertiary) is not a panacea against overeducation. Had it been, (1) we would observe the lowest overeducation among European countries in Germany and Switzerland, (2) there would be a strong negative effect on the probability of overeducation from vocational education relative to academic education at tertiary level, or at least from the mixed track relative to fully academic track, or at least (3) there would be no difference in the effect from overeducation on the propensity to quit at different educational tracks. The contrary is actually observed. Not only fully vocational graduates are more likely to become overeducated than fully academic graduates are, but also they are much more likely to quit than fully academic graduates once they become overeducated.

4.3 Choices Related to Getting a Doctoral Degree in the Baltics

Doctoral education plays two roles in society. Firstly, it enhances the research potential of the country, which has a positive effect on its exporting possibilities and, in a more general sense, competitiveness. Many countries recognise it and provide support for doctoral students via grants and/or scholarships. That is also reflected in the Europe 2020 Strategy of the EU, where the goal of enhancing R&D is among the five main goals put forward. Besides, the importance of doctoral degree for individuals has been increasing. Some graduates are keen on pursuing further studies to help self-realisation, but doctoral degree also gives more space for manoeuvring in the labour market due to the abundance of bachelors and masters and the related credential inflation. Indeed, while previously, one would hardly find a doctoral degree holder outside universities, research institutions and hospitals, it is not uncommon nowadays to see them in all different kinds of positions, including those in management and politics. Put shortly, doctoral degree allows to positively distinguish its holder from other candidates for a good job.

In this study, I compare two decisions prospective doctoral students make across three Baltic countries: Estonia, Latvia and Lithuania. The first decision is whether to pursue the opportunity to get a doctoral degree. The discussion will centre on the reasons behind choosing to get it, rather than on the differences between those who chose to pursue it and those who chose to stay out. The second decision is about the university (or, put more broadly, the HEI) where to study. This might be related to the first decision, so that the characteristics of the HEI important for a particular person depend on why (s)he wants to get a doctoral degree, and this relationship is studied here.

Most of the literature on doctoral education has dealt with shortening time to completion and decreasing the number of drop-outs. The motivation to enrol has been out of scope of the mainstream, perhaps due to successfully attracting own master's students to advanced-level programmes and seeing no major reason to boost the number of doctoral students. The few studies that are available on the topic typically focus on the proportion of graduates influenced primarily by personal or intrinsic factors (such as the interest in the topic of studies or in the study process itself, personal development or contributing to society, science or government) and extrinsic factors, related mainly to improving the current or future position in the labour market (such as career or salary) but also keeping up with professional requirements or acquiring the skills necessary for the current or future job (e.g., at the university). European studies showed that around half of respondents are motivated intrinsically and the other half extrinsically (Leonard, Becker, & Coate, 2005; Del Carmen Calatrava Moreno & Kollanus, 2013). In the US, academic (PhD) students in education science are mainly motivated by extrinsic factors, while professional (EdD) students mentioned intrinsic factors most frequently (Biddle, 2013). In Australia, doctoral students in history were mostly motivated by intrinsic factors (Brailsford, 2010).

No comprehensive research on the motivation of enrolment at doctoral level has been done in the Baltic countries. Although the administrative staffs of some doctoral programmes at some HEIs do ask new doctoral students about the reasons of enrolment, these findings are specific to the concrete institution and the concrete field of study, which limits their generalizability. Moreover, the results of such studies are typically unpublished and kept for internal use. The only comprehensive study was done in Latvia by Tarvid (2014), and this study builds heavily on the methodology developed therein. In fact, this paper takes Latvian data from Tarvid (2014) and compares them with similar data from Estonia and Lithuania.

The structure of the rest of the section is as follows. The next subsection presents a general view of the current state and recent dynamics of doctoral education in the Baltic countries. Then Section 4.3.2 describes the dataset that will be analysed in Section 4.3.3. The last subsection discusses the findings and concludes.

4.3.1 Doctoral Education in the Baltic Countries at a Glance

Estonia, Latvia and Lithuania are small countries, compared to most EU Member States. Nevertheless, their landscapes of doctoral education have certain important specifics, which will be shortly discussed in this subsection. This background discussion will start with general demand for doctoral degrees and continue with the relative popularity of fields of study at doctoral level.

The overall demand for doctoral studies soared in 1998–2012 in all three countries, but the dynamics were very different (see Figure 4-2). The absolute number of doctoral students grew more than 3 times in Estonia, around 2.5 times in Latvia and twice in Lithuania. By 2012, albeit being the smallest country of the three, Estonia attained the largest number of doctoral students, exceeding the 3000 threshold and moving above Lithuania on this indicator. The difference is reflected in the trend of the share of doctoral students from the size of the population aged 25–49, which has a higher slope in Estonia.

The growth patterns of the number of doctoral students differ markedly. In Estonia, the growth was mainly linear, except for the stable plateau in 2001–2004 and a slowdown in 2009. The same dynamics can also be observed at bachelor and master's levels, as the line representing the share of doctoral students from all tertiary students (see Figure 4-2) also flattened in the mentioned years. Lithuania presents a contrasting example: the dynamics there follow a jump-and-plateau pattern. The number of doctoral students jumped in 1999 and again in 2004, while between the jumps there was either slow linear growth (1999–2003) or stability (2005–2012). Years 1999 through 2003 were also marked by a steady decline in the relative share of doctoral

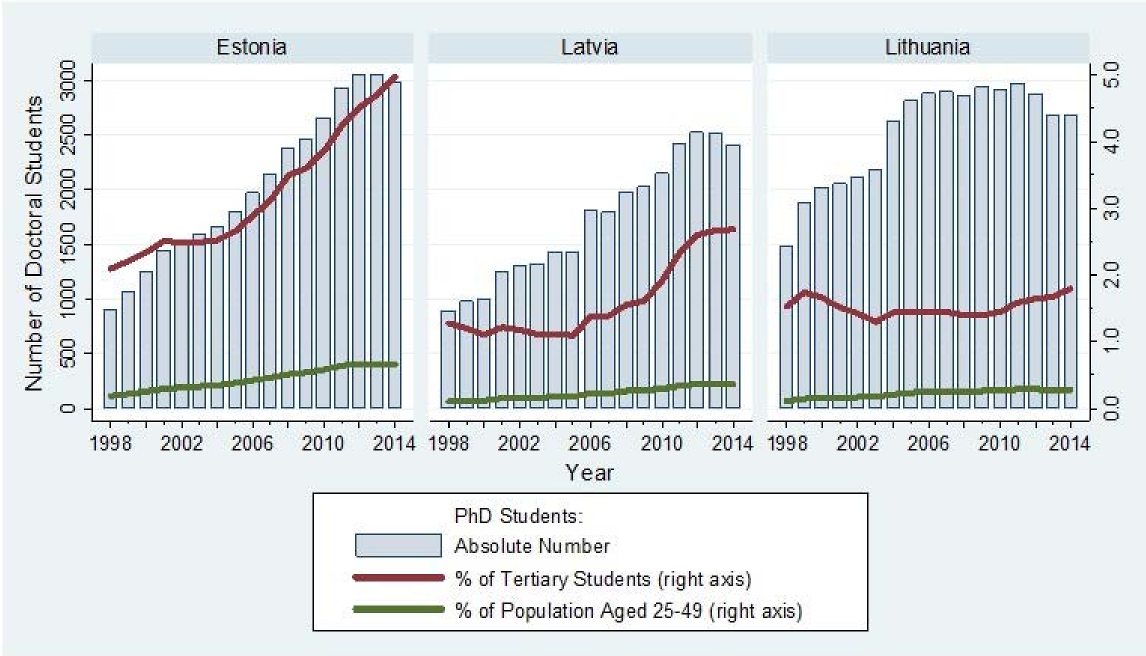

Figure 4-2 – Demand for Doctoral Studies in the Baltic Countries, 1998–2012

Source: Eurostat data

students in the higher education system, despite growth in absolute numbers; at the same time, while the absolute number of doctoral students was stable in 2009–2012, their share among all students increased, reflecting a drop in the demand for lower-level tertiary degrees. Latvia is somewhere between Estonia and Lithuania. There were relatively strong jumps in the number of doctoral students in 2001, 2005 and 2011, while in other years the growth was sluggish. The share of doctoral students from all students was declining up to 2005 (except for a small increase in 2001), but the decline was not as strong as in Lithuania. The indicator started a stable growth only in 2005, and its slope became steeper in 2009: in just three years (2009–2012), it soared from 1.6 per cent to 2.6 per cent. In this latter respect, Latvia is similar to Estonia: in 2009–2012, both increased the number of doctoral students by 1000 and their share in the higher education system by 1 percentage point; but Estonia started 2009 with around 2500 doctoral students and 3.6 per cent share, while in Latvia, these indicators were only at the levels of around 2000 and 1.6 per cent, respectively, so the relative change in Latvia is bigger.

In 2012–2014, however, the direction of the trend reversed in all three countries: the number of doctoral students started declining. At the same time, in Lithuania and, especially, in Estonia, the share of doctoral students from all tertiary students continued to rise, meaning that the drop in the number of bachelor and master’s students was stronger. In general, the future of the higher education market in the upcoming 15–20 years in the Baltic countries appears to be quite

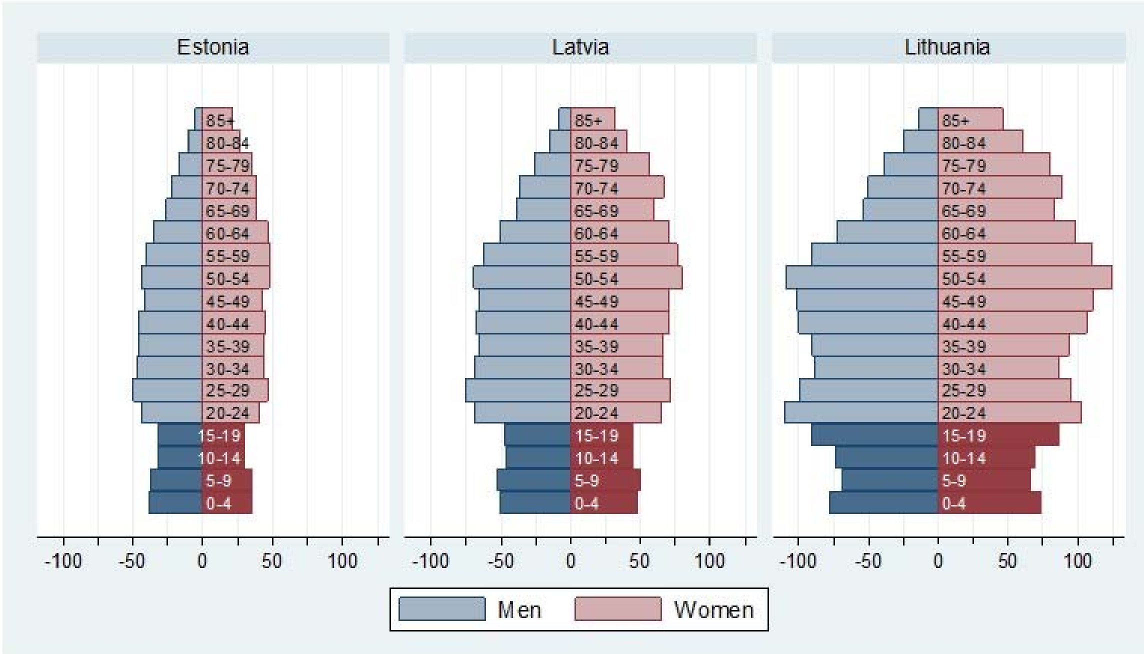

Figure 4-3 – Population Pyramid in the Baltic Countries, 1 Jan 2014

Source: Compiled from Statistics Estonia, Central Statistical Bureau of Latvia and Statistics Lithuania data. Population numbers shown in thousands.

bleak due to demographic problems (see Figure 4-3). In 2014, the number of inhabitants aged 15–19, who form the cohort of future bachelor students in the coming five years, was 30 per cent smaller than that of inhabitants aged 20–24, who form the majority of current bachelor students. In Lithuania, the drop is only 20 per cent, but compared to Estonia and Latvia, there is an additional contraction in the younger 5-year group: the size of the 10–14 age group is 30 per cent smaller than the size of the 20–24 age group, so that Lithuanian higher education system will run into the same problems, but later.

This demographic structure affected the HEIs providing doctoral degree programmes already in 2013–2014, as was shown in Figure 4-2, and the demand for doctoral studies will continue dropping. The competition among HEIs will increase, and HEIs understanding their market better will be in a more favourable position.

As fields of study will play an important role in the following discussion, it is necessary to understand how different the three countries are with respect to relative demand for fields at doctoral level (see Figure 4-4). In 1998, the largest number of doctoral students chose life sciences (52%) in Estonia, engineering (38%) in Latvia and engineering or social sciences (both 24%) in Lithuania. In 2012, the most important field in Estonia still was life sciences (23%), in Latvia still engineering (19%), but in Lithuania life sciences moved to the first place (23%, as in

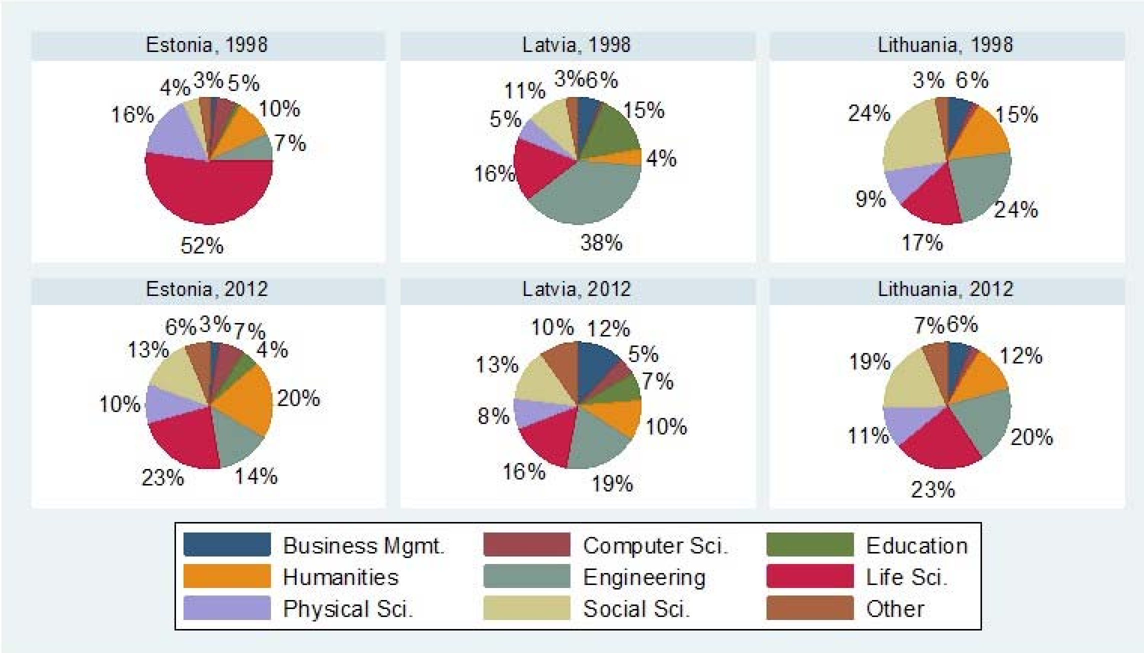

Figure 4-4 – Popularity of Fields of Study at Doctoral Level in the Baltic Countries, 1998 vs. 2012

Source: Eurostat data

The standard fields of study used by Eurostat were recombined to the extent possible to match the field groups used later in the study. Business mgmt. maps to Eurostat's Business and administration; Computer sci. to Computing; Education to Teacher training and education; Engineering to Engineering, manufacturing and construction; Life sci. to the sum of Life science, Agriculture and veterinary to Health and welfare; Physical sci. to the sum of Physical science and Mathematics and statistics; Social sci. to the sum of Social and behavioural science and Journalism and information; and Other to the sum of Law and Services. Eurostat does not have data for the number of doctoral students (nor graduates) in teacher training and education in Lithuania (Lithuanian statistical office also does not have this field of study in its data). Only sectors with at least 3% market share were labelled.

Estonia), although engineering and social sciences remained at the second place (each with roughly 20%). Thus, with some exception for Lithuania, the leading fields of 1998 were able to keep their positions, though not relative popularity, in 2012.

Compared to the list of most popular fields, the relative popularities of fields changed more markedly. In Estonia, the share of social sciences more than tripled and the shares of humanities, engineering and other fields (mainly, law) doubled, all principally at the expense of life sciences (whose share dropped 2.25 times) and, to a smaller extent, physical sciences. In Lithuania, changes have been much less radical – the only big jump happened in other fields, and the popularity of humanities, life sciences and physical sciences increased only by a few percentage points. All these changes came at the expense of engineering and social sciences. In Latvia, other fields had a stunning increase of more than 3 times, humanities 2.5 times and business management and education both doubled their relative popularity, while social and physical sciences had more modest changes. Latvia is the only Baltic country where the relative weight of business management in the palette of fields of study changed. At the same time, engineering and education each lost half of their respective shares by 2012.

4.3.2 Data

Two special-purpose online surveys were done to gather the necessary data. The first survey targeted doctoral students in Latvian HEIs; these are the same data as used by Tarvid (2014), although some modification and recoding was applied, as described below. It was later generalised and sent to doctoral students of Estonian and Lithuanian HEIs. Survey links were sent to all current doctoral students of participating HEIs through HEI administration (sending links directly would not be possible for privacy considerations), and HEI management support was reserved to enhance the response rate of doctoral students. Overall, seven HEIs participated from Estonia, 15 from Latvia and 14 from Lithuania.[38]

Both questionnaires had the same structure with four groups of questions. The first group collected contextual information about the doctoral studies of respondents. The second group then probed into the decision of enrolment into the current doctoral programme. The third group asked respondents about choosing the HEI for studies at doctoral level. The questionnaires ended with questions about the background information on respondents.

The most significant change, as compared to the questionnaire distributed to Latvian HEIs, was applied to the question defining the field of study of the respondent. In the first version, respondents were choosing one field from a list of fields constructed to match the broad titles of the doctoral programmes actually provided by the participating Latvian HEIs. That was superseded in the second version by allowing to select several fields from the list of fields adapted from the Survey of Earned Doctorates used in the United States. In particular, it resulted in a certain change of grouping of narrower fields, which was subsequently applied to Latvian data to make the three countries comparable along this dimension. There are, hence, nine field of study groups: business management, computer science, education, humanities, engineering,

[38] The participating HEIs in Estonia were: Estonian Academy of Arts, Estonian Academy of Music and Theatre, Estonian Business School, Estonian University of Life Sciences, Tallinn University, Tallinn University of Technology and University of Tartu. In Latvia: Art Academy of Latvia, BA School of Business and Finance, Baltic International Academy, Daugavpils University, Latvia University of Agriculture, Latvian Academy of Sport Education, Liepaja University, Riga International School of Economics and Business Administration, Rezeknes Augstskola, Riga Stradins University, Riga Technical University, Transport and Telecommunication Institute, Turiba University, University of Latvia and Ventspils University. In Lithuania: Aleksandras Stulginskis University, ISM University of Management and Economics, Kaunas University of Technology, Klaipeda University, Lithuanian Academy of Music and Theatre, Lithuanian Sports University, Lithuanian University of Educational Sciences, Lithuanian University of Health Sciences, Mykolas Romeris University, Siauliai University, The General Jonas Zemaitis Military Academy of Lithuania, Vilnius Gediminas Technical University, Vilnius University and Vytautas Magnus University.

life sciences, physical sciences (incl. mathematics), social sciences and other fields (architecture, law, library and information science, psychology, public administration, social work and sports).[39]

Another change was the substitution of the questions on the concrete field(s) of study at bachelor and master's levels by the questions on the respondent's perception of the extent to which the field of study was changed between the consecutive levels of education (in Tarvid (2014), the fact of change was determined by the author based on a comparison of fields at different levels). It is possible to change the field of study even while staying in the same broad field, so asking for a subjective opinion should have provided data that are more accurate. Besides, this change helped to keep the fatigue level of respondents moderate; it would certainly have increased, had they browsed two more times through the long list of fields.

The core questions of both surveys are about the main and second main goals pursued by the respondent when deciding to enrol in the current doctoral programme. Besides two respective multiple-choice questions, the respondents were asked to provide a short description of an occurrence or observation that made them believe that having a doctoral degree would help them in achieving each of these two goals. Answers to these questions will be analysed in the following subsection.

The total sample size is 971 respondents, all of which answered the core questions on motivation. This includes 287 respondents from Estonia, 330 from Latvia and 354 from Lithuania, which is 9.6%, 13.7% and 12.3% from the total population of these countries' doctoral students in 2014 (data from Figure 4-2), respectively. Of all respondents, 726 answered all questions.

4.3.3 Results

4.3.3.1 Goals Pursued when Enrolling in Doctoral Studies

Respondents were asked to choose a main and a second main goal they pursued when they chose to get a doctoral degree from nine pre-specified goals. Their descriptions in the "Other" answer option were recoded into one of these goals.

Figure 4-5 shows the details by country. Combining the main and second main goals, these goals may be divided into three groups by their country-specific popularity. The three countries are similar in the way respondents mentioned the top four goals, the middle three goals and the

---

[39] As compared to the grouping used by Tarvid (2014), psychology was moved from education to "other fields," law was moved to "other fields," economics was moved to social sciences, computer science and engineering were separated and a part of respondents originally put in the physical sciences group were reclassified as engineering students.

two least important goals. Indeed, in all these countries, respondents referred most frequently to (1) continuing learning and/or research experience, (2) contributing to the development of science or society or world in general, (3) new achievements or (4) better career prospects.

Respondents mostly found better competitive position in the labour market, demand for a doctoral degree by employers or a long desire to get a doctoral degree less important than the top goals, but these were still mentioned quite frequently. In Estonia and Latvia, each of these was marked by around 20 per cent of respondents, but only the "always wanted" option reached that level of popularity in Lithuania, where the other two goals had only around 10%–15% active seekers. Finally, social status and better salary are each mentioned by less than 10 per cent of respondents, and in most cases, only as a second main goal.

Following Tarvid (2014), I combine the nine goals in two groups: personal goals and labour-market goals. Labour-market goals are strategic goals related to the future success of the respondent in the labour market. These include enhancing career prospects, strengthening competitive position, moving to a higher salary range and reacting to employer's demand for employees with doctoral degrees. Personal goals include both long-term, strategic goals related to self-development (new achievement, realisation of a long desire and lifting social status), as well as searching for possibilities to improve the current state of affairs in science, society or

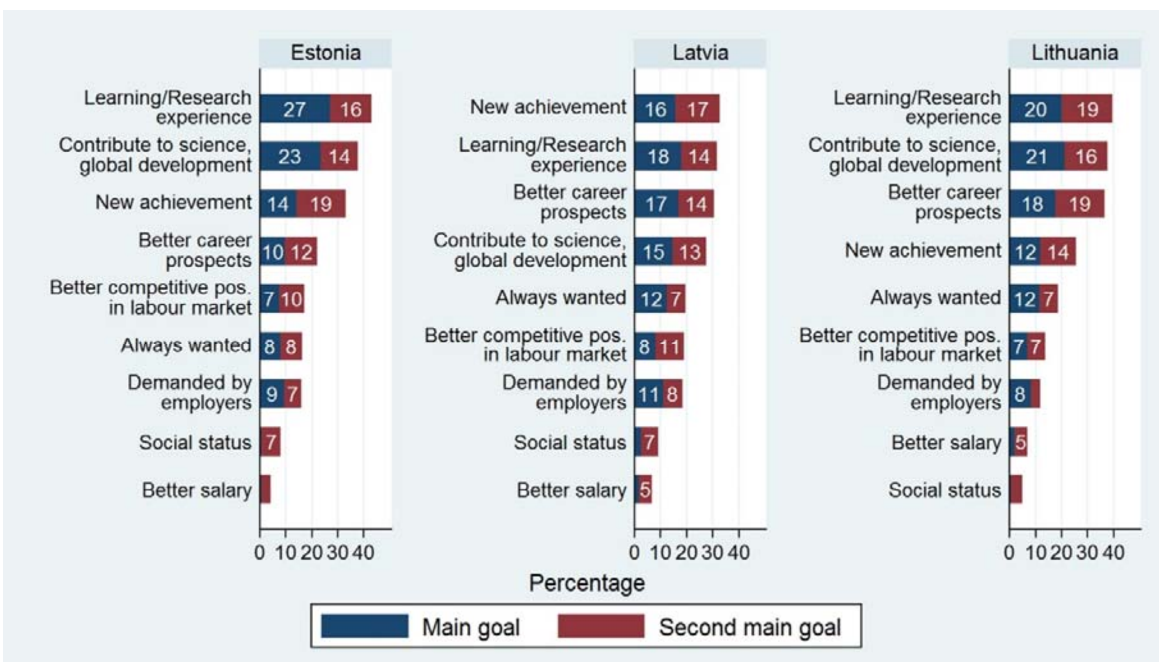

Figure 4-5 – Goals Pursued when Enrolling in Doctoral Studies, by Country

Only bars of length at least 5% labelled.

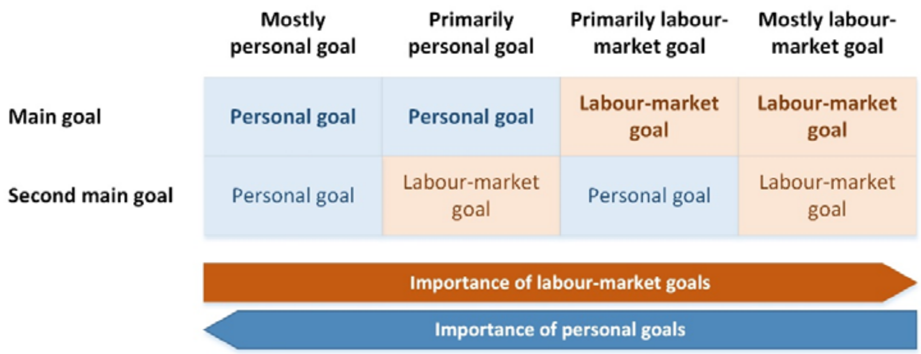

Figure 4-6 – Definition of Goal Types

global development[40] and simply continuing their previous learning or research experience (sometimes, because there is not much else to do). The respondents can now be divided into four types based on their main and second main goals and, hence, sorted on the prevalence and importance of personal or labour-market goals, as shown in Figure 4-6.

Most respondents who pursued labour-market goals commented that they were either already employed by a higher education or research institution or planning to work there. These students knew that a doctoral degree was a prerequisite for career advancement and higher salary in the field of higher education and/or research. For some Estonian lecturers, however, it was a question of *keeping* their job in the first place, as their university had informed them that all lecturers would have to have a PhD degree.

Several respondents who were working outside academia (typically, in engineering or industrial R&D) observed that in their field, people with master’s degree could only be involved in sales or do very basic experiments, so they saw a PhD degree as a prerequisite for a more interesting and demanding job. MD degree holders mentioned that they would have access to better positions if they had a PhD. Some Lithuanians stated that the degree was a substitute for examinations to higher-level legal positions such as judge or prosecutor, which was a major motivating factor for them. Those with substantial labour market experience viewed doctoral studies as an opportunity to deepen knowledge, widen view and learn theoretical frameworks to enhance their leadership abilities or the quality of their services. Learning practical skills – such as independent planning and analysis, decision-making, ability to concentrate and conducting experiments – is another benefit from getting the degree, which, as respondents assume, will enhance their competitive position in the labour market. Finally, several respondents were tired

[40] One could argue that this reason should be put in a separate category, as it is not as respondent-centred as all other mentioned goals. While there might be certain benefits in following this logic, it would complicate the analysis of the first- and second-order goals, as having three groups of goals at each level gives nine groups instead of four, as in the setup used in this study.

of their current job due to poor appreciation from others, low salary, not using the knowledge gained in higher education or simply being too monotonous as compared to doctoral studies, which “are always challenging, dynamic, [making] your brain work all the time,” so they saw doctoral studies as a pathway to better jobs.

Respondents working both in and outside of academia (primarily, studying in Latvia or Lithuania) quite frequently mentioned that a doctoral degree was necessary to lead projects and participate in grant competitions, both of which, in turn, were connected to higher salary.

At the same time, some respondents (mostly, Estonian students) expressed scepticism regarding the value of a doctoral degree outside academia: “it is less valued and [understood], and may even damage one’s career prospects.” A Lithuanian student noted that the payoff from doctoral studies might actually be negative: while some are studying for the degree, others “gain practical experience and become more competitive in the labour market.”

Respondents frequently mentioned credential inflation, as they observed that the number of bachelor and master’s graduates was increasing consistently, so they saw a doctoral degree (and the knowledge they gained and the skills they developed while getting it) as a possibility to differentiate themselves.

One of the alternative mechanisms of reaching the decision to get an advanced degree is through influence from others. Typically, other PhD students, colleagues with a doctoral degree or the supervisor persuaded the respondent that it was beneficial to get a doctoral degree. Alternatively, respondents observed that people with doctoral degrees had better careers and salary, worked in the best jobs or in the companies admired by the respondent and had better job security during crises.

Moving to personal goals, several respondents mentioned a family factor – such as a tradition of doctoral education in their family or, on the contrary, that none of their relatives had a PhD degree and, thus, getting one would be a great achievement for the respondent. Continuing comments on the degree as an achievement, respondents typically referred to it as the highest education level – the “gold medal” of education – so simply getting it was already an achievement for them. In addition, learning something new in the context of self-development, like frustration tolerance, was mentioned – but respondents stressed that it was the *process* of getting the degree, rather than the degree itself, that provided them with this possibility. The degree was also viewed as a signal of higher status (“entrance ticket to the elite”), and many respondents simply wanted to have “PhD” written near their name because they believed it to be, or had experienced that it was, respected by others.

Another popular category of personal comments was about continuing the research that respondents did at master's level – they typically viewed their research as unfinished and wanted to work more on it, and doctoral studies provided excellent opportunities to do it. Doctoral studies also typically motivate students to travel extensively – and indeed, travelling to participate in conferences (to meet with other researchers or potential investors) or as an exchange student was mentioned frequently as a complement to the possibility of collaborating with researchers working abroad. Nevertheless, some comments showed that the scholarships were inadequately small and neither allowed for extensive travelling nor even covered living costs, forcing students to work in parallel. In 2009–2013, a few scholarships were available through European Social Fund, and these were substantial, compared to travelling costs and average salary, but when this period ended, all the old problems with inadequate government financing reappeared.

The largest category of personal comments can be named "had nothing else to do." Some respondents confessed that they were "lazy" or "not ready" to participate in the labour market, saw "no [other] attractive options" or had "good financial position" and, apparently, preferred studies to work. Others viewed doctoral studies as an opportunity to do something while on a child leave (sometimes, because they were unable to find a job while their kid was small) – or as a hobby to escape from boring work and family life.

Other comments were related to staying up-to-date with the latest research in the field, the time efficiency of doing research simultaneously with getting a degree, promoting a new field or method (such as space technology, body and movement psychology, new methods in economics and linguistics) and adding weight to own opinion to be considered an expert and "be heard" by other experts.

Figure 4-7 shows the distribution of the four goal types (defined in Figure 4-6) by field of study in the three Baltic countries. Overall, it appears that students of Latvian HEIs are more focused on the labour market than students in Estonia or Lithuania. Indeed, more than 15% of students in all fields except for social sciences and humanities pursued mostly labour-market goals, and even in the latter two fields, the share of such students was around 10%. The fields where labour-market goals are most popular in Latvia are physical sciences (33% had mostly labour-market goals and 55% had a labour-market goal as the main goal) and education (28% had mostly labour-market goals), although life sciences should also be mentioned here, as it has 54% of students who mentioned a labour-market goal as the main goal. This should be compared to the maximum

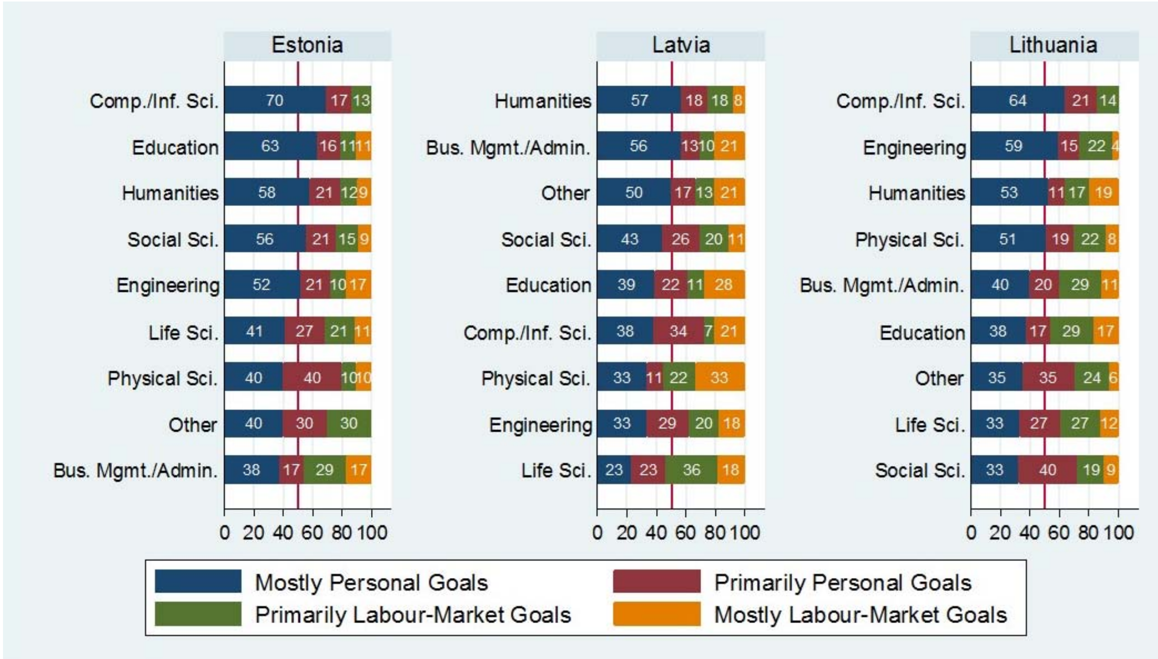

Figure 4-7 – Goals Pursued when Enrolling in Doctoral Studies, by Country and Field of Study

of 17% choosing mostly labour-market goals in Estonia (management and engineering) and 19% in Lithuania (humanities).

On the other side of the spectrum are fields whose students have a personal goal as the main goal or even as both the main and the second main goals. Here, the fields also do not overlap much across countries. The three fields with the largest share of students pursuing mostly personal goals are computer science (70%), education (63%) and humanities (58%) in Estonia, humanities (57%), management (56%) and other fields (50%) in Latvia and computer science (64%), engineering (59%) and humanities (53%) in Lithuania. Only humanities appear in the top three fields on this indicator in all three countries, and computer science holds the first place both in Estonia and in Lithuania, but is only in the middle of all fields on this indicator (with 38%) in Latvia. Looking on the share of students having a personal goal as the main goal (with any goal as the second main goal), the top three fields are computer science (87%), physical sciences (80%), education and humanities (both 79%) in Estonia, humanities (75%), computer science (72%), management and social sciences (both 69%) in Latvia and computer science (85%), engineering (74%) and social sciences (73%) in Lithuania. Only computer science was mentioned in all three countries, but humanities and social sciences were mentioned in two of the three. Regarding computer science, note again that Latvian students, even when having a personal goal as the main goal, tend to still have a labour-market goal as the second main goal much more frequently than those in Estonia or Lithuania (34% versus 17% and 21%, respectively). Notable is also a strong personal focus of physical sciences students in Estonia and Lithuania, which should be contrasted with a similarly strong labour-market focus of students of the same field in Latvia.

4.3.3.2 Reasons of Pursuing Particular Goals: Econometric Analysis

I now look in more detail on whether there exists some regularity between the characteristics of respondents, their chosen field of study at doctoral level and the motivation they had when enrolling into their study programme. I will use the four types of goals introduced in the previous subsection and cut the sample in two directions. Firstly, I will look on the differences between countries. The sample size does not allow using all four goal types in the dependent variable in each country. Thus, I run the standard logistic model with the dependent variable showing whether the main goal is a labour-market goal. Table 4-12 shows the results for these models. Secondly, I look on the differences between men and women. The sample size allows using the four-category dependent variable defining the type of goal in these models. Recall from Figure 4-6 that the four types may be arranged in a logical order. This advocates the use of generalised ordered logit (see Section C.5), where the proportional odds assumption of standard ordered logit is relaxed for some variables. Tables 4-13 and 4-14 show the results of sex-specific models.

Field of study, as expected, has one of the strongest relationships with the motivation of the prospective doctoral student. Business management and life sciences students are typically more labour-market oriented than humanities students. Both fields have significant effects in the whole sample, in Latvia and for females, and the life sciences field is also significant in Estonia. While not significant at country-level, fields combined in the "other" group[41] attract more labour-market oriented students than humanities in the overall and female-specific models. For males, motivation does not differ across the mentioned fields, but there are effects from engineering (near the border of significance) and social sciences, both again pointing at higher importance of labour-market goals than in humanities. In Lithuania, there are no significant effects from fields, but the more personally-oriented effect from computer science is near the border of significance.

Motivation also depends on how the field of study at doctoral level compares with the previous educational background, although it is generally visible only in Latvia and in the model for females. In both cases, respondents who changed field of study both at master's and at doctoral levels are significantly more personally oriented than those who stayed in the same field, and in the model for Latvia, a similar effect is from changing the field only at master's level.

[41] Recall that this group of fields contains architecture, law, library and information science, psychology, public administration, social work and sports.

Table 4-12 – Odds Ratios after Country-Specific Logistic Regressions of Having Labour-Market Primary Goals

	Estonia	Latvia	Lithuania
Field of doctoral studies (rel. to Humanities)			
Business Management	2.206	3.914**	1.019
Computer Science	0.648	1.696	0.201†
Education	0.662	2.005	1.103
Engineering	1.154	1.873	0.532
Life Sciences	2.219*	6.865***	1.138
Physical Sciences	0.762	2.612	0.794
Social Sciences	1.032	1.813	0.855
Other	1.588	1.303	0.613
Change of field of study (rel. to Never changed)			
Changed at master's level	1.286	0.265*	0.876
Changed at doctoral level	0.786	0.738	2.081†
Changed at both master's and doctoral levels	0.397	0.346*	1.817
Started doctoral studies at age > 30 years	0.897	4.314***	2.420**
Single female	1.131	1.677	0.484*
Decided that would get doctoral degree (rel. to After getting master's degree)			
Before/during secondary school	0.281	0.329	0.620
Before/during/after bachelor studies	1.355	1.721	0.348**
During master's studies	1.205	2.111*	0.707
Motivated by family	1.532	0.420**	0.824
Motivated by social circle	0.516*	1.533	0.978
≥ 6 friends with doctoral degree	1.510	2.543*	1.596
Last occupation before doctoral studies (rel. to Science/Engineering/Teaching professionals)			
Manager	1.285	0.604	0.337**
Other professional	0.889	0.307**	0.434*
Associated professional	0.614	1.094	1.650
Overeducated	0.689	0.603	0.582
Never employed	0.791	0.249	0.700
Constant	0.312**	0.284*	0.881
N	242	213	272
Pseudo R^2	0.0656	0.1797	0.1243
Hosmer–Lemeshow test, p-value (20 groups)	0.2217	0.3447	0.2217
Area under ROC curve	0.6863	0.7818	0.7309
Correctly classified	73.97%	72.77%	70.96%
Sensitivity	7.81%	54.88%	38.46%
Specificity	97.75%	83.97%	87.29%
AIC	311	283	354
BIC	398	367	444

*** p < 0.01; ** p < 0.05; * p < 0.10; † p < 0.15 Sensitivity is the probability of a true positive. Specificity is the probability of a true negative.

The difference between Latvia and the other two countries might be related to different construction of this variable: post-coding and self-reporting, respectively, see Section 4.3.2.

There is no systematic difference in motivation between males and females. In Lithuania, single females tend to pursue personal goals more, but in the overall model, this effect is counterbalanced by those observed in Estonia and Latvia.

The majority of respondents decided that they wanted to get a doctoral degree after getting a master's degree. In cases where this decision was made early – before or during secondary school in the overall and female-specific models or between graduating the secondary school

Table 4-13 – Odds Ratios after Sex-Specific Generalised Ordered Logistic Regressions of the Type of Goal: Constrained Part of the Models

	Total	Male	Female
Field of doctoral studies (rel. to Humanities)			
Business Management	1.795**	1.245	2.276*
Computer Science	1.043	0.703	1.142
Education	0.917		0.926
Engineering	1.114		1.538
Life Sciences		1.463	2.829***
Physical Sciences	1.191	1.236	1.208
Social Sciences	1.216		1.071
Other		0.786	
Change of field of study (rel. to Never changed)			
Changed at master's level	0.807	0.683	0.890
Changed at doctoral level	0.961	0.811	0.938
Changed at both master's and doctoral levels	0.715	0.984	0.579*
Started doctoral studies at age > 30 years	1.488**	1.694†	
Single female	1.342†		
Decided that would get doctoral degree (rel. to After getting master's degree)			
Before/during secondary school	0.323***	0.485	
Before/during/after bachelor studies	0.923	0.757	0.919
During master's studies	1.122	0.833	1.315
Parent has doctoral degree		1.305	0.466*
Motivated by family	0.943	0.780	1.086
Motivated by social circle	0.942	0.938	1.013
$\geq$ 6 friends with doctoral degree		1.406	1.617*
Unemployed $\geq$ 6 months when deciding on doct.st.	1.280	3.030***	0.771
Last occupation before doctoral studies (rel. to Science/Engineering/Teaching professionals)			
Manager	0.637*	0.558	0.599†
Other professional	0.567**	0.760	0.469**
Associated professional	0.961	1.031	0.866
Overeducated	0.617*	0.784	0.548*
Never employed	0.614	0.620	0.618
Country of doctoral studies (rel. to Latvia)			
Estonia	0.479***	0.407**	0.537**
Lithuania	0.671**	0.401**	0.855
Constant	1.654*	2.383*	1.618
N	726	233	487
Pseudo R^2	0.0495	0.0845	0.0653
Correctly classified	39.48%	23.62%	30.03%
Correctly classified: multinomial logit	41.28%	25.31%	31.27%
AIC	1845	594	1250
BIC	2024	701	1401

*** p < 0.01; ** p < 0.05; * p < 0.10; † p < 0.15 The table shows the variables where the proportional odds assumption holds, see Table 4-14 for the rest of results. The base goal type is *mostly personal*. Respondents studying in the education field removed from the male-specific model due to low number of observations. Classification accuracy determined by comparing the goal type with the highest predicted probability with the actual goal type. Being much more parsimonious than multinomial logit, the models have only slightly smaller classification accuracy.

and starting master's studies in Lithuania – respondents tended to pursue personal, rather than labour-market, goals. In Latvia, however, making this decision during master's studies shows a stronger labour-market orientation than for respondents who decided after graduating their master's programme.

Table 4-14 – Odds Ratios after Sex-Specific Generalised Ordered Logistic Regressions of the Type of Goal: Unconstrained Part of the Models

	Mostly Personal	Primarily Personal	Primarily Labour-Market
Total model			
Field of study: Life sciences	2.627***	2.317***	1.333
Field of study: Other	1.814*	1.268	0.620
Parent has doctoral degree	0.490***	0.877	1.114
$\geq$ 6 friends with doctoral degree	1.656*	1.710**	0.924
Male model			
Field of study: Engineering	0.737	0.737	2.218†
Field of study: Social Sciences	2.835*	1.054	1.054
Female model			
Field of study: Other	2.366**	1.361	0.481
Started doctoral studies at age $>$ 30 years	1.110	1.776**	1.745*
Decided that would get degree at/before sec. school	0.137***	0.174**	0.007***

*** p < 0.01; ** p < 0.05; * p < 0.10; † p < 0.15 See the constrained part of the models in Table 4-13. The base goal type is *mostly personal*.

The suggestions or other kinds of inducement given by family in the Latvian model or social circle in the Estonian model tend to be related with pursuing personal, rather than labour-market, goals. In other models, however, these two variables are not significant. A similar effect is observed for respondents having at least one parent with a doctoral degree in the overall and female-specific models. On the contrary, a sizeable presence of people with doctoral degrees in the social circle of respondents reflects a larger focus on the labour market in Latvia, as well as in the overall and female-specific models.

Labour-market experience is another factor having strong ties with the goals pursued when enrolling in doctoral studies. Respondents aged above 30 when starting doctoral studies are strongly more labour-market oriented than their younger colleagues in all models except the Estonian and the male-specific. Unemployment of at least 6-month duration before starting doctoral studies is a very important factor for males: such respondents favour labour-market goals much stronger than males who were employed at that time. Finally, the last occupation before doctoral studies is another factor related to the type of goal. Compared to respondents who were employed in a scientific, engineering or teaching profession, managers (overall and in Lithuania), other professionals (overall, in Latvia, Lithuania and for females) and those who were overeducated (overall and for females) have a strong tendency to pursue personal goals.

Finally, comparing the three countries, in accordance with the descriptive analysis in Section 4.3.3.1, Estonian doctoral students are the most personally-oriented and Latvian students are the most labour-market oriented when choosing to study for a doctoral degree.

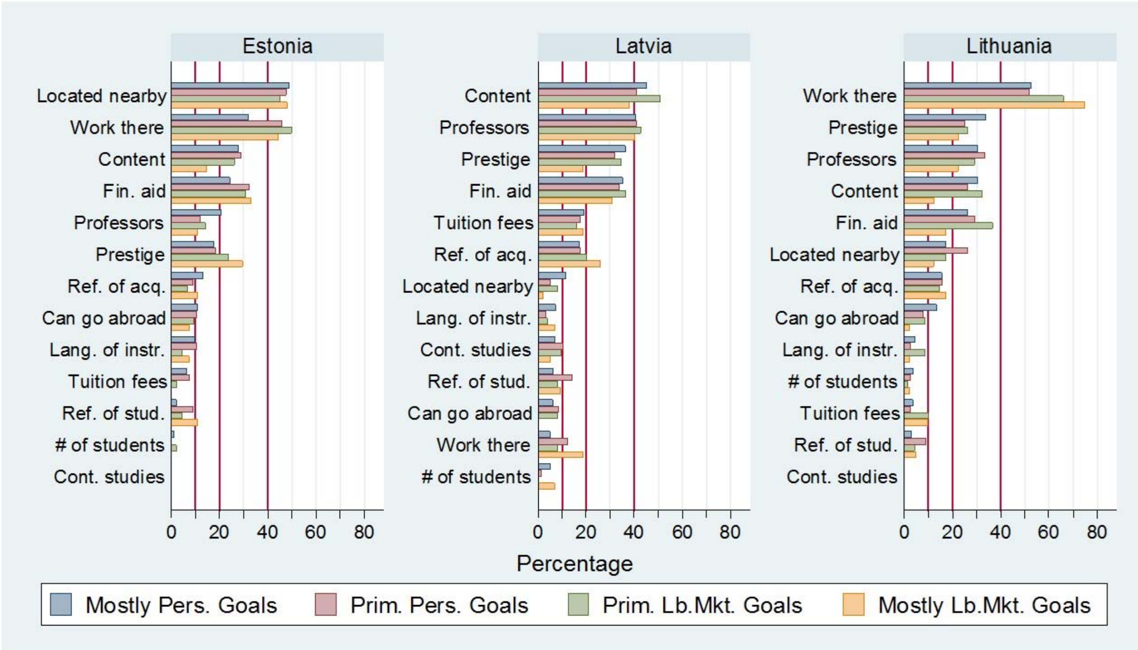

Figure 4-8 – Most Important Factors when Choosing an HEI for Doctoral Studies, by Country and Goal Type

Abbreviations: “Fin. aid" stands for Financial aid, “Ref. of acq.” for References of acquaintances, “Lang. of instr.” for Language of instruction, “Ref. of stud.” for References of students and “Cont. studies” for Continued studies.

4.3.3.3 Choosing the Higher Education Institution for Doctoral Studies

Baltic HEIs should be interested in not only what goals potential students pursue when deciding whether to enrol in a doctoral programme, but also in the HEI’s attributes they value or the information sources they would be using to support their decision on the place for studies. This subsection sheds light on these issues.

Figure 4-8 allows creating four groups of important HEI characteristics for each country: very important (mentioned by around 40% of respondents), moderately important (mentioned by around 20% to 30% of respondents), less important (mentioned by around 10% of respondents) and not important (mentioned by less than 10% of respondents). In most cases, respondents pursuing different types of goals have close opinions on the level of importance of each of the 13 factors mentioned in the figure, but for certain factors, respondents focused mostly on the labour market have substantially different opinions from others.

The most important factors are nearby location and working at the HEI (but not for the fully personally-oriented) in Estonia; programme content, professors’ reputation, HEI prestige and financial aid (both latter not for the fully labour-market-oriented) in Latvia; and working at the HEI and financial aid (only for the primarily labour-market-oriented) in Lithuania.

In Estonia, the second level of importance is attributed to working at the HEI (in case of pursuing mostly personal goals), programme content (except for the fully labour-market-oriented), financial aid, prestige and professors' reputation (only for the fully personally-oriented). In Latvia, these are prestige and financial aid (both for fully labour-market-oriented), as well as tuition fees and references of acquaintances, but also working at the HEI if pursuing mostly labour-market goals. In Lithuania, the moderately important goals are HEI prestige, the reputation of its professors, financial aid, references of acquaintances, programme content and HEI location (except for the fully labour-market-oriented for both latter factors).

Estonian students find professors' reputation (if not pursuing mostly personal goals), references of acquaintances, possibility to go abroad, language of instruction, tuition fees (not important if the main goal is a labour-market goal) and references of students as less important goals. Latvian students assign to this category close location (although the fully labour-market-oriented find it unimportant), language of instruction, references of students and possibility to go abroad (not mentioned by the fully labour-market-oriented). In Lithuania, the less important factors are location (for those pursuing mostly labour-market goals), possibility to go abroad (except for the fully labour-market-oriented), tuition fees (if pursuing a labour-market main goal), language of instruction and references of students (both for selected subgroups of respondents).

Several issues have to be noted in these results. Firstly, in Estonia and, especially, in Lithuania, students most frequently mention working at the HEI as the main reason they chose it for doctoral studies, while in Latvia, this factor is near the bottom of importance. For Estonians, whether the HEI is located nearby is at least as important as whether the respondent works there, but this factor is much less important in the other two countries. On the contrary, Latvians are more focused on the quality of knowledge they would gain (programme content and professors' reputation). This should be contrasted with fully labour-market oriented individuals in Estonia and Lithuania, for whom programme content is a less important factor, mentioned only by 10–15 per cent of respondents.

Secondly, financial aid was mentioned by 25 per cent to 35 per cent of respondents in all three countries, but tuition fees were mentioned less often and twice as much in Latvia (20%) than in Estonia or Lithuania (10% or less).

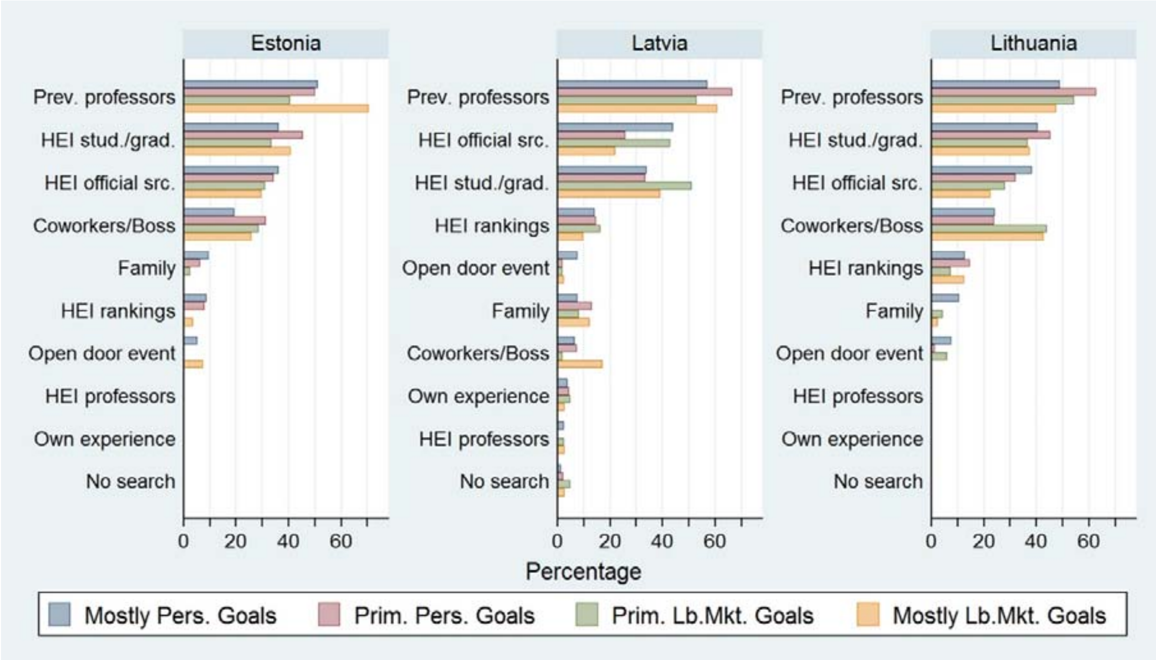

Figure 4-9 – Most Important Information Sources about the HEI for Doctoral Studies, by Country and Goal Type

Thirdly, around 10 per cent of Latvian respondents mentioned that they simply continued studies at the HEI where they had already studied before, so that they did not evaluate other places for studying; there are no such respondents in Estonia or Lithuania.

The most important sources of information about the HEI(s) or the doctoral programme(s) respondents considered choosing are similar in all three countries (see Figure 4-9). Professors whom respondents knew from their master's studies are the most frequently mentioned source of information, especially by Estonian students pursuing mostly labour-market goals. The students of and graduates from the HEI and HEI's official sources of information (website, accounts in social networks, etc.) follow, but it should be noted that the fully labour-market oriented students pay up to twice less attention to official information sources and trust the experience of other people more. The opinion of co-workers or superiors at work are a similarly important factor in Estonia and Lithuania, but less so in Latvia. Respondents having a labour-market goal as the main goal more frequently mention this source of information. In Lithuania, this source is nearly as important for such respondents as professors whom they know from master's studies. Less important sources, but still mentioned by 10–15 per cent of respondents, are family, HEI rankings and open door events.

4.3.4 Discussion and Conclusions

As discussed in Section 4.3.1, the demand for doctoral programmes has started falling recently, and the trend will, most likely, continue to face downwards purely for demographic reasons. Combined with dropping demand for lower-level tertiary education, this presents certain challenges for HEI management. The competition in the market for doctoral studies will increase within each of the Baltic countries, but also across these countries, as HEIs will seek to attract students from abroad. This study showed particular points of interest of individuals studying in each of these countries that HEIs should pay attention to when creating their marketing strategies.

Firstly, HEIs should understand why the demand for doctoral studies exists. In general, in all three Baltic countries, doctoral studies primarily provide opportunity to continue learning and/or research experience, contribute to the development of science or society or world in general, achieve something new and ameliorate career prospects. From these four goals, the first three can be considered as personal goals and only the last one as a labour-market goal. Thus, in general, HEIs should primarily appeal to the intrinsic benefits of getting a doctoral degree in their marketing communication.

Communication should, of course, be tailored to the target market, and field of study is an important factor that should be taken into account. In most fields in all three countries, personal goals are more important than labour-market goals, although Latvian students put more importance on labour-market goals than Estonians or Lithuanians. Econometric analysis showed that Estonian and Latvian life sciences students and Latvian business management students are significantly more focused on labour-market goals than humanities students, while Lithuanian computer science students are significantly more focused on pursuing personal goals. Men studying in engineering tend to emphasise labour-market reasons of the need for a PhD degree, as do female students in management, life sciences and a field group that contains architecture, law, library and information science, psychology, public administration, social work and sports. Male social science students, on the contrary, tend to want the degree for personal reasons.

The regularities in the goals prospective doctoral students pursue across fields of study are the most convenient for HEIs to take into account, but these are not the only ones. The choice and ranking of goals also depend on previous educational track, labour market experience and family and social circle. In particular, it was found that compared to doctoral students who worked as science, engineering or teaching professionals when starting doctoral studies, managers, other professionals and overeducated individuals tend to pursue personal goals rather than labour-market goals. This result for the overeducated is somewhat unexpected, as one might have anticipated that they would like to improve their labour-market position in the first place.

Secondly, after understanding which goals individuals with certain background and willing to study in a certain field pursue, HEIs should emphasise those of their attributes that are valued by these individuals. These vary strongly by country. In Estonia and Lithuania, students generally choose the HEI where they have been working and, specifically in Estonia, which is located nearby. Latvian students, in contrast, have programme content and professors' reputation as the factors of primary interest. Financial aid is more important than tuition fees, so the former should be emphasised more than the latter.

Finally, after understanding what to emphasise and for whom, HEIs should choose appropriate channels to communicate this information. Two of the three channels considered the most important by prospective doctoral students are not directly controllable by HEIs. These are (1) the professors whom students know from their master's studies and (2) the students of and graduates from the HEI. Both reflect the importance of the trust factor: there are solid grounds to trust the opinions of such professors and students/graduates for clear reasons. Official information sources of the HEI are only on the third place by importance, but this source can be controlled by the HEI and this is exactly the place where it can publish the outcomes of its doctoral programme related to personal and/or labour-market goals, whichever are most important for the target market. HEIs should also take into consideration that the opinions of the co-workers and the superior of the prospective doctoral student are very important for him/her if his/her primary goal is a labour-market goal. This implies that HEIs should work closely with the potential employers of their graduates to explain the benefits of a degree not only at lower education levels, but also at doctoral level.

4.4 Summary

This chapter considered some of the effects of overeducation on the position of individuals in LEMS. Its most important results are:

- Overeducation generally does not affect the job satisfaction of bachelors, but strongly decreases that of masters

- Most important primary job-related factor groups affecting job satisfaction (in order of importance) are content, risks and compensation; risks are, thus, very important and should not be overlooked, as typically done in job satisfaction studies
- Support activities (help from co-workers and ability to manage own working time) are also important for job satisfaction, as they allow to mitigate stressful situations
- Masters are more sensitive than bachelors to career opportunities, variety in work, and whether learning is required in the job; while bachelors are more sensitive to the risk of moving to a less interesting job and monetary compensation
- Overeducation increases the propensity to quit and affects the probability to quit stronger than overskilling; the strongest effect is on individuals who are both overeducated and overskilled
- Vocational education is not necessarily a shield from overeducation and its effect on the propensity to quit: fully vocational graduates are more likely to both become overeducated and quit if becoming overeducated than fully academic graduates are
- The most popular goals individuals pursue when choosing to get a doctoral degree are (1) continue learning/research experience, (2) contribute to the development of science, society or world in general, (3) new achievement and (4) better career prospects
- Most doctoral students pursued personal goals when enrolling into their doctoral programme in the Baltic countries
- Latvian doctoral students are stronger focused on the outcomes of a PhD degree related to the labour market than Estonian or Lithuanian students
- The main factors on which the type of goal pursued when enrolling into PhD programmes depends are field of study, previous educational track, labour market experience and family and social circle
- When enrolling into PhD programmes, overeducated individuals pursue personal goals more frequently than science, engineering or teaching professionals
- HEI attributes valued by prospective doctoral students vary strongly by country: Estonians choose the HEI located nearby and where they have been working, Lithuanians the HEI where they have been working, while Latvians choose the HEI based on the content of the doctoral programme and the reputation of HEI's professors

- The most important channels for the information on doctoral programmes are (1) the professors whom students know from their master's studies, (2) the students of and graduates from the HEI and (3) HEI's official information sources
- Prospective doctoral students pursuing a labour-market goal put substantial weight on the opinion of their co-workers and the superior when choosing the HEI for doctoral studies

5 SOCIAL NETWORKS IN LABOUR–EDUCATION MARKET SYSTEM: AGENT-BASED MODELLING

Previous chapters were dealing with the empirical analysis of different phenomena of LEMS, primarily related to the causes of and effects from overeducation. In this chapter, I will employ agent-based modelling (see Appendix D for the background on this modelling method) to study the implications of the dependence of agent decisions on social networks on the dynamics of LEMS. Three models will be discussed. The first model considers a labour market where job satisfaction is a unified mechanism for guiding agent behaviour in choosing among vacancies and deciding to quit the current job. The second model defines the mechanisms of choosing the field of study and determining the quality of match between an agent and a job and studies the resulting dynamics of LEMS, as well as migration decisions. The third model studies the effectiveness and efficiency of one of the possible policy responses to the problem of overeducation: restricting access to higher education on a random basis. As previously, the final section summarises the chapter.

5.1 Job Satisfaction as a Unified Mechanism for Agent Behaviour in a Labour Market with Referral Hiring

Existing agent-based models of labour market, even if include referral hiring or job search through social networks, impose a simplistic choice mechanism among vacancies on their agents. According to it, agents either accept a random vacancy or choose the best vacancy only based on the proposed wage. The only models that do introduce a dependence of agent's choice on non-wage characteristics of the job are Ballot (2002) and Ballot, Kant and Goudet (2013), but they include one overall non-wage factor and do not model different non-wage factors separately.

Individuals use a more sophisticated mechanism to choose the vacancy to apply for. Speaking abstractly and assuming perfect rationality of individuals, it would be plausible to assume that the individual operates with a certain utility function defined over different vacancies and chooses to apply for a vacancy that maximises this function. While wage is an important factor when selecting a vacancy, it is not the only one. Hence, the utility function depends also on other characteristics of the vacancy.

The assumption of perfect rationality in this case has two weaknesses. Firstly, it implies that individuals are willing and able to find all existing vacancies. This might be relaxed by allowing individuals to filter the vacancies during formal search based on some characteristics – most easily, the proposed wage, which individuals should be able to easily compare with the reservation wage. This assumption is plausible, given the abundance of electronic job search catalogues with filtering capabilities. Moreover, individuals might instead use their social networks and inquire about open jobs in the firms where their friends or acquaintances work. A related assumption might be that individuals apply for all vacancies that maximise their utility functions, but this might be relaxed by putting a cap on the number of jobs an individual applies for in one period.

Secondly, perfect rationality implies that individuals are able to correctly compute the utility of the given vacancy. This might be relaxed by providing a mechanism for them to regularly reassess the utility of the current job. In other words, individuals assess the vacancy based on the information they have *a priori*, but after becoming employed in that job, they see the reality, adjust the parameters of the job and re-compute the utility function.

It is desirable to use a unified mechanism for modelling individual behaviour in the labour market. In other words, it is desirable to model the *same* utility function to guide individuals' decisions on both selecting the vacancy to apply for and quitting the current job. Job satisfaction (JS) is known to be a strong predictor of the decision to quit and has been intensively studied in the occupations with high turnover (Parry, 2008). It is tempting to use JS also as the mechanism for selecting the best vacancy. Indeed, if JS is a result of an interplay between the values workers attach to job facets and the extent to which these values are satisfied (Kalleberg, 1977), then a given vacancy appears to be the best for a given individual if he/she *expects* it to satisfy these values to the highest extent or, in other words, it has the highest *expected* JS. Thus, JS appears to be a good candidate for a unified mechanism of both selecting the best vacancies and deciding to quit.

The first attempt to implement this idea and introduce JS in an agent-based simulation of the labour market with referral hiring was done in Tarvid (2011). There, JS depended on the relative wage and the social network component of the vacancy/job. However, JS has many more components than financial rewards and social support. There is general agreement that, besides these, an important role is played by intrinsic job attributes, career growth, job security and working conditions (see Section 4.1.1).

In this section, I take the idea of Tarvid (2011) further and propose to include other important facets of JS: job content and career opportunities. This mechanism is then integrated into a labour-market model with referral hiring and informal job search. The aim is to see how the structure of JS affects such key characteristics of the output of the model as the distribution of firm size, labour-market status of individual agents and wage distribution and, hence, whether it pays off to increase the complexity of JS structure that affects the decision-making processes of individual agents in the labour market.

The generalisation of the JS function is not the only contribution of this work over Tarvid (2011). Other contributions include the addition of job type, further elaboration on firm inside dynamics, adding the mechanism of birth and death of firms, changing vertex power distribution in social networks from lognormal to scale-free, more careful parametrisation of the model based on empirical data and other more minor changes in the model.

The section is structured as follows. The following subsection presents the job satisfaction mechanism used here in detail. Then Section 5.1.2 sets up the labour-market model and Section 5.1.3 sets parameter values based on existing empirical data. Results are discussed in Section 5.1.4. The last subsection concludes.

5.1.1 Job Satisfaction Mechanism

As in Tarvid (2011), I divide JS in two components: expected JS, s_{ijf}^{e}, and actual (or current, as called in Tarvid (2011)) JS, s_{ijf}^{a}. Both are defined for agent i relative to job j at firm f. In other words, I introduce the dependence of JS on the firm, whereby there is certain correlation in JS for jobs inside a firm. This reflects the perception that some companies are *in general* better employers than others.

JS is modelled as a multi-faceted concept. It consisted of wages and social support (mainly from co-workers) in Tarvid (2011). This does cover compensation and support facets, but does not take into account job content and career opportunities. Including these latter facets introduces substantial difficulties. The former two facets can be modelled objectively (with their relative importance depending on some agent-specific weight), in the sense that the agent can precisely measure the wage it will receive and the number of its friends (approximating social support) working in a concrete job. In contrast, job content and career opportunities are vague concepts, more related to perceptions rather than to hard data. Nowadays, nearly every job advertisement speaks about an “interesting” job with “ample” career opportunities, which individuals have to interpret in the context of their existing knowledge about the firm and the job under consideration.

To make the matters simple, I assume that there are two types of jobs. The first type has ample career opportunities and high variety (which approximates content), while the second has limited career opportunities and low variety. The former jobs can be represented (and will be calibrated) by non-manual jobs (International Standard Classification of Occupations (ISCO) major groups 1 (managers) through 5 (service workers)), while the latter by manual jobs (ISCO major groups 6 (skilled agricultural and fishery workers) through 9 (elementary occupations)).

While agents perceive, e.g., manual jobs as having limited career opportunities, the perception of career opportunities for a given vacancy for such job depends also on the firm that posted it. Again, I assume that the only characteristic of the firm important for the perception of both career opportunities and variety at a given vacancy is the size of the firm.

Thus, the perception of job content and career opportunities depend on job type (manual vs. non-manual) and firm size.

Hence, ***expected JS*** is a function of:[42]

- Monetary compensation defined as the ratio of the expected wage of agent i in job j in firm f to its reservation wage, w_{ijf}/w_i^r
- Social support defined as the ratio of the number of friends of agent i in firm f (which I will refer to as "local friends") to the total number of its friends (i.e., the share of its friends working in firm f), n_i^f/n_i
- Job variety defined as a function of job type and firm size, $v\{T(j), S(f)\}$
- Career opportunities defined as a function of job type and firm size, $c\{T(j), S(f)\}$

The functional form of expected JS is as follows:

$$s_{ijf}^e = \Lambda\left\{6\left[\frac{w_{ijf}}{w_i^r} - 1\right]\right\} + \left[2\Lambda\left\{\frac{6n_i^f}{n_i}\right\} - 1\right] + v\{T(j), S(f)\} + c\{T(j), S(f)\}, \tag{5-1}$$

where $\Lambda\{\cdot\}$ is the logistic function, whose range is $[0, 1]$. The logistic function makes JS increase with monetary compensation and social support, but the return to these factors in terms of JS is decreasing (any next dollar or local friend increases JS less). Importantly, it also bounds the range of JS. The factor of 6 appears because $\Lambda(6) \approx 1$ and $\Lambda(-6) \approx 0$. Thus, the first summand approaches zero when $w_{ijf} \ll w_i^r$ and one when $w_{ijf} = 2w_i^r$. The second summand is zero when the agent has no local friends ($n_i^f = 0$) and approaches one when it has all friends working with firm f. Functions $v(\cdot)$ and $c(\cdot)$ also have the range of $[0, 1]$; they will be defined in Section 5.1.3. Thus, s_{ijf}^e can take values in $[0, 4]$.

[42] I use parentheses and braces in the definition of functions and square brackets to group expressions; i.e., $f(x + y)$ and $f\{x + y\}$ should be read as "function f of $x + y$," while $f[x + y]$ should be read as "f multiplied by $x + y$."

Algorithm 5-1 – Monthly Actions in the Labour Market

```
if start of year then
   New population added
   Persons aged over ā retire
end if
Non-start-up firms with zero workforce die
if start of year then
   Firms select annual workforce change
   Firms select wage change factor
end if
Update labour-market experience of persons
Create new firms
Firms update wages for expiring contracts
Firms change workforce and/or publish vacancies
Persons update current job satisfaction and consider starting on-the-job search
Persons update reservation wage and apply for vacancies
Firms send acknowledgements to selected persons
Persons reply to the best acknowledgement, quit current job if needed and start working
Firms update or remove failed vacancies
```

Note that in its current form, (5-1) assumes that all four components of JS are equally important for the individual. This need not be the case, and in a more general model, the components have to be weighted. In Tarvid (2011), where JS had two components, the weight of the social component was equal to the importance of friends, an internal immutable characteristic of the agent, which also determined the maximum number of friends it could have. In this more complex structure of JS and with a different method of generating social network (see the next subsection), there appears to be no other way to distribute weights across agents than to do it randomly. This could be done in the future, but for now, I will keep the current setup with all components weighted equally.

Actual JS represents the dynamics in the facets of expected JS (mainly, wage and social support) and all other factors gauged by a normally distributed random disturbance ξ:

$$\Delta s^{a}_{ijf}(t) = \Delta s^{e}_{ijf}(t) + \xi. \tag{5-2}$$

Delta (Δ) is used here as standard difference operator for time-dependent functions, i.e., $\Delta f(t) \equiv f(t) - f(t-1)$. By construction, the expected JS is always in $[0, 4]$. The value of the actual JS is reset at the closest boundary of this interval if (5-2) gives out-of-boundary values.

5.1.2 Model Specification

There are two types of agents: persons and firms. The model includes only the labour market; in particular, the education market is ignored. The degree of match between the person and the job is controlled through job requirements published in vacancies, see the job search mechanism below.

Timing is discrete with one period representing one month. Most actions in the labour market are done on a monthly basis. The only exceptions are changes in population (inflow of N new school or university graduates and retirement after living for $\bar{a}$ years, the "retirement age," in the model) and in wages (standard assumption about wage stickiness), which happen annually (every 12 periods).

The overall view of the monthly labour-market actions of persons and firms related to job search are summarised in Algorithm 5-1. The rest of this subsection describes its steps in detail.

5.1.2.1 Job Search

There is a unique vacancy list in the market, which everyone is able to access (e.g., a country-wide job search website). Firms post vacancies on this list and persons may browse it to find new jobs, which is called ***formal job search***. Alternatively, persons can choose to search for jobs ***informally***, using their friends that are employed. Implicitly, I assume that all employees in a given firm are informed about all its vacancies.

A vacancy is a quadruple (f, T, x, w), where

- f is the firm hosting the vacancy
- T is the type of job (manual or non-manual)
- $x \in \mathbb{Z},\ 0 \leq x \leq \overline{x},$ is the minimum required working experience measured in years; $\overline{x}$ is the sufficient experience, which is common for all vacancies
- $w \in \mathbb{Z},\ w \geq w_m,$ is the proposed wage rate at the required experience x; w_m is the minimum wage, which is common for all vacancies

Between the minimum required experience x and the sufficient experience $\overline{x}$, wage changes linearly with experience x_i:

$$w(x_i) = \begin{cases} w\big[1 + q(T)[x_i - x]\big], & x \leq x_i \leq \overline{x} \\ w\big[1 + q(T)[\overline{x} - x]\big], & x_i > \overline{x}, \end{cases} \tag{5-3}$$

where $q(T)$ is a constant specific for each job type T.

Every person i knows its actual experience x_i and reservation wage w_i^r and, for each vacancy, is able to find out the wage it will be paid. The person decides probabilistically whether to use formal or informal search. In both cases, the person creates an ***awareness list*** of matching vacancies (i.e., set $\{v | x(v) \leq x_i \wedge w(v, x_i) \geq w_i^r\}$), see below on the additional restrictions on the vacancies in this list for persons engaged in on-the-job search. As information-processing capabilities of agents are limited, they consider not more than K matching vacancies. If the person uses informal search, it picks its employed friends at random and asks for appropriate

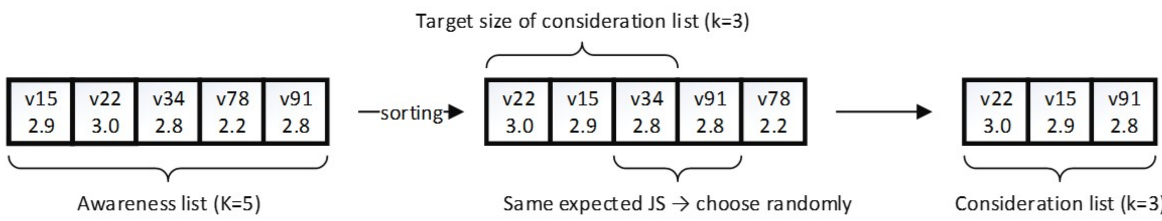

Figure 5-1 – Example of Creating of a Consideration List from the Awareness List

In this example, the awareness list consists of 5 vacancies: v15, v22, v34, v78 and v91. The consideration list should have 3 vacancies (because $k = 3$). After sorting by expected JS, the top two vacancies (v22 and v15) are added to the consideration list. But because v34, the kth vacancy after sorting, has the same expected JS as v91 (2.8), one of them should be chosen randomly as the third vacancy in the consideration list. In the example, v91 is chosen. Thus, the consideration list consists of v22, v15 and v91.

vacancies at their firms (all vacancies are published in the vacancy database, so there are no vacancies that employees know about but that are not published in the database). All appropriate vacancies of which the friend informs the person are put in the person’s awareness list. This process of asking random friends for vacancies continues until either all friends were asked or the length of the awareness list reaches K. If the person chooses to use formal search or it has not found any vacancies during informal search, it simply fills the awareness list by K random appropriate vacancies from the vacancy database.

The person then creates a ***consideration list*** from the awareness list. It sorts the vacancies from the awareness list descending by expected JS. Then top k vacancies form the consideration list (see Figure 5-1 for an example and further clarification). The person sends applications to all vacancies in the consideration list. Both k and K are the same for all persons.

Small (less than 25 employees) and medium-sized (25–499 employees) firms may probabilistically choose to use the referral hiring mechanism when choosing candidates to employ. In this case, if in the list of applications for the vacancy, there are candidates having friends employed in the firm, the firm chooses randomly from these candidates; otherwise, it chooses randomly from all candidates. Large (500+ employees) firms do not look on the existing social ties of candidates with their employees and choose randomly from all applicants.

Successful candidates receive acknowledgements. For a given vacancy, exactly one of the applicants receives acknowledgement. If a person receives acknowledgements from several vacancies, it chooses the job with the highest expected JS. It then starts working immediately in that job.

The reservation wage of a working person is equal to its current wage. For a person with no working experience, it is given by the minimum wage. For an unemployed, it decreases with the length of unemployment measured in months, starting from the last wage, but is bounded from below by the minimum wage. The decrease occurs with constant elasticity φ, which is the same for all persons. Thus, the longer the person is unemployed, the lower wage it is ready to accept.

If for an employed person, its current JS falls below the critical level, which is the same for all persons, it starts on-the-job search. Note that individuals start thinking about quitting the job when they feel that their job is "unsatisfactory." What different individuals mean by this is reflected by the combinations of the values of JS facets, but all these combinations lead to JS falling below certain boundary. Using a relative, rather than absolute, measure of JS, it seems realistic to assume that this boundary is the same for everyone. In contrast with the unemployed, who consider all matching vacancies, the employed consider only those matching vacancies with the expected JS being higher than their current JS. If accepted for a vacancy, such person quits the current job and immediately starts working in the new position.

5.1.2.2 Dynamics inside Firms

In the first month of the calendar year, firms plan changes in workforce and wages. Then, every month of that year, they implement these decisions.

If firms decide to change workforce by δ persons this year, they implement it by changing workforce every month by $\delta/12$ persons (rounded up or down as required). If the firm decides to expand this year, it publishes the according number of new vacancies every month throughout the year and all vacancies substituting the employees who quit after on-the-job search.

Each vacancy for a new position is created with required experience drawn from exponential distribution with mean λ and probabilistically chosen type of job (manual or non-manual). The corresponding proposed wage w is set to the average wage of the firm's employees with that experience and job type. If no such persons are currently employed in the firm, it makes interpolation from the average wage it pays employees with the experience nearest to x and the same job type. Firms having no employees of this job type take average current wages at experience x and the selected type of job in the labour market. If there are no such persons in the labour market, average wages are interpolated from employees of the same job type with nearest experience.

The vacancy may fail to attract applications if it has low expected JS. For a given person, the firm can increase the expected JS of the vacancy only by increasing the proposed wage, as other components – social support, job variety and career opportunities – are fixed. At the same time, the firm cannot decrease the required experience to attract additional candidates, as it needs qualified employees.

The length of vacancy life is l periods. After every period where the vacancy did not find a match, the firm hosting that vacancy checks whether the wage the vacancy proposes is not below the average wage in the labour market for the proposed experience and job type. If it is below the average, the firm sets it to the average for the next period. If it is already above the average, the firm increases the proposed wage by a constant multiple $\nu > 1$. After l periods without success, the vacancy is removed from the vacancy database.

Firms also publish vacancies to substitute employees who just quit the firm either after reaching the retirement age or due to low JS (employees who were laid off are not substituted). In this case, the experience is set at the level of the experience of the worker who quit, x' (but still keeping it in $[0, \overline{x}]$):

$$x = \min(x', \overline{x}).$$

The type of job is left the same, and the wage is set as described above.

If the firm decides to contract this year, its behaviour is slightly different. Every month, the firm has to lay off a certain number of employees. Before actually laying off agents, it computes how many employees left it last month and first tries to implement the change in workforce by not publishing substituting vacancies. For instance, if the firm has to contract by 5 employees but 3 employees left it last month, the firm does not publish these 3 substituting vacancies and lays off only 2 employees. All lay-offs are made randomly.

Every firm changes wages once a year, in accordance with standard economic results on the stickiness of wages. Wages are changed for all jobs in the firm by the same factor. That factor is chosen from the set $\{w_d, 1, w_u\}$, where $w_d < 1$ and $w_u > 1$ with the corresponding probabilities $\{\pi_d, 1 - \pi_u - \pi_d, \pi_u\}$. For a given employee, the wage is changed in the month it was hired in the current job (if it occurred in month 3 last year, it is changed in month 3 this year, although the decision to change wages was made in the beginning of this year). In other words, I assume yearly contracts with fixed wages.

5.1.2.3 The Birth and Death of Firms

Every month, every person that is employed or unemployed, but not an entrepreneur, can create a new firm with probability μ. In case the person becomes entrepreneur but is employed, it quits the current job. Entrepreneurs are not subject to on-the-job search or quitting their companies, which makes the definition of their wages and, more broadly, JS unnecessary.

New firms choose their initial size from the uniform distribution over the possible sizes of a small firm (i.e., from $[1, 24]$). Then they publish all these vacancies immediately (rather than distributing them over the year, as more mature firms do). The mechanism of setting vacancy's characteristics is the same as described above.

Every period, firms with no workforce disappear from the model. Their owners become unemployed and start searching for a new job. New firms, however, are allowed to exist for Δt_s periods without workforce, during which they should find first employees. When a firm's owner is removed from the simulation, the firm continues to exist without an owner: no other person is assigned as a new owner.

The simulation starts with M empty firms, which start behaving as new firms. These initial firms do not have owners from the set of persons in the model.

5.1.2.4 Dynamics of Social Networks

Some of the persons that enter the model in the same period are interconnected, forming an ***initial social network*** of, e.g., secondary school or university friends (friends in the broad sense, meaning both close friends and acquaintances). They also have friendship ties with those persons already in the labour market (irrespective of whether they are employed or not), forming a ***mature social network***. These two sets of connections form the social network with which the person enters the model.

This initial social network and the mature social network are separately generated using the Duplication model $\mathcal{D}(\rho)$ (Chung & Lu, 2006, Ch. 4) parametrised to build a scale-free social network, having many low-degree vertices and a few high-degree vertices (see Section 2.2.2). Empirical research indicates that such networks approximate real-world social networks quite well (see Section 2.2.2). Both networks are built with the same parameter $\rho \in (0, 1)$.

The person also changes its social network when it gets a new job. It tries to make new friendship ties with the employees of the firm hosting the new job and their friends. Naturally, the Duplication model with the same parameter ρ as above is used for creating it. An existing employee can refuse friendship with the new employee, so friendship ties with employees are created probabilistically. The number of friends the person makes depends on the size of its current network, n_i. The total number of new connections cannot exceed $\lceil \psi n_i \rceil, \psi \in (0,1)$. After the person added new friends to its network, it removes the same number of existing friends from its network, starting from those subject to the longest unemployment spell. The person tries to make friends only with those employees working in the same job type as its new job.

Table 5-1 – Distribution of Career Opportunities

Extent of Agreeing, ESS coding (Fully disagree 1..5 Fully agree)	1	2	3	4	5
Parameter Value in the Model	**0**	**0.25**	**0.5**	**0.75**	**1**
Non-manual occupations					
Small firms (< 25 employees)	17%	27%	26%	25%	5%
Medium & Large firms	11%	26%	27%	30%	6%
Manual occupations					
All firms	22%	29%	23%	22%	4%

Source: Calculated from European Social Survey round 5 pooled data from 24 European countries (mentioned in text) over respondents employed at the moment of surveying. The original question is "How much do you agree or disagree with the statement 'My opportunities for advancement are good'?". Design and population weights applied. Non-manual occupations are ISCO major groups 1–5. Manual occupations are ISCO major groups 6–9. The table also defines the values and the distribution of the function $c(\cdot)$. In this case, it should be read, e.g., "$\Pr(c = 0.25|\text{small firm, non-manual occupation}) = 27\%$."

Table 5-2 – Distribution of Job Variety

Extent of Agreeing, ESS coding (Fully disagree 1..5 Fully agree)	1	2	3	4
Parameter Value in the Model	**0**	**0.33**	**0.67**	**1**
Non-manual occupations				
Small firms (< 25 employees)	7%	26%	32%	35%
Medium firms (25–499 employees)	5%	21%	33%	41%
Large firms (≥ 500 employees)	3%	19%	32%	46%
Manual occupations				
All firms	17%	28%	28%	27%

Source: Calculated from European Social Survey round 5 pooled data from 24 European countries (mentioned in text) over respondents employed at the moment of surveying. The original question is "How true about your current job is the statement 'There is a lot of variety in my work'?". Design and population weights applied. Non-manual occupations are ISCO major groups 1–5. Manual occupations are ISCO major groups 6–9. The table also defines the values and the distribution of the function $v(\cdot)$, e.g., "$\Pr(v = 0.33|\text{small firm, non-manual occupation}) = 26\%$."

This dependence on unemployment spell length reflects both (1) homophily, i.e., that people make friendship ties with those close to them by some characteristics (in this case, higher tendency to connect with the employed if it itself is employed), and (2) the usefulness of the friend (the most useful friends in the labour market are employed friends, because they can inform about an open vacancy or refer to the employer about their friend pretending to get a job). The creation of ties with employees of the same job type also reflects homophily. The cap on the number of new friends and the substitution of new friends for existing friends also allow to constrain the size of the social network.

5.1.3 Parametrisation

European Social Survey (ESS) round 5 (Norwegian Social Science Data Services, 2010) has individual-level data on the perceptions of the employed in 24 European countries (Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Lithuania, the Netherlands, Norway, Poland, Portugal, Slovakia, Slovenia, Spain, Sweden, Switzerland and the UK) in 2010–11 about their jobs. In particular, there is information about their perceptions on job variety and career opportunities. Tables 5-1 and 5-2, respectively, show the distributions of respondents' perceptions grouped by firm size and occupations. Some distributions differ by firm size, while others – e.g., for manual occupations – do not depend on it. Functions $c(\cdot)$ and $v(\cdot)$ are defined probabilistically, given vacancy's characteristics, using these tables (see the notes below the tables).

Wage dynamics are set based on estimates available in the literature, see Table 5-3. The average effect on wages from an additional year of labour-market experience is around 2.5 per cent. Estimates based on US data show that it is around 1.5 times larger for the tertiary-educated than for the secondary-educated. Assuming that the former work in non-manual jobs and the latter in manual jobs, I set the wage–experience coefficient q at 3% for non-manual jobs and at 2% for manual jobs. The relationship between wage and experience is non-linear: returns are diminishing at higher experience levels and after 20–30 years (García-Mainar & Montuenga-Gómez, 2005), the wage–experience curve flattens (in some countries, wages start dropping afterwards). Here, I assume that wage changes linearly with experience (recall (5-3)) and stops depending on experience at 10 years of experience, which is the sufficient experience $\bar{x}$.

The distribution of annual workforce change is set in accordance with Amadeus data for European companies in 2010–2013, see Table 5-4. Firms randomly select the changes of their workforce, depending on their size and the direction of the actual workforce change in the previous year.

According to the job-search theory (Stigler, 1962), reservation wage falls with unemployment spell duration. Empirical studies used constant-elasticity models for quantifying this effect, but came to substantially different elasticities, ranging from -10% (Addison, Centeno, & Portugal, 2004) to -80% (Brown & Taylor, 2013). I set the elasticity at –20%, which means that for every 10 per cent increase in unemployment length, reservation wage falls 2 per cent.

The Duplication model parameter ρ is set so that the model generates a scale-free network with exponent $\beta \approx 2.61$, which is inside the range $[2, 3]$ typical for social networks (Chung & Lu, 2006). The number of direct connections in initial and mature networks are set so that the number of friends would be not more than 100–120, the maximum number of Facebook friends with which an individual interacted at least once (Wilson, Boe, Sala, Puttaswamy, & Zhao, 2009, Fig. 15).

Of other parameters shown in Table 5-5, the most difficult to set was the probability of starting a new firm, μ. Too high values of μ result in generating too many small firms, which also

Table 5-3 – Empirical Data on Wage Returns on Year of Experience

Author	Country	Sex	Education Level		
			Secondary	Tertiary	Total
Battisti (2013)	Italy	Men			2.9%
		Women			1.1%
Williams (2009)	UK	Men			2.2%
Myck & Paull (2001)	UK	All			7.4%
García-Mainar & Montuenga-Gómez (2005)	9 countries	All			1.06%–3.67%
Connolly & Gottschalk (2006)	USA	Men	1.8%	2.5%	
		Women	1.9%	3.2%	
Munasinghe, Reif, & Henriques (2008)	USA	Men	3.1%	5.8%	
		Women	3.7%	8.2%	

Estimates correcting for unobserved heterogeneity were taken where available. Where cumulative effect of experience over several years was given, it was converted into annual effect assuming that the effect from every additional year of experience is the same (e.g., a 20% cumulative effect over 10 years would be converted into a 2% annual effect). From García-Mainar and Montuenga-Gómez (2005), data were taken only on countries with positive relationship between wage and experience; the nine countries are Austria, Belgium, Germany, Greece, Ireland, Italy, Portugal, Spain and the UK.

Table 5-4 – Distribution of Annual Workforce Change

Firm Size / Distribution	Small ($\text{size} < 25$) absolute changes			Medium, subgroup 1 ($25 \leq \text{size} < 100$) absolute changes			Medium, subgroup 2 ($100 \leq \text{size} < 500$) % changes			Large ($500 \leq \text{size}$) % changes		
Δ Size	< 0	$= 0$	> 0	< 0	$= 0$	> 0	< 0	$= 0$	> 0	< 0	$= 0$	> 0
$[-50, -40)$								2				
$[-40, -30)$							1	2		1	1	
$[-30, -20)$				1			3	2	2	2	4	2
$[-20, -10)$				4	2	3	9	3	6	9	2	6
$[-10, 0)$	26	15	19	37	24	28	38	17	29	44	10	30
0	31	51	35	13	32	15	6	42	8	2	65	5
$(0, 10]$	40	31	42	38	36	42	32	21	35	34	13	40
$(10, 20]$	2	2	3	5	4	8	7	5	12	6	3	11
$(20, 30]$	1	1	1	2	1	3	2	2	5	2	2	4
$(30, 40]$					1	1	2	2	2			2
$(40, 50]$								2	1			
	100	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**

To be read: For a small firm whose size decreased last year, the probability of decreasing workforce in the range of 1 to 10 workers this year is 26%. For a medium-sized firm with at least 100 employees whose size increased last year, the probability of increasing workforce in the range of 0% (excluded) to 10% (included) this year is 35%.

Source: Amadeus database, accumulated over 2010–2013. The sample contains active firms operating in EU-28, Norway and Switzerland, but only those countries were taken in a particular year whose real GDP (Amadeus data) increased that year. Thus, the sample is based on firms operating in positive macroeconomic conditions. The distributions shown in the table are truncated versions of actual distributions with minor deviations after correcting for rounding errors. The truncations were done to eliminate extremely long tails, which are partly due to inaccuracies in reporting the number of employees by firms in consecutive years. Only 10-worker or 10-percent intervals with probabilities of at least 1% included in the distributions shown in the table.

remain small due to high competition for new employees. Too low values lead to extended periods of nearly full unemployment in the beginning of the model. The parameter was set at 10^{-5} to balance both effects.

Table 5-5 – Parameter Values

Parameter Name	Notation	Value
General		
Annual inflow of new persons	N	500
Retirement age, years	$\bar{a}$	30
Length of simulation, years		260
Initial number of firms	M	100
Job Satisfaction		
Current job satisfaction disturbance		
mean	$\mu(\xi)$	0
std.dev., % of JS interval	$\sigma(\xi)$	5%
Critical job satisfaction for on-the-job search		20%
Workforce Dynamics		
Prob. of referral hiring for small and medium-sized firms		0.40
Prob. of manual vacancy		0.30
Experience distribution parameter	λ	1.5
Wage multiplier when updating vacancy	ν	1.02
Wage–experience multiplier	q	
manual jobs		2%
non-manual jobs		3%
Sufficient experience, years	$\overline{x}$	10
Max vacancy age, periods	l	3
Job Search		
# of simultaneous applications	k	5
Max # of vacancies considered	K	50
Prob. of formal job search		0.30
Prob. of recommending friend to firm		0.90
Unemployment length elasticity of reservation wage	φ	-20%
Wage Specification		
Minimum wage	w_m	100
Wage dynamics		
Prob. of increasing wage	π_u	0.6
Factor of wage increase	w_u	1.02
Prob. of decreasing wage	π_d	0.1
Factor of wage decrease	w_d	0.98
Entrepreneurship		
Prob. of becoming entrepreneur	μ	10^{-5}
How long considered start-up, periods	Δt_s	12
Social Networks		
Duplication model parameter	ρ	0.45
Number of direct connections		
initial social network		30
mature social network		10
Prob. of creating social tie with existing employee when coming to new job		0.50
# of new friends to make on new job		
relative to # of current friends	ψ	0.50
if has no friends yet		5

5.1.4 Results

The model was implemented in Repast Simphony 2.1. The model started from scratch (N agents, M empty firms) and evolved throughout 250 years, which was sufficient to stabilise unemployment statistics. The following analysis is based on the data gathered during years 251 through 260. The analysis will compare the model where JS is based on (1) all four components

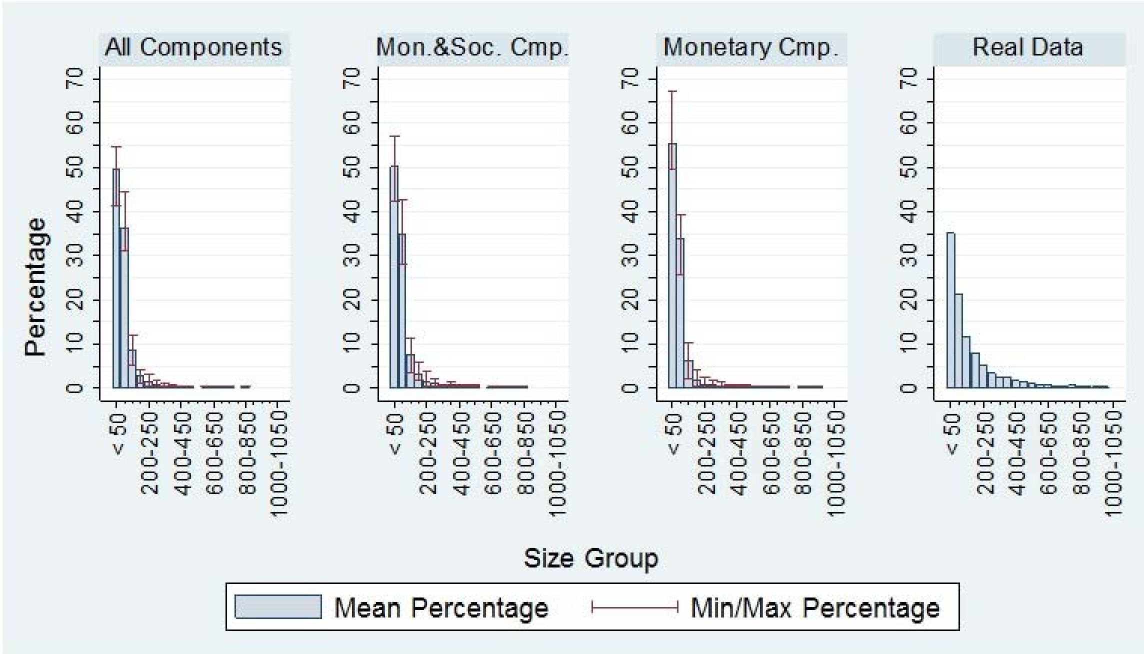

Figure 5-2 – Distribution of Firm Size at End-of-Year 260 Depending on the Composition of Job Satisfaction

Statistics calculated over 20 runs for each composition type of job satisfaction. The distribution shown only until the size of 1050 for compactness. No firms larger than that are generated by the FJS model. There are three larger firms in the MSJS model, with sizes of 1246, 1836 and 3541. The MJS model generates two larger firms, with 1161 and 2078 employees. Real data taken from Amadeus database, 2009–2013 (pooled data are not different from data in each specific year).

(abbreviated FJS for "full JS"), (2) only monetary and social components (MSJS) and (3) only the monetary component (MJS). For each of these JS structures, the model was run 20 times.

In this section, I will analyse the performance of the model along three lines. I will start with the distribution of firm size the model generates. Then I will continue to the status of persons in the labour market, including the relationship between unemployment and the number of vacancies known as Beveridge curve. Finally, I will look on wage distribution.

5.1.4.1 Firm-Level Statistics

The FJS model creates a more realistic distribution of firm size than the other two models (see Figure 5-2), which is formally supported by the Kolmogorov–Smirnov test (see Table 5-6). Compared with the MJS model, adding social component to JS decreases the share of firms with less than 50 employees. The largest firms are generated by the MSJS model, and this is, most likely, due to social network effects. Compared to the MSJS model, the FJS model decreases the range (but not the average) of firms sized less than 50 and pushes up the shares of larger firms.

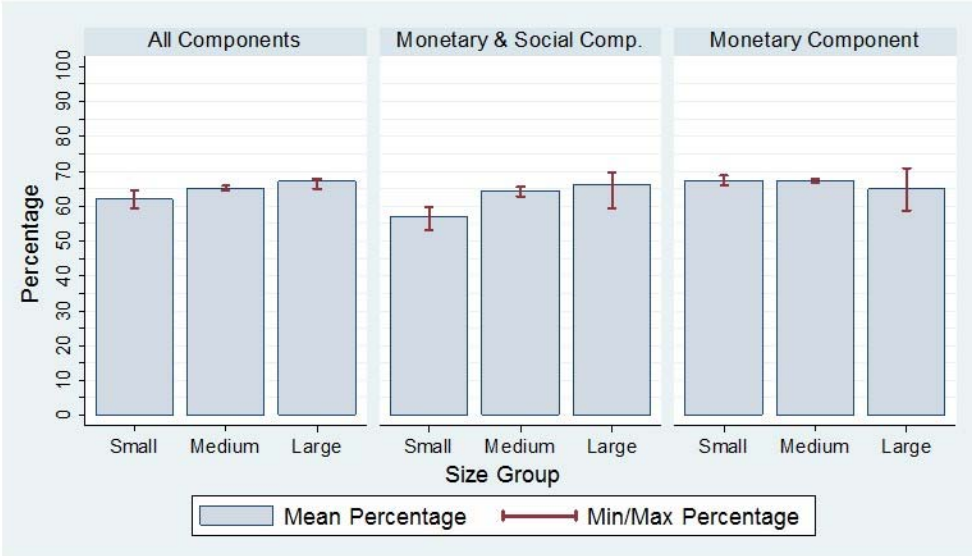

Figure 5-3 – Average Job Satisfaction at End-of-Year 260

Firstly, average job satisfaction in each firm was measured. Then, in each run and period, average was taken in each size group. The bars show averages over runs and the range plots show the ranges over runs at end-of-year 260.

The fact that it does not create firms with more than 850 employees (at year-end 260) might be the result of job variety and career opportunities dampening the social network effects.

This is also seen from the number of firms the three models generate (see Table 5-7). The MJS model generates the largest number of firms, because there are many small firms. The MSJS model generates the lowest number of firms, as it generates very large firms taking in many of the agents that could end up working in new firms. The FJS model is somewhere in-between.

It is illustrative to observe the differences in average job satisfaction for firms of different size depending on the structure of JS (see Figure 5-3). JS increases with firm size group only in the FJS model. While on average the behaviour is similar in the MSJS model, the range of average JS for

Table 5-6 – Results of Kolmogorov–Smirnov Test on the Equality of Firm Size Distribution

Full JS vs. Real Data, exact p-value	0.151
Monetary & Social Components vs. Real Data, exact p-value	0.125
Monetary Component vs. Real Data, exact p-value	0.037

Generated distributions are averaged over 20 runs. Test results show that the firm size distribution of the MJS model does not correspond to real data. Distributions generated by MSJS and FJS models are not statistically different from the real distribution, but the latter of the two is closer to the actual distribution than the former.

Table 5-7 – Number of Firms at End-of-Year 260

	Average	Minimum	Maximum
Full JS	237	210	260
Monetary & Social Components	219	197	246
Monetary Component	253	223	273

Statistics calculated over 20 runs for each composition type of job satisfaction.

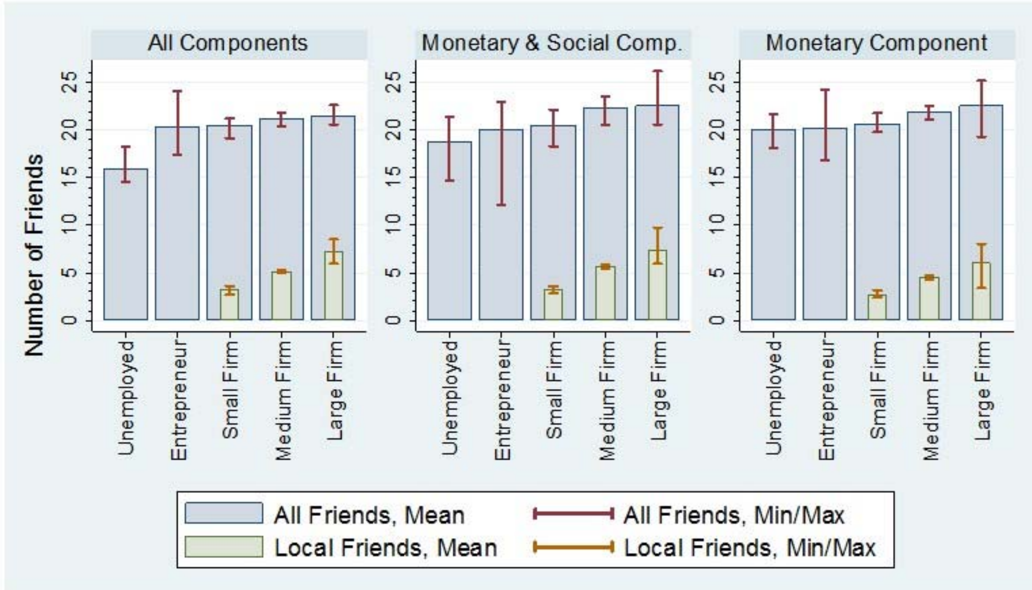

Figure 5-4 – Average Number of Friends at End-of-Year 260

Firstly, average number of friends in each firm and for unemployed and entrepreneurs was measured. Then, in each run and period, average was taken in each category. The bars show averages over runs and the range plots show the ranges over runs at end-of-year 260.

large firms is substantial and fully includes that of medium-sized firms. At the same time, average JS in small firms is 7 percentage points below that in medium-sized firms (as compared to the difference of 3 percentage points in the FJS model). In the MJS model, average job satisfaction is the same for small and medium-sized firms but on average *smaller* for large firms, although the range is wide, similarly to the MSJS model.

The average number of local friends increases with firm size, which is logical, as in a larger organisation, there are more possibilities to create more social ties (see Figure 5-4). Adding the social component to JS increases the number of local connections (by 1 in FJS and MSJS models, as compared to the MJS model), but not the overall number of connections. Adding job variety and career components to JS decreases the range of the overall number of connections and, except for the unemployed, who have only 16 friends on average in the FJS model, compared to 19–20 in the other two models, does not affect the average number of friends.

5.1.4.2 Labour-Market Status of Persons

The model generates the Beveridge curve, as shown in Figure 5-5. In all three models, it is non-linear and downward sloping, as expected. The figure also shows that the FJS model generates a more compact distribution of unfilled vacancies at year-ends, while the range of

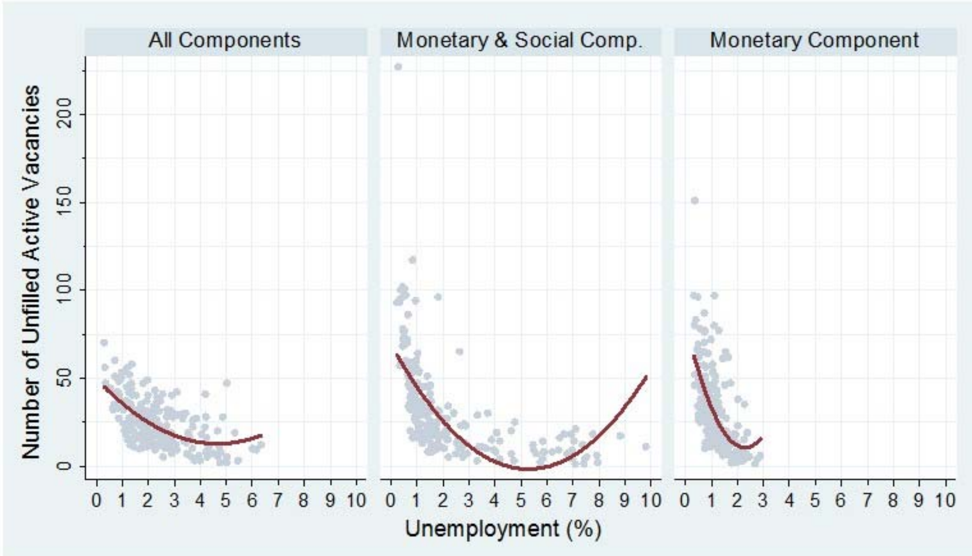

Figure 5-5 – Beveridge Curve at End of Years 251–260

Each point is a combination of unemployment and number of vacancies in a particular run of the model. Lines are predicted values of quadratic regression.

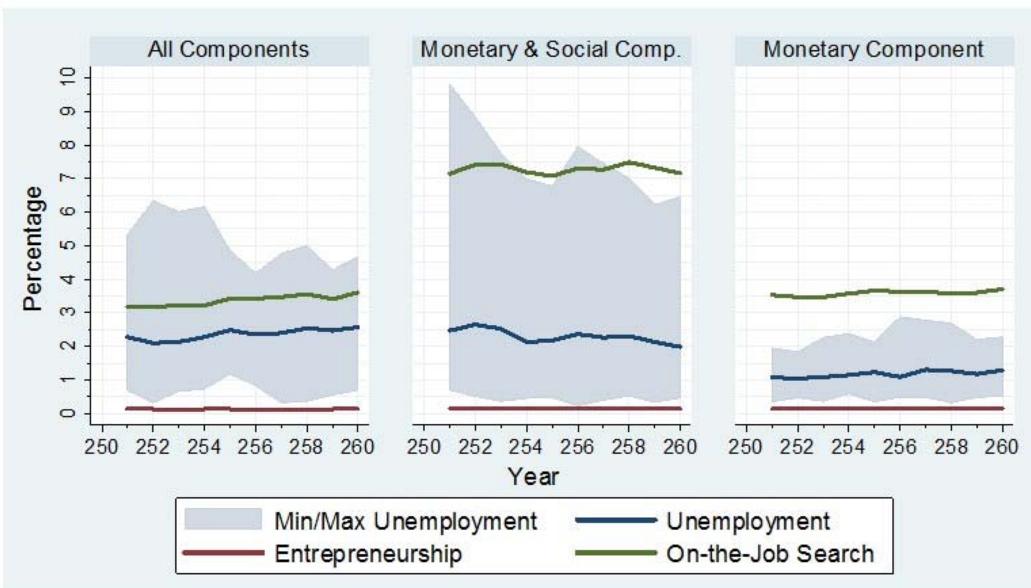

Figure 5-6 – Unemployment, Entrepreneurship and Agents Involved in On-the-Job Search for at Least One Month as Percentage of All Agents

The figure shows averages over runs at year-ends. On-the-job search data exclude persons who started on-the-job search and moved to the new job within the same month.

unemployment distribution at year-ends is larger than for the MJS model but smaller than for the MSJS model.

While adding the social component in JS increases both average unemployment and its range, average unemployment levels in the MSJS and FJS models are similar (see Figure 5-6). This,

Table 5-8 – Statistics on Unemployment Duration

	Prob. of Various Durations				Mean	Median
	3 months	6 months	9 months	1 year	(months)	
FJS	50%	73%	85%	92%	5.5	4
MSJS	42%	65%	78%	86%	7.3	4
MJS	46%	68%	82%	90%	6.2	4

Means over runs reported.

Table 5-9 – Statistics on On-the-Job Search Duration

	Prob. of Various Durations				Mean	Median
	3 months	6 months	9 months	1 year	(months)	
FJS	51%	72%	83%	88%	6.2	3
MSJS	41%	60%	70%	77%	10.4	5
MJS	44%	63%	75%	81%	8.4	4

Means over runs reported.

though, cannot be said about unemployment duration. On average, persons are unemployed for 5.5 months in the FJS model and 7.3 months in the MSJS model (see Table 5-8). While the level of unemployment is similar in the FJS and MSJS models, finding a new job is, thus, more difficult in the latter model.

The extent of on-the-job search continuing at least one month is very similar in the MJS and FJS models, but it doubles in the MSJS model (see Figure 5-6). Table 5-9 also shows that the MSJS model generates the highest on-the-job search duration and the FJS model – the lowest. Half of respondents are able to find a new job within the first 3 months of on-the-job search when JS includes the job variety and career components, compared with 40 per cent if it is based only on monetary and social components. Similarly, median duration is half a year in the FJS model, compared to 10 months in the MSJS model and 8 months in the MJS model.

In all three models, around 80 per cent of jobs in firms of any size type are non-manual jobs, although the probability of publishing a manual vacancy was 30 per cent. This higher popularity of non-manual jobs should be related to non-manual workers receiving much higher wages than manual workers, which will now be discussed in more detail.

5.1.4.3 Wage Distribution

The wage distribution has very long right tail, so it is shown on logarithmic scale in Figure 5-7. During the evolution of the economy, wages strongly surpassed the initial minimum wage of 100, and no agent received less than 8000 by the end of year 260 (the cumulative annual growth rate in actual minimum wage is, thus, $\sqrt[260]{8000/100} - 1 = 1.7\%$).

There are in fact two distributions on each plot of Figure 5-7: for manual jobs (left) and for non-manual jobs (right). Because wages rose with experience (parameter q in Table 5-5) faster

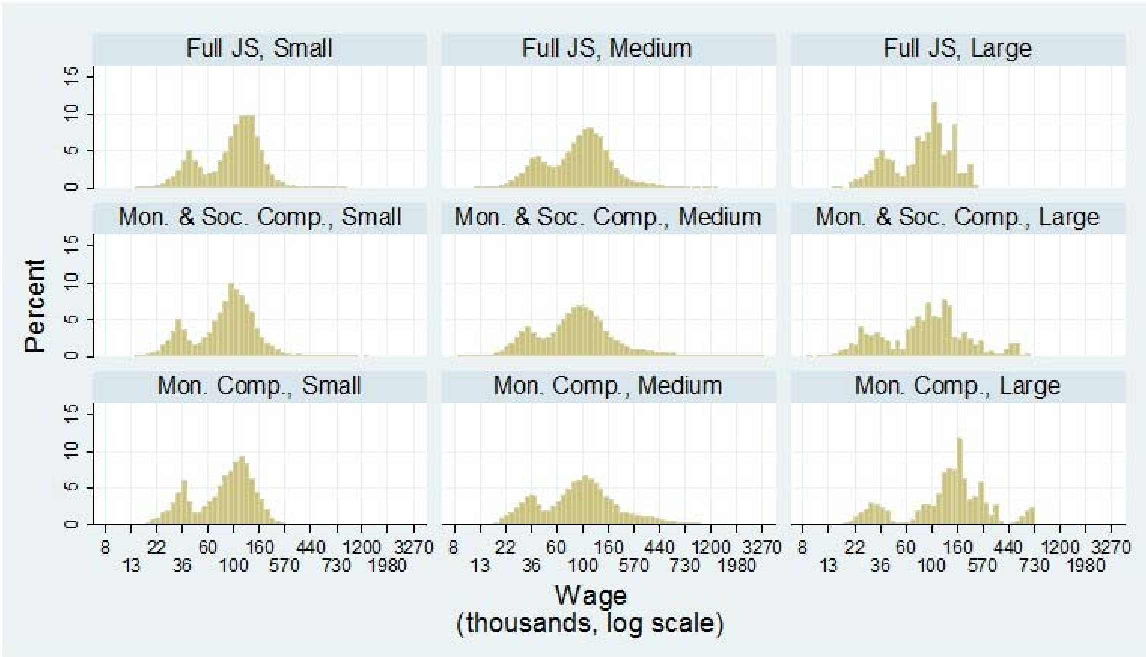

Figure 5-7 – Histograms of Wages at End-of-Year 260

Histograms created on pooled data from 20 runs.

for the latter than for the former, it is no surprise that non-manual workers generally have higher wages. The two distributions, however, intersect (except for large firms in the MJS model), so manual workers are sometimes paid more than non-manual workers.

The distributions for manual and non-manual workers are close to normal distributions when log-wages are used, meaning that all models generate lognormal distributions of wages.

Median wages in the firms of same size group sometimes substantially differ across the models (see Table 5-10). The FJS model generates the highest wages for manual jobs across all firm size groups and for non-manual jobs in small firms. At the same time, it has nearly the same level of wages as the MJS model for non-manual jobs in medium-sized firms and the lowest wages for those jobs in large firms.

The table also shows that in the FJS model, wages in both types of jobs *decrease* with firm size group. This is also observed for manual jobs in the other two models, but *not* for non-manual

Table 5-10 – Median Wages by Job Type and Firm Size at End-of-Year 260

	Manual Jobs			Non-Manual Jobs		
	Small	**Medium**	**Large**	**Small**	**Medium**	**Large**
FJS	40	38	35	119	112	99
MSJS	33	32	28	98	102	120
MJS	34	33	32	107	115	188

Wages reported in thousands. Firstly, median wages in each run were computed. The table presents medians of these values over runs.

jobs. Thus, medium-sized and large firms in the FJS model can attract non-manual workers without proposing them higher wages. In MSJS and MJS models, these firms compete more intensively by increasing their wage proposals.

5.1.5 Discussion and Conclusions

In this section, I proposed to use job satisfaction (JS) in agent-based models as a unified mechanism for guiding individual agent's behaviour in the labour market. In my illustrative model, JS affects agents' decisions on which vacancies to apply for, which of them to select in case of receiving several acknowledgements from firms and whether to quit the current job. I compared the performance of three models of JS: (1) MJS, where JS depends only on the monetary component (i.e., wages), (2) MSJS, where JS depends on monetary and social (i.e., number of friends) components, and (3) FJS, where JS depends not only on monetary and social components, but also on job variety and career opportunities. The performance is compared for firm-level statistics, labour-market status of individual agents and wage distribution.

The FJS model creates a more realistic firm size distribution, but creates no firms with more than 1000 employees. The reason of that is smaller importance of social network effects after adding job variety and career opportunities to JS. Recall that because the model assumes that all components are equally important, adding two more components decreased the weights of the social component from $1/2$ in the MSJS model to $1/4$ in the FJS model. This is exactly why the FJS model does not generate mutli-thousand-employee firms.

Only in the FJS model, average JS increases with firm size. In FJS and MSJS models, there are more local friends than in the MJS model, which is attributed to social networks affecting JS. Unemployed agents have fewer friends in the FJS model than in the other models, but the number of friends of entrepreneurs and workers of firms of different size types are similar across models.

All three models generate the Beveridge curve, but the distribution of vacancies is more compact in the FJS model. Average unemployment is higher in FJS and MSJS models than in the MJS model. The share of agents involved in on-the-job search is higher in the MSJS model than in the other two. The FJS model has the lowest duration of on-the-job search and unemployment, meaning that it is easier for individual agents to find a new job in this model than in the other models.

Wage distributions for manual and non-manual workers are close to lognormal. Wages are generally higher for non-manual workers, but the two distributions intersect. Cumulative annual growth rate of the actual minimum wage during the 260 years of evolution of the model was 1.7 per cent (from 100 to around 8000). The FJS model generates the highest wages for manual workers and for non-manual workers in small firms, but the lowest wages for non-manual workers in large firms. Wages decrease with firm size for manual jobs in all models, but for non-manual jobs only in the FJS model.

Three limitations of the model should be addressed in the future. Firstly, it is the absence of weights on JS components, so that currently, all components are equally important to agents. It is very interesting to see how the relative performance of FJS and MSJS models changes when the distribution of weights across agents is introduced. Secondly, there are only two types of jobs representing broad groups of occupations. Possibilities of operating with narrower occupations should be considered. Thirdly, firm behaviour with respect to publishing vacancies is simplistic. It could be enhanced by creating different types of firms pursuing different strategies and adapting their strategies based on their success. The model could also be integrated into a larger model with production and consumption sides.

5.2 Field of Study Choice and Migration: Effects of Social Networks

Causes of emigration and factors influencing the choice of migration target have long been under study. Recent empirical evidence shows that, in addition to labour market effects such as differences in wages and unemployment levels, social networks play an important role in decisions on migration. In particular, it was shown that the size of social capital in a country is positively related to emigration intentions to that country (Haug, 2008; Mayda, 2010; Pedersen, Pytlikova, & Smith, 2008), while having strong social networks in the current place of residence decreases the probability of emigration (Michaelides, 2011). The most frequently used explanation is that future emigrants obtain useful information about the target country through social networks, as well as use their contacts to enhance their adaptation speed after migrating.

The current section studies the dynamics of migration in a set of artificial countries[43] using agent-based simulations. Decisions on migration are made probabilistically after a period of being unable to find a job in the local labour market. The target country for emigration is chosen by considering an indicator consisting of economic (wage) and social (number of friends) factors. In empirical macro-level research, the share of emigrants from own country is frequently used to

[43] Countries are used here abstractly. Hence, the same model could be used to study regional migration.

measure social network size driving the intention to choose a particular migration target. I instead focus on the use of individual social networks.

Attention has recently been brought to the position of tertiary graduates in the labour market based on the field of study they graduated. The main decision here is to choose the "right" field of study; if they fail to do it, graduates will face problems to find a suitable job, which leads to overeducation or other kinds of mismatch. In models of field of study effects in the labour market, authors frequently neglect the *quality of match* between the characteristics of job applicant and requirements of job. I incorporate it into my model by making (1) the probabilities of finding and losing a job and (2) wage rates conditional on it.

Moreover, I conceptually distinguish (1) the factors agents take into account when choosing the field of study from (2) the factors that determine how successful they will be in the chosen profession after graduation. Agents choose a field of study based on subjective information available from their friends, not using any population-wide statistic.

Thus, migration flows among countries in my model are studied in the context of the quality of match between the agent and the profession where it is searching for a job, taking into account potentially adverse information obtained from the social network that biases one's choice of the field of study.

The section is structured as follows. The next subsection specifies the agent-based model. Section 5.2.2 presents parametrisation and Section 5.2.3 discusses results. The last subsection concludes.

5.2.1 Model Setup

Agents representing individuals enter the model after graduating secondary school. They decide whether to continue studies at the university, start working immediately or emigrate from the country. I assume that there is a single university in the country (or, equivalently, that the quality of education in a given field of study is the same in all universities), which provides access to higher education in $\Phi \in \mathbb{N}_+$ fields of study, where $\mathbb{N}_+$ is the set of positive natural numbers. There is also the same number Φ of professions in the skilled labour market. Each field of study $\varphi \in \{1, \dots, \Phi\}$ prepares students for the corresponding skilled position type $\psi = \varphi$.

5.2.1.1 Agents' Characteristics

Agents are described by two sets of characteristics: $\mathcal{F}$ and $\mathcal{H}$. ***Set*** $\boldsymbol{\mathcal{F}}$ contains characteristics that are important for success in particular fields of study and, subsequently, in the corresponding professions in the labour market. The importance of each $\mathcal{F}$-characteristic for fields of study is measured in three levels: Low, Medium or High. For instance, mathematical skills might be of low importance to a linguist, of medium importance to a programmer and of high importance to a physicist. These mappings from fields of study to the levels of $\mathcal{F}$-characteristics are stored in matrix $M = (m_{\oint\varphi})$, where $\oint \in \mathcal{F}$, $\varphi = \overline{1, \Phi}$ and $m_{\oint\varphi} \in \{L, M, H\}$. Note that it is quite realistic to assume that the requirements summarised in matrix M are stable over time (of course, provided that $\mathcal{F}$-characteristics and fields of study are sufficiently general).

Set $\mathcal{H}$ contains characteristics that are irrelevant for one's success in any profession but, nevertheless, influence the choice of field of study. For instance, the importance of friends' opinions does not influence the success in studies and in the labour market; it may, however, distort the choice of field of study for agents that are easily influenced by their friends.

For education policy-makers, it is important to know the distributions of both $\mathcal{F}$- and $\mathcal{H}$-characteristics. The former show which professions would be demanded in the ideal case. This could be used in changing the secondary (and, perhaps, primary) education policy in accordance with government's vision of future economic development so that future students have necessary abilities, knowledge and skills to be able to study in the priority fields. A simple example could be the decision to increase the number of hours of mathematics lessons at secondary schools with the aim to enhance mathematical skills if the vision is that the share of engineering professions should increase.

The latter show why field-of-study decisions are distorted. These factors are important because they may prevent education policy from reaching its aim. For instance, if a large share of secondary-school graduates is extremely interested in prestigious positions, they may choose to study economics and business over engineering if the former are considered much more prestigious than the latter in society. Increasing scholarships in engineering in this case might fail to provide enough incentives for students to change their choice. It would instead be appropriate to run marketing campaigns emphasising the prestige of engineering professions.

One can, thus, also view $\mathcal{F}$-characteristics as objective indicators of the fields/professions where it would be best for a given agent to study/work, while $\mathcal{H}$-characteristics – as subjective factors the agent takes into account when selecting the field of study.

5.2.1.2 Labour Market Dynamics

In the real world, one can make two observations. Firstly, the fact that individuals get their higher education in a particular field of study does not prevent them to work in a profession that

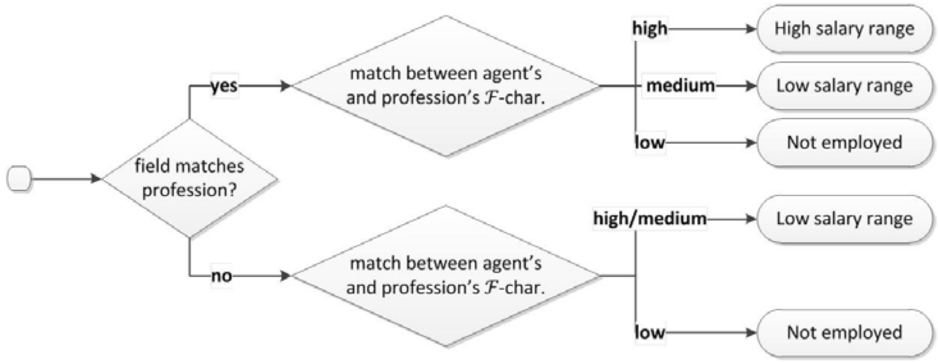

Figure 5-8 – Determination of Employment Possibilities and Salary in the Skilled Labour Market

is weakly related or unrelated to that field. Secondly, employers generally require candidates for a technical or high position inside the profession to hold a tertiary degree in the respective field. To incorporate these two features in my model, I introduce the scheme shown in Figure 5-8. Thus, while $\mathcal{F}$-characteristics do not play any role in the field-of-study choice, they are an important factor affecting graduates' wages and employment possibilities in the profession.

Recall that field of study $\varphi \in \{1, \dots, \Phi\}$ matches profession $\psi \in \{1, \dots, \Phi\}$ if $\varphi = \psi$.

The quality of match $q_{i\psi} \in \{H, M, L\}$ between $\mathcal{F}$-characteristics of agent i and profession ψ is defined as follows. Note that if all $\mathcal{F}$-characteristics are positive in the sense that more is better, the fact that a particular characteristic $\mathscr{f}$ is considerably lower for the agent than for the profession ($\mathscr{f}_i = L$ but $\mathscr{f}_\psi = H$) means that the agent will have significant difficulties in this profession. When the characteristic is moderately lower for the agent, however, the problem is less pronounced, especially if the difference is in a small number of characteristics. Formally, defining the function $n: \mathcal{F} \to \mathbb{N}$ as $n(L) = 0, n(M) = 1, n(H) = 2$, one can define the distance

$$d(\mathscr{f}_i, \mathscr{f}_\psi) = n(\mathscr{f}_i) - n(\mathscr{f}_\psi).$$

Thus, if characteristic $\mathscr{f}$ is lower (higher) for the agent than for the profession, the respective distance is negative (positive). I can now define the ***quality of match*** as

$$q_{i\psi} = \begin{cases} L, & \left(\exists \mathscr{f} \in \mathcal{F}: d(\mathscr{f}_i, \mathscr{f}_\psi) = -2\right) \vee \left(\left|\left\{\mathscr{f} \in \mathcal{F} \middle| d(\mathscr{f}_i, \mathscr{f}_\psi) = -1\right\}\right| > q^L\right) \\ M, & \left(\forall \mathscr{f} \in \mathcal{F}\ d(\mathscr{f}_i, \mathscr{f}_\psi) > -2\right) \wedge \left(q^H \leq \left|\left\{\mathscr{f} \in \mathcal{F} \middle| d(\mathscr{f}_i, \mathscr{f}_\psi) = -1\right\}\right| \leq q^L\right) \\ H, & \left(\forall \mathscr{f} \in \mathcal{F}\ d(\mathscr{f}_i, \mathscr{f}_\psi) > -2\right) \wedge \left(\left|\left\{\mathscr{f} \in \mathcal{F} \middle| d(\mathscr{f}_i, \mathscr{f}_\psi) = -1\right\}\right| < q^H\right), \end{cases}$$

where $|\{\cdot\}|$ denotes the cardinality of the respective set.

In the beginning of every period, agents that did not study at the university either keep their job with probability p_u or lose it with probability $(1 - p_u)$. In case of losing the job, unskilled agents try to find a new job in the same period. They succeed with probability p_u or remain

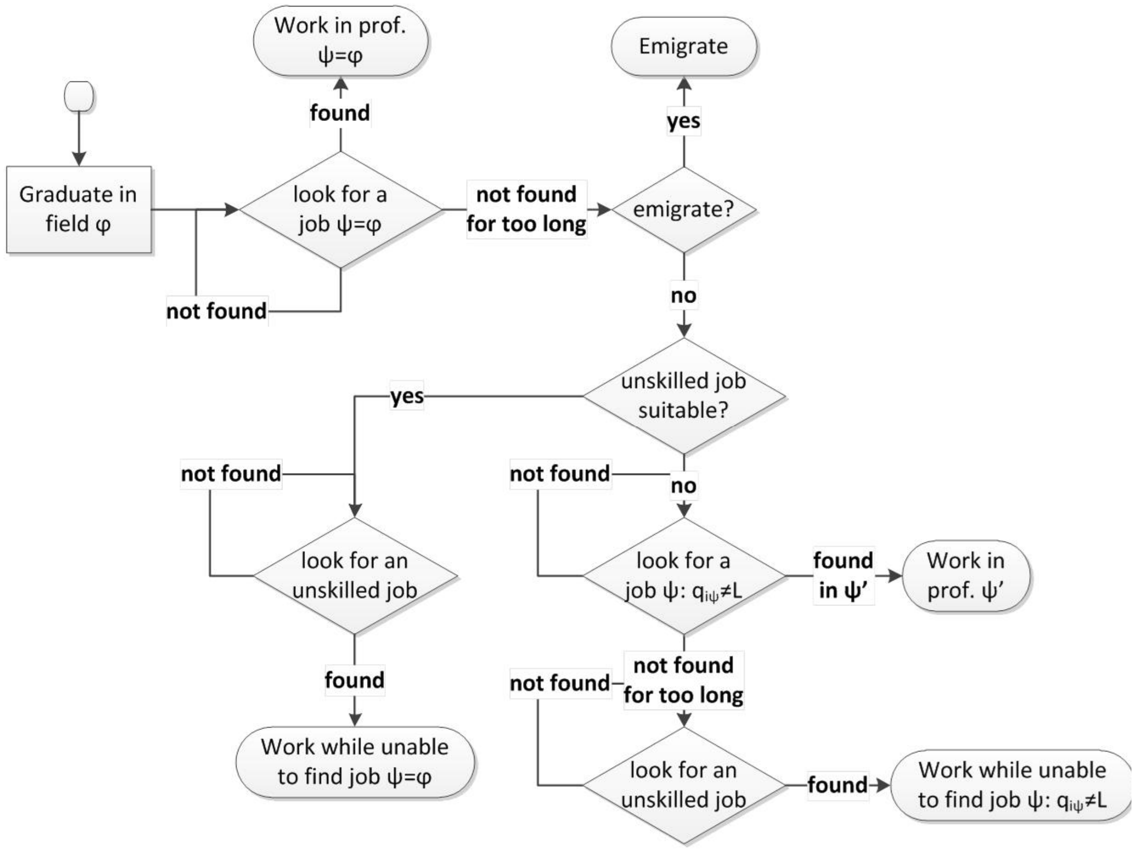

Figure 5-9 – The Process of Finding a Job for University Graduates

unemployed until the next period. If they are unemployed for more than τ periods, agents emigrate[44] with probability p_e^u.

The labour market dynamics of university graduates are more complex. The possible paths from graduation to finding a job are displayed in Figure 5-9. All choices are probabilistic. Agents find a job in profession ψ with probability $p_{\psi M}$ if the quality of match $q_{i\psi}$ is medium and $p_{\psi H} > p_{\psi M}$ if it is high (having a diploma in field $\varphi = \psi$ does not change the probability, but recall that it influences the potential wage rate). Agents lose their jobs with probability $(1 - p_{\psi M})$ or $(1 - p_{\psi H})$, depending on the quality of match. "Too long" means longer than τ periods. Graduate agents emigrate with probability p_e^s. Unskilled job is suitable with probability p_u^s. A skilled agent finds/loses an unskilled job with the same probabilities as an unskilled agent.

The possible paths after losing one's job are nearly identical to those after graduation, with the only difference in that agents firstly try to find a job in *the last profession where they worked* as opposed to *the profession that corresponds to their field of study*, as in Figure 5-9. As in the

[44] Note that regional migration inside a country is not modelled; it is, thus, implicitly assumed that agents consider employment opportunities in the whole country, and not only in their current region of residence.

Algorithm 5-2 – Choosing the Field of Study

```
x ← random(0, 1)
if (x < ρ)
  for each field φ
    for each ℋ-characteristic ℏ
      m[ℏ][φ] ← measure ℏ in φ
  for each ℋ-characteristic ℏ
    l[ℏ] ← all fields φ ∈ arg max_φ m[ℏ][φ]
  for each field φ
    score[φ] ← number of times φ found in l
  maxscorelist ← all fields φ:(score[φ] = max_φ score[φ])
  choose random field from maxscorelist
else
  do not enter the university
```

case of unskilled agents, graduates lose jobs in the beginning of the period and can find another job already in the same period.

Denote the current country of residence as c_r, the mean wage rate of country c as $\overline{w}_c$, and the number of friends of agent i living in country c as f_{ic}. Agent i emigrates to the country with the maximal sum of relative mean wage rate and relative number of friends, combining economic and social criteria:

$$\max_{c \neq c_r}\left(\overline{w}_c/\overline{w}_{c_r} + f_{ic}/\max_{d \neq c_r} f_{id}\right).$$

5.2.1.3 Choice of Field of Study

The actual choice of field of study is made *exclusively* based on $\mathcal{H}$-characteristics. In other words, I assume that right after graduating secondary school, agents are either very optimistic regarding their capabilities to succeed in the profession that is the best, according to their subjective opinions, or are simply unable to assess their $\mathcal{F}$-characteristics and those of different fields and, consequently, unable to identify the field for which they are objectively best suited.

Not everyone, however, decides to continue studies at the university. I assume that this decision is unrelated to $\mathcal{F}$- or $\mathcal{H}$-characteristics. An agent decides to enter the university with probability ρ.

The overall algorithm for choosing the field of study is presented in Algorithm 5-2.

5.2.1.4 Social Networks

Right after birth, agents born in the same period randomly build a social network using the Duplication model with parameter $p \in (0,1)$ (Chung & Lu, 2006, Ch. 4). They also become friends of approximately[45] m agents who graduated the university and are currently employed, also in accordance with the Duplication model with the same parameter p. I use the Duplication model for creating a social network because it produces a scale-free network, with many low-degree vertices and a few high-degree vertices, which is common to many real-world social networks (see Section 2.2.2).

5.2.2 Parametrisation

In this study, I assume that the joint distribution of $\mathcal{F}$-characteristics in the population and the mapping matrix M are given. I study the effects of $\mathcal{H}$-characteristics and, in particular, take social networking characteristics as primary factors of interest, as (1) there is evidence that they play an important role in decision-making, (2) the possibilities to express and discuss one's opinions have been increasing tremendously in the last years and (3) these are very difficult to study analytically.

The matrix M, which maps fields of study/professions to $\mathcal{F}$-characteristics, is given in Table 5-11. I chose to model $\Phi = 4$ fields of study/professions and $|\mathcal{F}| = 6$ characteristics. The mappings are chosen so that Fields 1 and 2 have a considerable overlap, Field 3 has very low requirements of success in the labour market, while Field 4 emphasises different characteristics in agents, as compared to the other three fields.

The distribution of $\mathcal{F}$-characteristics in the population of secondary school graduates are presented in Table 5-12. Each characteristic is of medium level with probability 0.6. Characteristics differ on the probabilities of high and low levels. For instance, characteristics 5 and 6, which are required for success in Profession 4, are high quite rarely (in 10% of cases). On the contrary, characteristics 1 and 2, which are required in Professions 1 and 2 (and partly in Profession 3), are high with probability 0.3. The other two characteristics have equal probabilities to be high or low.

Wage intervals (see Table 5-13) are set so that employees in skilled positions receive much more than those in unskilled positions. Inside professions, intervals for agents with medium match and with high match do not overlap. Employees in Professions 1 and 2 receive wages drawn from the same distributions. Agents working in Profession 3 receive significantly lower wages, which is logical due to very low requirements for success in this profession. On the contrary, high-quality-matched agents in Profession 4 receive a much higher wage than in other

[45] Technically, the two duplication steps are executed until the number of non-student friends is at least $\min\{LF/2, m\}$, where LF is current labour force size (i.e., the number of non-student agents).

Table 5-11 – $\mathcal{F}$-characteristics and Their Mappings to the Fields of Study Used in the Model

	Char. 1	Char. 2	Char. 3	Char. 4	Char. 5	Char. 6
Field/Profession 1	H	H	L	H	L	L
Field/Profession 2	H	H	H	L	L	L
Field/Profession 3	H	L	L	L	L	L
Field/Profession 4	L	L	M	M	H	H

Table 5-12 – The Distribution of the Levels of $\mathcal{F}$-characteristics in the Population

	Char. 1	Char. 2	Char. 3	Char. 4	Char. 5	Char. 6
Low	0.1	0.1	0.2	0.2	0.3	0.3
Medium	0.6	0.6	0.6	0.6	0.6	0.6
High	0.3	0.3	0.2	0.2	0.1	0.1

Table 5-13 – Wage Intervals of Professions

		Unskilled	Profession 1	Profession 2	Profession 3	Profession 4
Quality of Match	**Medium**	$[200, 300]$	$[500, 800]$	$[500, 800]$	$[500, 600]$	$[500, 800]$
	High		$[800, 2000]$	$[800, 2000]$	$[600, 800]$	$[1500, 3000]$

professions, as very few agents have the characteristics necessary for such quality of match in this profession. Wage distributions are lognormal inside intervals.

The $\mathcal{H}$-characteristics that are used in this section are listed in Table 5-14. There are several comments to be made to this list. Firstly, agents in the model have no access to aggregate information but can gather statistics over their friends. For instance, agents do not know the popularity of a field of study in the population; they, thus, make an approximation by measuring it among their friends. Secondly, agents are quite optimistic about their wage prospects and compute the average over the top 50 per cent of the wages of their friends.

Values of the other parameters are shown in Table 5-15. I now briefly comment on these.

The number of periods between two consecutive graduations is set to four. Thus, if an inflow of secondary school graduates happens once in a year, as in the real world, then four periods snap one year and one period represents one quarter. Studies continue for 12 periods (or three years) in all fields, typical for bachelor studies in Europe, which last for 3–4 years.

Probabilities of finding/keeping job were set using Eurostat data. In particular, the probability of finding/keeping an unskilled job was set to 0.70, the average employment level of the secondary-educated in 2007–2009. According to the same data, this probability was near 0.85 for the tertiary-educated. However, this is the average figure over those who fit the profession perfectly (high quality of match) and well enough (medium quality of match). To distinguish between the two, I set the probability for the former to 0.95 and for the latter to 0.80. Probability of starting to look for an unskilled job instead of looking for a skilled job in a different profession

Table 5-14 – $\mathcal{H}$-characteristics Used in the Model

Popularity of the Field among Friends
Share of friends who graduated the field among all graduate friends
Labour Market Success of the Friends Who Are Graduates in the Field
Average wage, calculated over the top half of employed friends
Current employment rate
Labour Market Success of the Friends Working in the Profession Corresponding to the Field
Average wage, calculated over the top half
Current employment rate

Table 5-15 – Other Parameters

Parameter	Notation	Value
Timing		
Length of simulation, periods		400
Number of periods between graduations		4
Population & Social Networks		
Max age, periods		60
Inflow of new graduates, number of agents		500
Max number of friends from the same graduation period		100
Approx. number of mature friends of new graduates	m	50
Duplication model parameter	p	0.45
Education		
Prob. of going to university	ρ	0.9
Length of studies, same for all fields φ, periods		12
Labour Market Dynamics		
"Too long," periods	τ	2
Non-exact lower bound of the number of $\mathcal{F}$-characteristics where the distance is -1 for having a low quality of match	q^L	4
Non-exact upper bound of the number of $\mathcal{F}$-characteristics where the distance is -1 for having a high quality of match	q^H	1
Prob. of finding/keeping an unskilled job	p_u	0.70
Prob. of starting to look for an unskilled job in parallel to looking for a job in the priority profession	p_u^s	0.15
Prob. of finding/keeping a medium-match skilled job, same for all professions ψ	$p_{\psi M}$	0.80
Prob. of finding/keeping a high-match skilled job, same for all professions ψ	$p_{\psi H}$	0.95
Migration		
Prob. of emigration, skilled agents	p_e^s	0.05
Prob. of emigration, unskilled agents	p_e^u	0.02

is set to 0.15, the highest average European overeducation rate in 2002–2011, according to the European Social Survey.

Migration probabilities for skilled and unskilled labour force are approximated from world average emigration rates of tertiary- and secondary-educated individuals, respectively (Docquier & Marfouk, 2005, p. 18 (Table 2), data from 1990 and 2000).

I simulate agent behaviour in four identical countries.

5.2.3 Results

The simulation was written in *Repast Simphony*. I perform the analysis on periods 100–400, as period 100 is sufficiently far from initial conditions. The simulation was run 50 times.

Figure 5-10 presents a typical pattern of the dynamics of demand for different fields of study. It shows that the demand for Fields 1 and 2 are approximately equal, while the demand for Fields 3 and 4 is significantly lower. This can be explained by recalling two facts. Firstly, while it is very easy to get a job in Profession 3 with wage in the upper range, wages there are much lower than in the other professions. Apparently, while secondary-school graduates have no access to the statistical information on wages in the whole population, they are able to extract it from their social network. Secondly, while the upper wage range in Profession 4 is far higher than in any other profession, it is much harder to qualify for high-quality match with this profession than, for instance, with Professions 1 and 2. Consequently, it may be that most agents employed in Profession 4 are actually of a medium-quality match and, thus, their wages are in the lower range.

I, thus, look on the statistics of employees with high-quality match (HQM) (see Table 5-16). My hypothesis that there are much fewer HQM agents in Profession 4 than in all other professions is supported. However, it is not only the quality of match that determines the wage range – the field–profession match also plays an important role, but the dynamics of the share of field–profession matches among HQM agents employed in the profession under consideration are very volatile: its median fluctuates between 40 and 90 per cent. The reason for this is a very

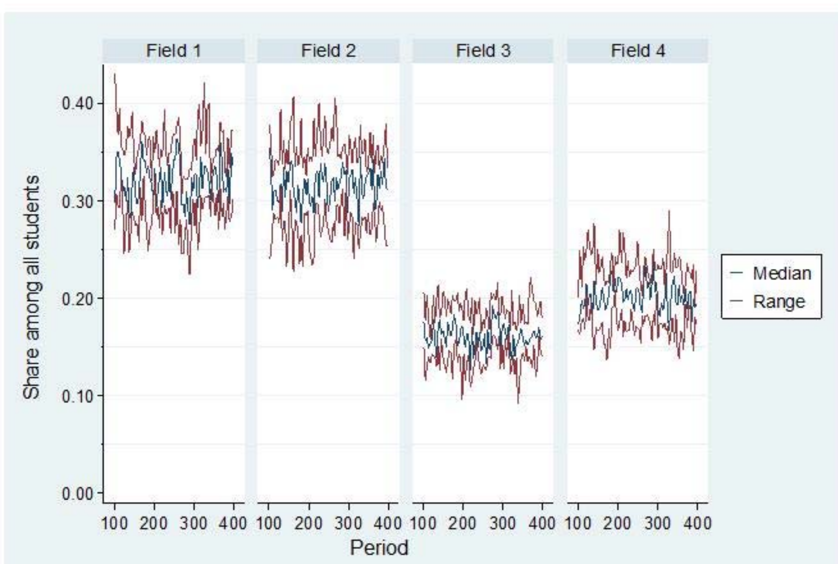

Figure 5-10 – Demand for Fields of Study in Country 1

Table 5-16 – Statistics of Employees with High-Quality Match with Their Current Profession

	Prof 1	Prof 2	Prof 3	Prof 4
Share of HQM among all employees	$3.0\% \pm 0.6\%$	$3.0\% \pm 0.6\%$	$35.4\% \pm 1.5\%$	$1.2\% \pm 0.5\%$
Number of HQM	34 ± 7	35 ± 7	455 ± 28	9 ± 4
Share of field–profession matches among HQM	$85\% \pm 7\%$	$85\% \pm 7\%$	$56\% \pm 4\%$	dynamics is unstable

Results reported within professions. HQM stands for “high-quality matched.”

Table 5-17 – Overeducation Rates by Field of Study

Field 1	Field 2	Field 3	Field 4
$8.3\% \pm 1.0\%$	$8.3\% \pm 1.0\%$	$4.8\% \pm 1.0\%$	$10.5\% \pm 1.0\%$

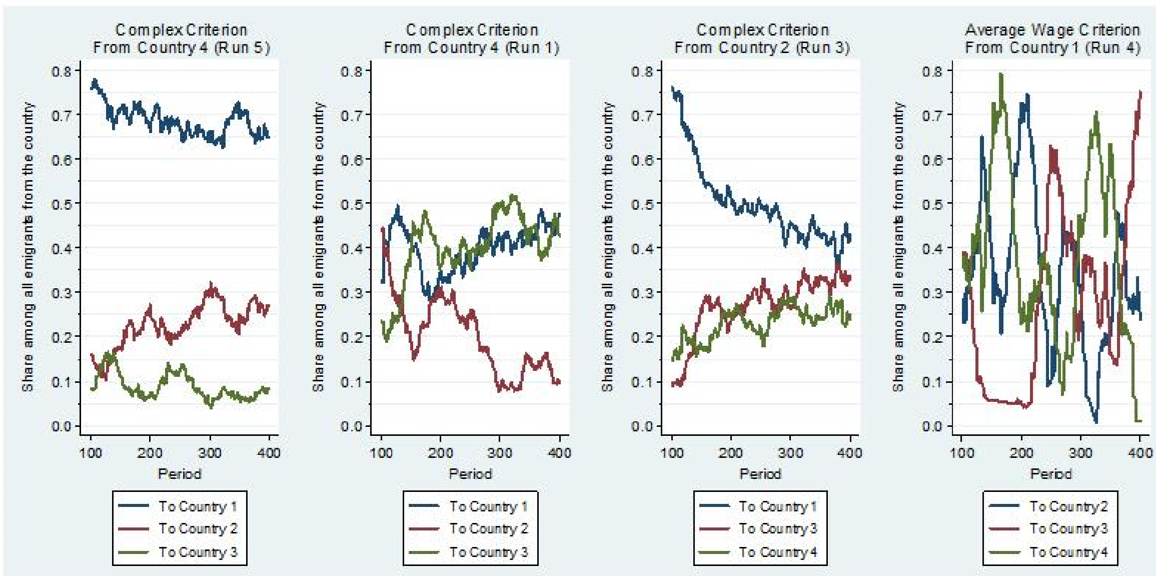

Figure 5-11 – Selected Examples of Emigration Patterns

low absolute number of HQM agents in Profession 4 – only around ten, so that small absolute changes significantly affect the share indicator.

Another interesting finding from Table 5-16 is that the share of field–profession matches among HQM agents in Profession 3 is much lower than in Professions 1 and 2. This is because agents that are well-suited for Profession 3 graduate the university in other fields and, being unable to find a job in the profession that matches their field, finally enter Profession 3.

Of course, another reason of choosing Field 4 less frequently might be that it leads to a higher risk of overeducation (hence, to lower average wages of this field’s graduates). This is indeed the case, as reported in Table 5-17. Note also that Field 3 has the lowest probability of having to move to an unqualified position after graduating it, but this does not increase its popularity.

Finally, I consider effects of field and profession popularity among friends on the choice of field of study. For this purpose, I run a multinomial logistic regression, the dependent variable

Table 5-18 – Marginal Effects of Social Network Characteristics on Choosing Fields of Study after Multinomial Logistic Regression

	Field 1	Field 2	Field 3	Field 4
Decile of friends who graduated the field	0.140^{***}	0.155^{***}	-0.191^{***}	-0.104^{***}
Decile of friends working in the profession corresponding to the field	-0.002^{**}	-0.008^{***}	0.160^{***}	-0.167^{***}

*** and ** mark significance at 1% and 5% level, respectively. Model pseudo R-squared is 0.3383. Cumulative data from 5 runs, periods 100–400.

Table 5-19 – Share of Emigrants from Country by Education

Unskilled	Field 1	Field 2	Field 3	Field 4
$3.3\% \pm 2.3\%$	$31.9\% \pm 7.0\%$	$31.1\% \pm 7.0\%$	$5.9\% \pm 3.5\%$	$28.0\% \pm 6.1\%$

being four fields of study (see Table 5-18). Firstly, note a quite high pseudo R-squared of 0.34 coming exclusively from the popularity of the field or the corresponding profession (these being the only regressors, I did not take into account friends' labour market characteristics). Field popularity among friends increases the probability of choosing Fields 1 and 2 and decreases that for Fields 3 and 4. The popularity of the corresponding profession among friends increases the likelihood of choosing Field 3, decreases that of choosing Field 4, but has a quite minor influence on Fields 1 and 2.

Figure 5-11 presents examples of emigration patterns encountered in the simulation. In the vast majority of cases, there is a single leading target country for emigrants, which attracts around 60 to 80 per cent of all emigrants (see the left panel). Rarely there are two target countries, each attracting 40 to 45 per cent of migrants (the middle-left panel) or three target countries with similar chances to attract migrants from the given home country (the middle-right panel). The right panel shows the typical pattern encountered when the target country is selected only based on its relative average wage. Comparing these graphs, one can conclude that the use of the complex criterion, when social network aspects influence decision-making, brings more structure to emigration patterns.

Table 5-19 shows who tends to emigrate from the native land. While the chance of finding an unskilled job are much lower than that of finding a skilled job, the unskilled are the least likely to emigrate, due to lower probability of making the emigration decision. Graduates in Field 3 are also quite unlikely to emigrate, this time because they have lower requirements for being admitted to work and, hence, higher chances to find one. Unexpectedly, graduates in Field 4 are less likely to emigrate than graduates in Field 1 or Field 2 (the differences in means are statistically significant).

5.2.4 Conclusions

In this section, I study a model that allows to assess the macro-dynamics emerging exclusively through (1) interactions within social networks and (2) the quality of match between the agent's characteristics and those required by the profession.

Overall, my model gives plausible results about education and labour markets. This, however, does not mean that all results are as expected. For instance, surprising results were found among the effects of popularity factors on the choice of field of study, where effect directions differ by field, or in emigration, where fewer emigrants are educated in a more difficult field of study than expected.

Statistics of agent behaviour in choosing field of study and succeeding in the labour market were very similar in all four countries, which was expected. Nevertheless, it was shown that social networks play a significant role in shaping emigration patterns.

The current work might be extended in several directions. Firstly, heterogeneous countries might be introduced. Secondly, one could analyse the explicit and implicit requirements for jobs in different (broad) professions to make the $\mathcal{F}$-characteristics and the requirement matrix M more tangible.

5.3 The Effectiveness of Access Restriction to Higher Education in Decreasing Overeducation

As noted in Section 1.2, there are several solutions to the overeducation problem. The fastest and simplest, and thus, most appealing, solution could be restricting access to higher education. The reasoning might be that because there is an obvious excess supply of tertiary graduates, the government should bring that supply in line with demand. It is, however, not obvious that this is the best solution, and whether it is a good solution at all.

In this section, I study the effectiveness of this policy in decreasing overeducation and consider the potential side effects of the policy. I simulate the choice of studying at university and subsequent employment chances. The decision to continue to tertiary-level studies depends on the composition of an individual's social network. Because analytical methods are not applicable to the analysis of models where the dynamics depend on social networks, I use agent-based modelling as the study method. I then study how system behaviour changes if the government creates a ceiling to the number of admissions to university and compare its behaviour to the situation where all applicants continue to be admitted.

In the next subsection, I specify the agent-based model. Then Section 5.3.2 presents the mathematical analysis of a simplified version of the model without social network influence. Section 5.3.3 then discusses the parameter values chosen for simulations and Section 5.3.4 presents the results. Section 5.3.5 analyses the sensitivity of the model results to alternative parameter specifications. The final section concludes.

5.3.1 Model Specification

Agents enter the model with secondary education. Every y model periods, there is an inflow of N new secondary school graduates. New entrants decide whether they continue studies at university; if they do, they graduate after τ model periods. Agents whose age exceeds the maximum-age threshold, $\bar{a}$, are removed from the simulation.

5.3.1.1 Agents, Jobs and Overeducation

Agents, with either secondary or tertiary education, enter the labour market (I assume that students are not allowed to work while they study[46]). There are two types of jobs: high-qualified (HQ), where tertiary education is required, and low-qualified (LQ), where secondary education is sufficient.

There are two types of agents corresponding to these jobs. H-type agents have internal characteristics important for success in HQ jobs. These agents feel comfortable in HQ positions, but have access to them only if they graduate from university.[47] L-type agents do not have these characteristics. Even if they obtain a tertiary degree, employers would screen them out during the selection process. These agents, hence, can only occupy LQ positions. Thus, the type of agent is a simple version of $\mathcal{F}$-characteristics determining the quality of match between an agent and a job as was introduced in Section 5.2. Agent type is assigned probabilistically when the agent enters the system; type H is assigned with probability h.

Table 5-20 shows the types of agent–job correspondence. Note that it gives two kinds of overeducated. Chevalier (2003) also divides the overeducated into two types: the apparently overeducated (having similar unobserved skills as matched graduates) and the genuinely

---

[46] Allowing students to work during university studies will not change the results of the model, as work experience and income are not modelled. Even if they were modelled, the only change would be an increase in work experience by one or two years at the same age. On the contrary, allowing students to drop out of the university (due to having a time-consuming job in parallel to studies or for other reasons) would make a substantially larger change in the results, as it would affect the distribution of education levels in the population.

[47] In principle, an agent without higher education could become an entrepreneur. Thus, he would be classified as a manager, which is a typical HQ position. In denying access to HQ positions to the secondary-educated, I implicitly assume that the majority of these entrepreneurs would continue studies in the near future.

Table 5-20 – Agent–Job Correspondence

	L-Type Agent		H-Type Agent	
	Secondary Educ.	**Tertiary Educ.**	**Secondary Educ.**	**Tertiary Educ.**
LQ Job	LQ matched	(genuinely) overeducated	LQ matched	(apparently) overeducated
HQ Job	—	—	—	HQ matched

overeducated (having much lower skills). Here, the former are H-type overeducated, and the latter are L-type overeducated. I use both terminologies interchangeably.

It is important to relate both types of overeducation to the theories discussed in Section 2.1.1. The distinction of agents into L- and H-type agents and the inability of L-type agents to work in HQ jobs, regardless of their education level, is consistent with human capital theory and the unobserved heterogeneity view of mismatch. Thus, preventing L-type agents from entering university is clearly relevant from the policy perspective, as it makes public investment in education more efficient, especially in continental Europe, where "education is heavily subsidized" (Leuven & Oosterbeek, 2011, p. 287).

The model also takes into account theories assuming the imperfectness of the labour market. It does so implicitly, however, because labour market dynamics are modelled probabilistically depending on the agent's characteristics and its status in the labour market, see Section 5.3.1.3. In particular, the probabilistic structure of labour market dynamics reflects (1) firms' general interest in employees with different education levels, which is associated with job competition theory, and (2) the discrepancies between labour demand and supply at different education levels, which is associated with assignment theory. In addition, the possibility of the existence of apparently overeducated workers is most related to assignment theory. As described in Section 5.3.1.3, this type of overeducation arises mainly because of the complexity of the assignment process (difficulty of finding a matching job fast enough) and factors compensating for the negative effects from overeducation (it is better to work in a job where the agent is mismatched than to have no job at all).

5.3.1.2 Social Networks and Higher Education

Overeducation could be decreased by L-type agents choosing not to enter university studies. It is, however, difficult for them, because agents do not know their type when entering the model. Instead, they use their social networks to guide them in their decision of whether to continue studies.

When entering the model, agents form a random social network with other agents that entered the model in the same period (***same-period network***). They also create a network with older agents (***adult network***). Both networks form the general social network of an agent.[48] Research shows that real-world social networks are scale-free networks, having many low-degree vertices and a few high-degree vertices (see Section 2.2.2). Both networks in this model are built using the Duplication model, which generates scale-free networks (see Section 2.2.2 for details). Both networks are built with the same parameter $p \in (0, 1)$.

Agents decide whether to enter university by the popularity of higher education in the adult network, which is measured by the share of adult friends who are either tertiary graduates or students. This represents the homophily principle (see Section 2.2.2). In particular, if a large part of its adult friends are secondary-educated, the agent would, most probably, also not choose to continue studies. Technically, I set three thresholds on the share of friends with tertiary education, which form four intervals of higher education popularity. The probability of applying to university then depends on the interval where this indicator falls for a given agent. In accordance with the homophily principle, probabilities are higher for intervals covering higher popularity.

One could argue that, given the same popularity of higher education in their social networks, H-type agents should be more willing to enter the university than L-type agents. The argument could, e.g., go through higher aspirations, different (from L-type agents) social background and/or previous educational decisions. While the latter two aspects are not modelled here, the implication that, other things equal, H-type agents have higher desire to get higher education is incorporated in the model.

After making that decision, the agent adds another block of friends to its social network. If it succeeded in entering university, it makes friends with current students (without age limitations). Otherwise (if unwilling to continue studies or denied access to university), it makes friends with other agents currently in the labour market (regardless of their employment status). This reinforces the homophily principle, allowing the tertiary-educated to have a higher share of tertiary-educated friends than the secondary-educated have. In both cases, the network is created by the Duplication model with the same parameter p as before.

5.3.1.3 Labour Market Dynamics

The artificial society in this model believes that having a matched job is always better than having a mismatched job. Agents, thus, always first consider the matching vacancies, and only an

---

[48] I call members of this network as agent's "friends" for short, but they represent both its close friends and acquaintances.

inability to be hired for a matching job for too long (Δt periods) makes them consider other vacancies; in particular, tertiary-educated agents start looking for an LQ job. If they get it, they still continue to search for an HQ job, and if they find one, they leave the LQ job for the HQ job they found.

I do not model separate vacancies or employers. Instead, I operate with a set of probabilities of finding and losing job. The probabilities of losing job are time-independent, while the probabilities of finding job depend on how long the agent has been searching for a job. Time dependence was introduced, because in reality, the probability of exiting from unemployment is unevenly distributed across unemployment spell length (see Section 5.3.3 for more details). Thus, the probability of losing job is denoted by q, while for finding job at k-th, $k \in \mathbb{N}$, period of unemployment by p_k.

Further, these probabilities depend on the quality of match between an agent and the job sought. For instance, the probability of losing an LQ-matched job is denoted by q^l, an HQ-matched job by q^h and a job for which the agent is overeducated by q^o. The same superscripts also apply to the probabilities of finding job.

Tertiary-educated agents face the same probability of finding an LQ job, regardless of their type (i.e., whether they are apparently or genuinely overeducated for it). On the contrary, the chances of quitting or losing an LQ job depend on the type of tertiary-educated agent. In particular, I assume that tertiary-educated H-type agents feel uncomfortable in LQ jobs and are more likely to quit than tertiary-educated L-type agents ($q^{oh} > q^{ol}$). Moreover, I assume that the genuinely overeducated are as comfortable with working in LQ jobs as the LQ-matched. Hence, $q^o \equiv q^{oh} > q^{ol} \equiv q^l$. Recall from Section 2.1.2 that the fact that $q^{oh} > q^l$ has empirical grounding.

5.3.1.4 Government Policies: OEM and REM

I compare two policy regimes in this paper. In the first regime, which I name ***open education market*** (OEM), there are no barriers to university entry. This corresponds to the current situation in Europe, where higher education frequently is state-funded and student loans are easily available for non-funded study programmes. In the second regime, the government recognises that overeducation rate is high and decides to act to reduce it. From the potentially helpful policies described in Section 1.2, it chooses reducing access to higher education. I name this option ***restricted education market*** (REM).

Under REM, some agents wishing to continue studies are denied access to higher education. The reaction of an agent to this restriction depends on its type and aspirations. L-type agents are not very much depressed about it, which means that their labour market dynamics probabilities, p_k^l and q^l, stay unchanged. H-type agents that wish to have higher education, on the contrary, feel frustrated that they do not have access to HQ positions. They, thus, behave in the labour market as the apparently overeducated. In other words, under REM, H-type agents aspiring to higher education but ending up in LQ jobs behave as apparently overeducated, regardless of their actual education level. Note that H-type agents that *choose* to remain secondary-educated behave in LQ jobs as LQ-matched.

The government is unable to determine agent type, nor does it ask the university to run admission tests to select only H-type applicants[49]. It simply introduces a ceiling on the number of applicants the university could admit, which is computed as follows. Assume that in the last year before the period when the government decides to introduce the change in education policy, the number of admitted students was A. The ceiling is then set to

$$\omega A. \tag{5-4}$$

In reality, the university would either actually run admission tests or admit applicants randomly without exceeding the ceiling. I assume that the university uses the latter option; for instance, because they believe that available admission tests might be inadequate for this purpose or because when several applicants have the same score on the test, the university might still have to admit them randomly if it would otherwise exceed the ceiling.

5.3.2 Analysis of REM without Social Networks

The government would, most probably, build some model before it decides on the value of the restriction parameter, ω. Assume that government analysts ignore social network effects and build a purely probabilistic model of agents' decisions with respect to higher education. In this case, agents are assigned the willingness to continue studies with probability u. Assume it is the same for agents of both types.

Denote by $M = N\bar{a}$ the total number of agents in the system. Then the number of tertiary-educated agents under OEM is Mu. If the government implements the restriction in accordance with (5-4), the number of agents with tertiary education will eventually fall to

$$Mu\omega. \tag{5-5}$$

[49] One of the reasons is that a fraction of H-type graduates also become overeducated, so not all H-type applicants should be admitted.

The question is: how will the overeducation rate change as a result?

The total number of the overeducated is the sum of the number of the L-type overeducated and the number of the H-type overeducated. Not going into full details, this number is given by the following expression:

$$M(1-h)uF + MhuG = Mu[(1-h)F + hG].$$

Here, F and G are probabilistic expressions reflecting the length of looking for a job until getting an LQ position for L- and H-type agents, respectively.

It is then clear from (5-5) that under REM, this would eventually fall to

$$Mu\omega[(1-h)F + hG].$$

In other words, the rate of overeducation would drop from o to $o\omega$. Given this result, the government might then decide on the target overeducation rate and set ω accordingly.

In measuring the performance of its policy, the government should take into account that the target overeducation rate, as measured over the total population, will not be achieved until the share of the tertiary-educated in the economy falls from u to $u\omega$, which will happen exactly after $\bar{a}$ years. If unwilling to wait for that long, the government should measure the results of its policy in the young subpopulation. Practically, it is complicated by the labour-market success of the young also depending on how firms view the lack of work experience, which, in turn, depends on whether the economy is expanding or contracting. This is not taken into account in the current model.

Given this analytical result, I now study whether and how the influence of social networks on agent decision-making affects the effectiveness and efficiency of REM.

5.3.3 Parametrisation and Calibration

Parameters are divided into three groups: common parameters, labour-market dynamics parameters and parameters subject to calibration. These are shown in Tables 5-21, 5-23 and 5-24, respectively. This section describes the choice of the pre-set parameters and the calibration procedure.

5.3.3.1 Setting Common Parameters

A model period represents one month. Because secondary school graduation happens every year, $y = 12$. Typical bachelor studies in Europe last for 3–4 years; hence, $\tau = 3y = 36$.

Initially, no agents have adult friends; in the very first period of the model, I, thus, assign entrance to university purely probabilistically. The Duplication model parameter, p, was set to

Table 5-21 – Common Parameter Values

Parameter Name	Notation	Value
Length of simulation, periods		3540
REM introduced in period		3001
Time between inflows of new agents	y	12
Number of agents in each inflow	N	250
Maximum age, years	$\bar{a}$	40
Higher education popularity thresholds		
Threshold 1		30%
Threshold 2		50%
Threshold 3		70%
Prob. of going to university		
In period 1		55%
Length of studies, periods	τ	36
Number of friends		
Same-period network		30
Adult network (min.), when entering the model		10
Adult network (min.), after education choice		10
Duplication model parameter	p	0.45
"Too long", periods	Δt	2

the value that would generate networks with structure similar to social networks.[50] Then, the number of friends the agent makes was chosen for it to have 130–150 friends on average. This is approximately equal to the maximum number of Facebook friends with which an individual interacted at least once (Wilson, Boe, Sala, Puttaswamy, & Zhao, 2009, Fig. 15).

The government in this model chooses the restriction ω equal to one minus the overeducation rate observed in the last period before introducing the reform.

5.3.3.2 Setting Labour Market Dynamics Parameters

Labour market dynamics parameters are set using my own estimates based on micro-level data from the European Social Survey (ESS), round 5 (Norwegian Social Science Data Services, 2010). This round was fielded in 2010–2011. It covers 27 European countries, from which I exclude Russia, Ukraine and Israel. All estimates are done on the pooled sample, applying both design weights and population weights. Thus, the contribution of each of the 24 countries is weighted by its population size. Estimates are done on the sample including both men and women. Round 5 was chosen because at the moment of writing the paper, it was the only ESS round that contained detailed statistics on the distribution of the length of unemployment spell.

[50] If $p = 0.45$, the Duplication model generates scale-free networks with the exponent $\beta \approx 2.61$, which is inside a typical range of $[2, 3]$ reported for scale-free social networks.

In the model, agents can either work or be unemployed and actively seeking job. In other words, there are no inactive agents. Hence, labour market parameters are also set based on respondents who are active in the labour market. In ESS, they are identified based on answers to the question "Which of these descriptions applies to what you have been doing for the last 7 days?". Respondents who marked "paid work (or away temporarily) (employee, self-employed, working for family business)" are deemed employed and those who marked "unemployed and actively looking for a job" are deemed unemployed.

The values of labour-market parameters depend on the measurement of overeducation. The typically mentioned measurement methods are normative, statistical and subjective (see Section 2.1.3). Each has its advantages and disadvantages, but none of these is considered the most appropriate one. I use variants of both objective approaches: ISCO-based approach from the set of normative methods and mean-based approach from the set of statistical methods. Unfortunately, it is impossible to accurately apply the subjective method to ESS data.[51]

For the ISCO-based approach, I use the variables about the level of education (ISCED levels) and occupation (ISCO88 major groups). According to the standard mapping between ISCO major groups and skill levels represented by ISCED categories, higher education (ISCED 5–6)[52] is required in occupations from ISCO major groups 1–3, secondary (ISCED 3–4)[53] in major groups 4–8, and primary (ISCED 1–2) in major group 9. An individual with a higher level of education than required in their occupation by this matching is considered overeducated. Thus, a tertiary-educated is overeducated when employed in major groups 4–9 and a secondary-educated when employed in major group 9.

The conformance between the type of agent–job correspondence in the model and individual–job correspondence in ESS is as follows. HQ jobs are ISCO major groups 1–3, while LQ jobs are ISCO major groups 4–8.[54] Then HQ-matched agents appear in ESS as tertiary-educated

[51] In Round 5, there is a question "If someone was applying nowadays for the job you do now, would they need any education or vocational schooling beyond compulsory education? If yes, about how many years of education or vocational schooling beyond compulsory education would they need?". Instead of comparing the job to the respondent's own education level, the question asks for a general comparison of the number of years of education, not even the education level itself. Making an accurate assessment at this level of detail might be difficult for the respondent, and it introduces additional measurement error. Because of that, even if respondents believe that the number of years one would need at their job is less than their own years of education, they may not necessarily feel overeducated. For these reasons, I consider this question as unsuccessful for gauging subjective mismatch. A more direct question about feeling overeducated or overskilled would be a better candidate, and this is what is typically used in studies involving the self-assessment of mismatch.

[52] This includes vocational higher education (ISCED 5B).

[53] This includes post-secondary non-tertiary programmes.

[54] Elementary occupations (major group 9) were excluded here because the secondary-educated are overeducated in these occupations but the model assumes that the secondary-educated cannot be overeducated.

Table 5-22 – Distribution of the Length of the Longest Unemployment Spell Experienced by the Respondent

Length (months)	ISCO-Based Mismatch Measure			Mean-Based Mismatch Measure		
	LQ Matched	HQ Matched	Overeducated	LQ Matched	HQ Matched	Overeducated
1	7.9%	9.4%	10.3%	7.1%	9.6%	10.2%
2	9.4%	13.6%	11.8%	8.6%	14.8%	9.7%
3	11.9%	12.8%	14.9%	11.4%	13.2%	14.2%
4–6	17.6%	26.2%	21.2%	18.6%	25.3%	23.2%
7–12	21.1%	21.1%	19.9%	21.3%	19.6%	23.0%
13+	32.0%	16.9%	21.9%	33.0%	17.5%	19.7%

Source: Estimated from the European Social Survey's question "Thinking just of the last 3 years, what was the longest period in months, if any, that you were continuously unemployed and seeking work?". Estimated only over respondents who had a period of unemployment in the last three years before the survey.

Table 5-23 – Labour Market Parameters

Parameter	Notation	ISCO-Based Mismatch Measure			Mean-Based Mismatch Measure		
		LQ Matched	HQ Matched	Overeducated	LQ Matched	HQ Matched	Overeducated
Prob. of losing job	q	0.8%*	0.5%	1.0%**	0.8%*	0.5%	0.6%**
Prob. of finding job after unemployment of							
1 period	p_1	7.9%	9.4%	10.3%	7.1%	9.6%	10.2%
2 periods	p_2	10.2%	15.0%	13.2%	9.3%	16.4%	10.8%
3 periods	p_3	14.4%	16.6%	19.1%	13.6%	17.4%	17.8%
4–6 periods	p_4, p_5, p_6	9.1%	16.0%	12.8%	9.3%	15.9%	13.5%
7–12 periods	$p_7, \ldots, p_{12}$	8.1%	12.6%	10.2%	8.0%	11.8%	12.1%
13+ periods	$p_{13}, \ldots$	50.0%	50.0%	50.0%	50.0%	50.0%	50.0%

* Includes genuinely overeducated.

** Only apparently overeducated.

respondents employed in ISCO major groups 1–3 (recall that while any agent can graduate university, only an H-type agent can be employed in an HQ job). LQ-matched agents are secondary-educated respondents employed in ISCO major groups 4–8. Overeducated (both apparently and genuinely) agents are tertiary-educated respondents employed in ISCO major groups 4–8.

For the mean-based approach, I use the variables regarding the number of years of education and two-digit ISCO88 occupation groups ("minor groups"). Inside each minor group, separately for each country, the mean and standard deviation of the number of years of education is measured. Then agents employed in the minor group and having more than the mean plus the standard deviation years of education are considered overeducated. There is no clear analogy to HQ and LQ jobs in this case as there was in the previous case. Here, HQ jobs are those where all possible numbers of years of education of university graduates are inside the mean $\pm$ standard deviation interval. The HQ-matched and LQ-matched are, then, tertiary- and secondary-educated, respectively, non-overeducated individuals.

Table 5-22 shows the length of the longest unemployment spell in the last three years (before completing the survey) depending on the current (for the employed) and last (for the currently unemployed) type of match. These data do not fully correspond to what I am trying to measure (the question measures the duration of the *longest* spell *in the last three years*, not the *last* spell), but may, nevertheless, be used as an approximation.[55]

Original data from ESS show that the distribution of the length of unemployment spell is highly uneven. While having detailed data from ESS about the share of the sample experiencing k-month unemployment spell for each k up to 36, there are major distortions at this level of detail: large spikes are observed at multiples of 6 months. For instance, the number of respondents who exited unemployment after 6 months is nearly twice as high as that of respondents who were able to find jobs after 4 or 5 months. One of the reasons for such spikes might be bounded duration of unemployment benefits. However, it is also quite probable that this happens because of inaccuracy, rather than because this is the true picture. To obtain smoother data, I group spell lengths, as shown in Table 5-22.

Based on Table 5-22, using formulae from Proposition E.1, the probabilities of exiting from unemployment depending on the length of unemployment spell are computed. The probabilities of losing job are set based on Proposition E.2. These parameters are shown in Table 5-23.

As the table shows, HQ-matched agents are always less likely to lose their job and more likely to find a new job, given the length of unemployment, as assumed in Section 5.3.1.3. The actual probabilistic dynamics for the overeducated differ somewhat from what one would assume. On the ISCO-based measure, the apparently overeducated are the most likely to quit (with the probability of 1.0 per cent), while on the mean-based measure, the probability to quit is nearly the same as for the HQ-matched and lower than for the LQ-matched (0.6 per cent). The probability of finding a job for which one is overeducated is higher than it is for the LQ-matched, while it is sometimes higher but sometimes lower than for the HQ-matched, depending on the length of unemployment spell.

[55] The overall (i.e., not dividing into the quality of match) distribution of this variable was compared to the distribution of unemployment duration available from Eurostat in 2010 (the year of fielding ESS round 5 questionnaire) over the same countries as used in ESS round 5 and applying ESS round 5 population weights to Eurostat country-level data. Comparison shows that ESS and Eurostat have similar shares of unemployment duration of 1 month, 6–11 months and 12+ months. The only discrepancy is in short-term unemployment: two-month unemployment was experienced by 10% of respondents in ESS and 17% in Eurostat, and 3–5 month unemployment by 22% in ESS and 16% in Eurostat.

Table 5-24 – Calibrated Parameter Values

Parameter Name	Notation	Value
Calibrated using ISCO-based mismatch measure		
Prob. that an agent is of H-type	h	70%
Prob. of going to university, H-type		
popularity below threshold 1		50%
popularity between thresholds 1 and 2		70%
popularity between thresholds 2 and 3		80%
popularity above threshold 3		90%
Prob. of going to university, L-type		
popularity below threshold 1		30%
popularity between thresholds 1 and 2		60%
popularity between thresholds 2 and 3		70%
popularity above threshold 3		80%
Calibrated using mean-based mismatch measure		
Prob. that an agent is of H-type	h	30%
Prob. of going to university, H-type		
popularity below threshold 1		60%
popularity between thresholds 1 and 2		70%
popularity between thresholds 2 and 3		80%
popularity above threshold 3		90%
Prob. of going to university, L-type		
popularity below threshold 1		30%
popularity between thresholds 1 and 2		60%
popularity between thresholds 2 and 3		70%
popularity above threshold 3		80%

5.3.3.3 Calibrating the Remaining Parameters

Two parameters have not yet been set: the probability that the agent is "born" with H-type and the dependence of the probability of the willingness to go to university on the popularity of higher education among the direct friends. Because there are no ready data on which to base these parameters, they have been calibrated to the values at which the model generates the empirically observed tertiary education attainment and overeducation rates in period 3000 (immediately before the REM policy is introduced, see Section 5.3.4). The target values of these variables were calculated using the same weighted ESS round 5 data as used in Section 5.3.3.2.

The results of calibration are shown in Table 5-24. Note that the probability of generating an H-type agent is more than twice higher in the model calibrated using the ISCO-based mismatch measure than when using the mean-based mismatch measure. H-type agents are also (nearly) twice as likely as L-type agents are to desire continuing studies even if higher education is unpopular among their direct friends.

5.3.4 Results

For each regime (OEM and REM) and mismatch measure (ISCO- and mean-based), the model was run 50 times to allow the distribution of overeducation rates given by different model runs

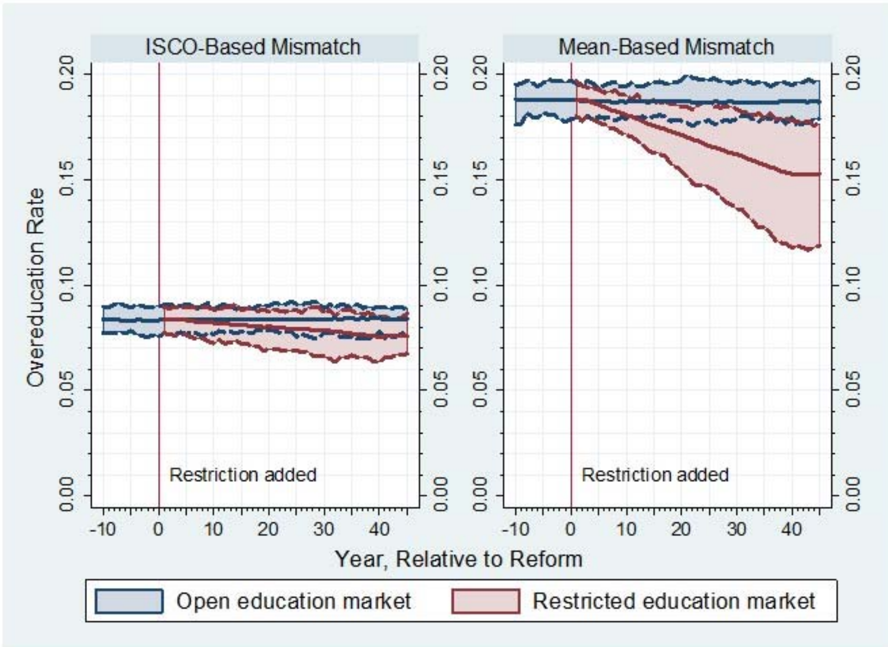

Figure 5-12 – Overeducation Rate Dynamics

Dashed lines represent minimum and maximum values across runs. Solid lines represent average values across runs. Overeducation is measured in the end of academic year (i.e., in the last period before the inflow of new secondary school graduates) to reduce seasonal fluctuations.

to stabilise. The restriction was introduced in period 3001 (250 years from the initialisation of the model) to give the system enough time to stabilise tertiary attainment rates and social network structures. System behaviour was tracked 10 years before the reform and 45 years (5 years beyond the length of life of one generation of agents) after it under status quo (OEM) and restricted admission (REM).

Figure 5-12 shows that the reform is able to decrease the average rate of overeducation, although it takes 40 years, or one generation of agents, for the labour market to stabilise it, which is in line with the theoretical analysis in Section 5.3.2. The range of overeducation distribution under REM, however, overlaps with that under OEM when overeducation is gauged by the ISCO-based measure. Thus, it is tempting to conclude that while in both ISCO-based and mean-based mismatch measurement cases, the reform is effective in terms of successfully decreasing overeducation, it is more efficient in decreasing mean-based overeducation.

The accuracy of this conclusion depends on how the efficiency of this reform is measured. One of the ways is to compare the decrease in overeducation 45 years from the introduction of the reform, o_{45}, relative to overeducation immediately before the reform, o_0, with the restriction on the number of admitted students, $(1-\omega)$:

Table 5-25 – REM Reform Efficiency

Measure of Overeducation	ISCO-Based	Mean-Based
% drop in overeducation	9.3	18.5
% denied access to education	8.3	18.8
Efficiency	**1.12**	**0.99**

Based on average overeducation over runs.

$$\frac{-(o_{45} - o_0)/o_0}{1 - \omega}. \tag{5-6}$$

The reform is effective if (5-6) is positive, but it is efficient only if it is larger than or equal to 1.0. If it is smaller than 1.0, it damages the education market more than it benefits the labour market. Recall that in this simulation, $1 - \omega = o_0$, and the theoretical analysis in Section 5.3.2, thus, shows that without the presence of social networks, the theoretical efficiency of this reform is exactly 1.0, because overeducation will eventually drop exactly by o_0.

According to Table 5-25, the reform is efficient for the ISCO-based model, but not for the mean-based model. In the latter case, the actual drop in overeducation is very close to the theoretical drop, i.e., social networks do not seem to play any major role in amplifying or dampening the theoretically expected result of the reform on average. Nevertheless, returning to Figure 5-12, the range of overeducation under REM on the mean-based measure is substantial, so that while on average, the reform is not efficient, it might be efficient in some cases.

On the ISCO-based measure, the share of apparently overeducated among the overeducated fluctuates around 5 per cent, and on the mean-based measure, it fluctuates around 1.25 per cent.

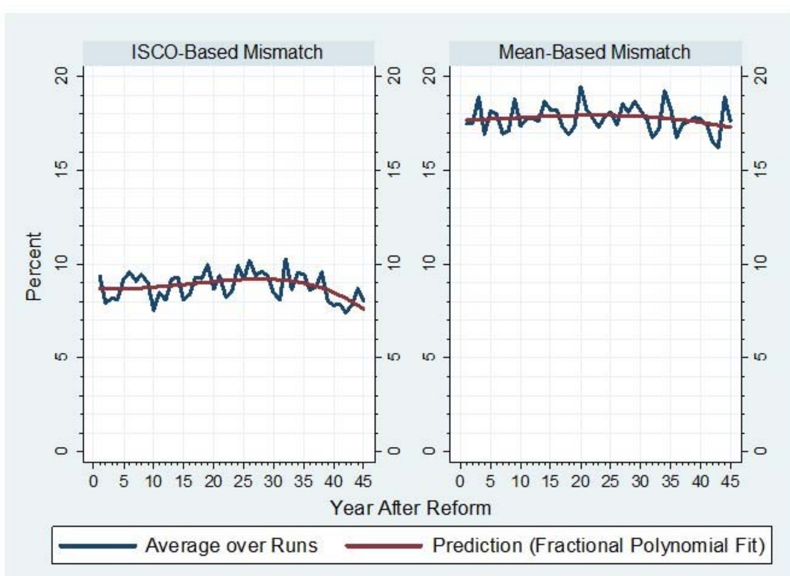

Figure 5-13 – Agents Not Admitted to the University from Fresh Graduates Desiring to Continue Studies

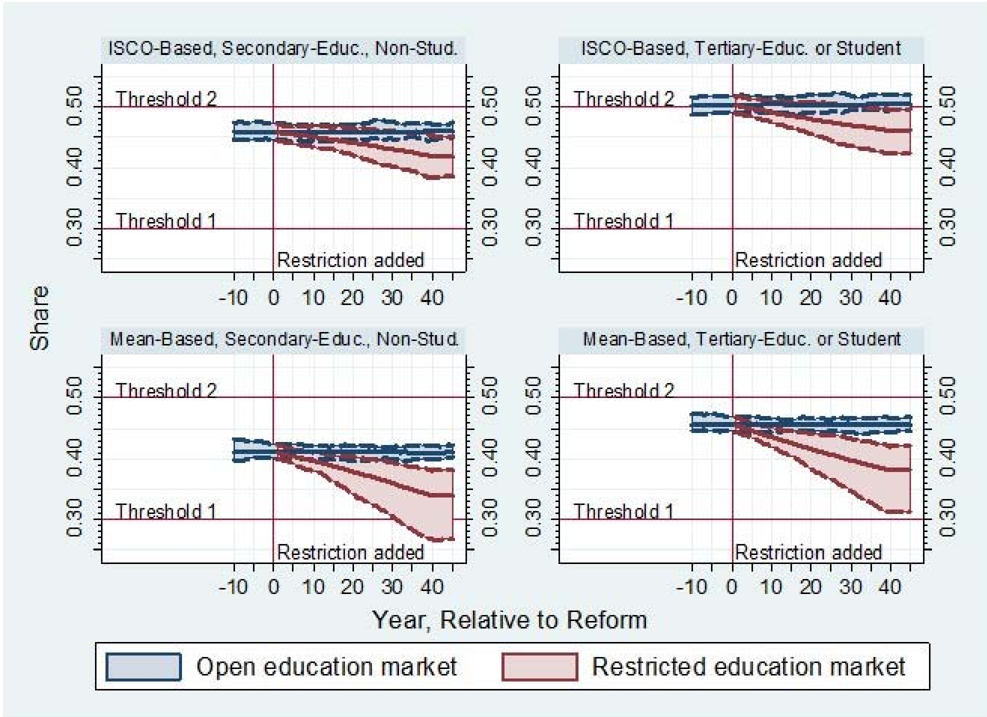

Figure 5-14 – Average Share of Students and Tertiary Educated Agents among Agent's Friends

Dashed lines represent minimum and maximum values across runs. Solid lines represent average values across runs. Statistics measured in the end of academic year (i.e., in the last period before the inflow of new secondary school graduates) to reduce seasonal fluctuations.

In both cases, thus, the vast majority of the overeducated are genuinely (L-type) overeducated. Unfortunately, the restriction does not change this distribution. Thus, while overeducation is decreased, apparent overeducation still exists.

Figure 5-13 shows that the average desire to go to university (represented by the share of agents who were denied access to university from all agents willing to get higher education when entering the model) remains stable during 30 years after the reform, after which it starts sliding down. The downward trend is, however, modest: on average from year 30 to year 45 after the reform, it declined only one percentage point in both models.

This long stability of the willingness to study at university could be caused by the inability of the reform to substantially change the composition of the social network of the secondary- and tertiary-educated. Figure 5-14 illustrates this. In the model where the mean-based measure was used for calibration, the average share of tertiary-educated friends in the agents' networks generally remains between the first and second thresholds for both secondary- and tertiary-educated agents. This seems to be the main cause of the inefficiency of this reform (recall Table 5-25). By contrast, where the ISCO-based measure was used for calibration, the share for tertiary-

Table 5-26 – Labour Market Parameters for Spain, France and the United Kingdom

	Spain						France						United Kingdom					
	ISCO-Based			Mean-Based			ISCO-Based			Mean-Based			ISCO-Based			Mean-Based		
Parameter	LM	HM	OE	LM	HM	OE	LM	HM	OE	LM	HM	OE	LM	HM	OE	LM	HM	OE
$\boldsymbol{q}$	2.0	0.6	1.0	1.8	1.0	0.6	0.6	0.8	1.6	0.7	0.9	0.8	0.7	0.5	1.0	0.7	0.3	0.8
$\boldsymbol{p_1}$	7.4	8.8	9.2	3.3	9.7	7.6	7.4	6.6	19.1	6.9	8.5	12.5	10.7	9.6	9.6	11.5	1.6	17.4
$\boldsymbol{p_2}$	13.8	18.4	6.8	13.4	24.1	6.8	7.7	14.1	16.7	6.8	13.0	22.9	19.9	16.0	25.3	15.9	29.5	14.1
$\boldsymbol{p_3}$	13.4	18.6	14.5	13.6	27.1	5.9	11.1	8.2	23.3	8.9	10.7	13.0	15.6	30.2	33.9	12.6	27.9	37.7
$\boldsymbol{p_4, p_5, p_6}$	8.1	18.5	6.8	8.1	19.8	8.5	5.5	14.0	7.0	6.2	11.5	16.9	12.0	16.1	37.6	14.2	31.4	25.1
$\boldsymbol{p_7, \dots, p_{12}}$	8.8	15.8	16.4	8.7	19.1	13.9	7.1	12.3	2.9	7.1	9.8	9.3	9.3	13.4	22.2	9.3	8.2	37.1
$\boldsymbol{p_{13}, \dots}$	50.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0

All numbers shown are percentages. Calculated from country-specific data from ESS Round 5 using the same procedure as in Table 5-23. Design weights were applied. The restrictions on the sample are as described in Section 5.3.3.2. Abbreviations: 'LM' – LQ Matched, 'HM' – HQ Matched, 'OE' – Overeducated.

Table 5-27 – Calibrated Parameters for Europe, Spain, France and the United Kingdom

	Europe		Spain		France		United Kingdom	
Measure of Overeducation	ISCO-Based	Mean-Based	ISCO-Based	Mean-Based	ISCO-Based	Mean-Based	ISCO-Based	Mean-Based
Prob. that agent's type $= H$	70%	30%	65%	45%	65%	40%	65%	55%
Prob. of going to university, H-type								
popularity $< T_1$	50%	60%	65%	65%	50%	60%	60%	60%
popularity $\in (T_1, T_2)$	70%	70%	70%	70%	70%	70%	70%	70%
popularity $\in [T_2, T_3)$	80%	80%	80%	80%	80%	80%	80%	80%
popularity $\geq T_3$	90%	90%	90%	90%	90%	90%	90%	90%
Prob. of going to university, L-type								
popularity $< T_1$	30%	30%	50%	55%	20%	30%	45%	50%
popularity $\in (T_1, T_2)$	60%	60%	60%	60%	60%	60%	60%	60%
popularity $\in [T_2, T_3)$	70%	70%	70%	70%	70%	70%	70%	70%
popularity $\geq T_3$	80%	80%	80%	80%	80%	80%	80%	80%

T_i denotes ith popularity threshold. European parameters repeated here for convenience.

educated agents dropped from mostly above the second threshold to between the first and the second thresholds. For secondary-educated agents, however, the share remained between the first and the second thresholds. Hence the efficiency of the reform in this case.

5.3.5 Sensitivity Analysis

After reading the previous section, two questions may have appeared in the reader's mind. Firstly, the model was calibrated to European "average" labour–education market system. What would change if the model was fit to other labour–education market systems? Secondly, the consequences of one concrete restriction size were analysed. What would change if the restriction size was different? This section sheds light on these questions.

5.3.5.1 Re-Calibrating to the Cases of France, Spain and the UK

Europe is, of course, not homogeneous with respect to the situation in its labour markets. Thus, there is a need to investigate how the results of the model change in labour market settings different from the overall European settings used so far. For this purpose, I compare the results for Europe with the independent results for Spain, France and the United Kingdom. These countries were chosen because they have enough observations in ESS round 5 to estimate all labour-market parameters. Table 5-26 presents parameter values for these countries. Using these, the calibration procedure from Section 5.3.3.3 was applied (see Table 5-27 for the calibrated parameter values).

According to ESS round 5 data, Spain, France and the UK are quite different with respect to tertiary attainment and overeducation. Tertiary attainment in France (37 per cent) is close to the European average (39 per cent), while more than half of population in Spain and the UK have tertiary education (56 and 51 per cent, respectively). Similarly, ISCO-based overeducation in France (6.7 per cent) is close to the European rate (8.0 per cent), while the rate is nearly twice higher in Spain (15.3 per cent) and the UK (14.1 per cent). The differences in mean-based overeducation are smaller. The rate is much higher in Spain (21.2 per cent) than in France (15.3 per cent), while that of the UK (18.5 per cent) is close to the European rate (17 per cent).

These countries also have different probabilistic structures of labour-market dynamics (see Table 5-26). The probability of quitting is higher for the overeducated than for the LQ-matched in France and the UK, but it is half that of the LQ-matched in Spain. Compared to the HQ-matched, however, the overeducated are twice as likely to quit in all three countries in the ISCO-based models and in the UK in the mean-based model, but less likely to quit in Spain and France in the mean-based models.

In France, the overeducated are generally able to find a job in the short- and medium-term (first 3–6 months) faster than either the LQ- or HQ-matched. In the ISCO-based model of the UK, the same situation holds in the whole first year of unemployment. In the mean-based model of the UK, the overeducated are more likely to find a job than the LQ-matched, but not always more likely than the HQ-matched. Compared to France and the UK, the probabilities of finding a job for the overeducated in Spain in months 1–6 are much smaller. It is always easier for a Spanish tertiary-educated to find a matched job rather than a mismatched job, according to the mean-based criterion. On the ISCO-based definition of mismatch, however, the Spanish overeducated are better able to find a job in the first month of unemployment or starting from month 7.

Figure 5-15 shows that there is a substantial overlap between the distributions of overeducation under OEM and REM policies in France when both ISCO- and mean-based mismatch measures were used for calibration. On the contrary, the ranges do not overlap in the

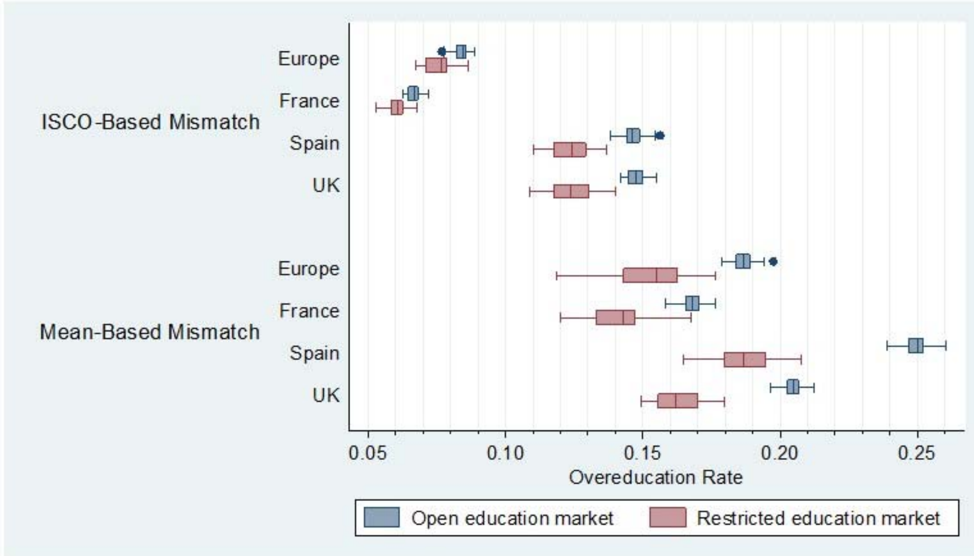

Figure 5-15 - Distribution of Overeducation Rate over Runs 45 Years after Restricting Access to Higher Education, by Country

cases of Spain and the UK, and when the mean-based measure was used for calibration, the drop in overeducation appears to be substantial.

As previously noted, this does not indicate high efficiency of the reform, as seen in Table 5-28. In models calibrated using the ISCO-based measure, the highest efficiency is observed in France, where the drop in overeducation is 1.5 times greater than the restriction applied to the university. In Spain and the UK, the efficiency is lower than in the European model, but still is higher than 1.0. In models calibrated using the mean-based measure, the efficiency of the reform in Spain and the UK, similar to the European case, is close to 1.0. The efficiency in France, however, is only 0.92: Overeducation dropped only 15.4 per cent after 16.7 per cent of potential students were denied access to higher education.

The efficiency of the reform turns out to be not necessarily dependent on whether the average popularity of higher education in agents' social networks moves below another threshold (see Table 5-29), as it seemed in Section 5.3.4. In France, ISCO-based case, for instance, the popularity indicator remained in the same interval $[T_1, T_2]$, but the efficiency was high, exceeding 1.5. By contrast, in the UK and Spain, mean-based cases, the indicator also remained in the same interval, and the efficiency was close to 1.0. In the French mean-based case, the indicator also remained in the same interval, but the reform was clearly inefficient.

Table 5-28 – REM Reform Efficiency in Europe, France, Spain and the United Kingdom

Measure of Overeducation	Europe		France		Spain		United Kingdom	
	ISCO-Based	Mean-Based	ISCO-Based	Mean-Based	ISCO-Based	Mean-Based	ISCO-Based	Mean-Based
% drop in overeducation	9.3	18.5	10.1	15.4	15.3	24.8	15.8	20.7
% denied access to education	8.3	18.8	6.7	16.7	14.6	24.8	14.7	20.5
Efficiency	**1.12**	**0.99**	**1.51**	**0.92**	**1.05**	**1.00**	**1.07**	**1.01**

Based on average overeducation over runs. Results for Europe repeated for convenience.

Table 5-29 – Average Popularity of Higher Education in the Social Network of Agent by Its Level of Education, by Country

Measure of Overeducation	Europe		France		Spain		United Kingdom	
	ISCO-Based	Mean-Based	ISCO-Based	Mean-Based	ISCO-Based	Mean-Based	ISCO-Based	Mean-Based
Secondary-educated, non-student								
Last year before reform	$[T_1, T_2]$	$[T_1, T_2]$	$[T_1, T_2]$	$[T_1, T_2]$	$[T_2, T_3]$	$[T_2, T_3]$	$[T_2, T_3]$	$[T_2, T_3]$
45 years after reform	$[T_1, T_2]$	$[0, T_2]$	$[T_1, T_2]$	$[T_1, T_2]$	$[T_1, T_3]$	$[T_1, T_3]$	$[T_1, T_3]$	$[T_1, T_2]$
Tertiary-educated or student								
Last year before reform	$[T_1, T_3]$	$[T_1, T_2]$	$[T_1, T_2]$	$[T_1, T_2]$	$[T_2, T_3]$	$[T_2, T_3]$	$[T_2, T_3]$	$[T_2, T_3]$
45 years after reform	$[T_1, T_2]$	$[T_1, T_2]$	$[T_1, T_2]$	$[T_1, T_2]$	$[T_1, T_3]$	$[T_2, T_3]$	$[T_1, T_3]$	$[T_1, T_3]$

Based on average overeducation over runs. Results for Europe repeated for convenience.

5.3.5.2 Varying Restriction Size

In a general case, the government might want to introduce a restriction of $1 - \omega = mo_0$, where m is some multiplier and o_0 is, as previously, the overeducation rate immediately before the introduction of the reform. Up to now, the analysis assumed that $m = 1$. Nothing prevents the government from considering other multipliers of course. The theoretical analysis still implies that overeducation would drop to ωo_0, but it is not obvious what the change will be in the presence of social networks.

Having a larger restriction should put more pressure on the share of tertiary-educated friends. As a result, the interest of agents to pursue higher education should decrease, resulting in lower overeducation than expected. In other words, the hypothesis is that efficiency should grow along with the restriction size, m.

Table 5-30 – REM Reform Efficiency by Restriction Size

Multiplier	Efficiency	
	ISCO-Based Measure	Mean-Based Measure
1.0	1.12	0.99
1.5	1.02	0.99
2.0	1.03	1.00
2.5	0.99	1.00
3.0	1.00	1.00
3.5	1.02	1.00
4.0	1.01	1.00

Table 5-30 shows how the efficiency of the reform changes with respect to the restriction multiplier. The efficiency of the reform on the ISCO-based measure quickly drops to nearly 1.0 starting from the multiplier of 2.5. In fact, it becomes only marginally more efficient than theoretically expected already at the multiplier of 1.5. Under the mean-based measure, the efficiency remains at 1.0 at all values of the multiplier.

Overall, it can be concluded that with larger restrictions, social networks cease to play any sizeable role, efficiency converges to theoretical efficiency and, thus, the analytic approach is fully sufficient to estimate the effects of the reform on the labour market.

5.3.6 Discussion and Conclusions

The analysis presented in this section showed that restricting access to university successfully decreases the overeducation rate. If agents' interest in applying for further studies is affected by their social networks, the introduction of admission ceiling sometimes is more efficient than theoretically assumed: the drop in overeducation rate exceeds the share of agents willing to study at university but denied access to it. At the same time, social network effects may also make the reform less efficient than expected, hurting the education market more than helping the labour market.

An important fact shown by the simulation is that the interest of agents in higher education starts falling 30 years after the introduction of the reform. While the trend appears to be moderate, it is, nevertheless, stable. Hence, it can be concluded that the message that the government is not interested in having that many tertiary graduates propagates through social networks and dampens the desire to continue studies. In the future, it might lead to difficulty in attracting students back to higher education. However, these are more long-term consequences, and many other things can happen before that.

One of these is changing demand for skills (e.g., due to technological change), which is not taken into account in the model and, thus, may bias its results. However, even taking into account that the skills demanded in the jobs and in different occupations change over time, there will certainly be space for workers with more or fewer skills than required by the job, i.e., some mismatch will always exist. From the point of view of theories assuming that mismatch is a persistent market failure, overeducation is not really a product of technological change (except as argued by the technological change theory, see Section 2.1.1), and so the presence and size of overeducation is not dependent on it. After all, together with changing demand for skills in the labour market, the skills taught at different education levels also change. Moreover, the job competition theory argues that overeducation is a product of credential inflation, which is even independent from what is taught at different levels of schooling. By focusing on education levels (or years of education), and not on skills, the model keeps a general view of mismatch, and its results should not be biased by changing demand for skills.

More substantial for the results is the assumption that the probabilistic structure governing labour market dynamics remains constant. The analysis in Sections 5.3.4 and 5.3.5 showed that the reform would lead to little or no change in the *overall* overeducation rate in the first 10 years after it is implemented. It was expected already in Section 5.3.2: The reform gives fast results only for youth and propagates through the whole population only after the whole generation changes. The reform will, however, fail to lower overeducation if it is hard to find *any* job for the youth, be it matched or mismatched. Employing labour market policies that boost youth employment, however, would change labour market dynamics probabilities, which would change the predictions of the model regarding the timing and size of the effect of the REM policy on overeducation. Because it makes sense for the government to change the whole labour–education market system, not the education system alone, it should predict the changes in these probabilities and analyse how these impact the effectiveness and efficiency of the changes they want to implement in the education market.

Another important result is that the vast majority (95 to 99 per cent, depending on the calibration assumptions) of the overeducated are L-type overeducated. Recall that this was not the assumption behind the model, but the result of setting up the empirically-based probabilistic structure of labour market dynamics and calibrating the remaining parameters for the model to generate the empirically observed higher education attainment and overeducation rates. This raises the concern that the benefits of lower overeducation, which, thus, mainly affect L-type agents, can fail to justify the negative consequences for those H-type agents that were willing to but were denied access to further studies if the reform is applied without admission tests.

One would claim that the introduction of an admission restriction would make access to higher education more difficult, but, hence, more appealing, as it would send a stronger signal to the labour market about the graduate's abilities. In the settings analysed in this paper, this logic is wrong, because it violates the basic assumption of the effectiveness of a signal (see Section 2.1.1), which is the negative correlation between the costs of education and the graduate's ability. If the restriction is applied on a random basis, then both H-type and L-type agents have equal chances to remain without higher education, and it will not make them want it more because of that.

The signal, on the contrary, would naturally be stronger if the restriction was coupled with admission tests. Indeed, the existing literature generally shows that admission tests (or, more generally, admission exams or background checks) allow the prediction of the student's performance in the university well enough in different fields of study for these to be adopted (see, e.g., Ali & Ali, 2010; Kuncel, Credé, & Thomas, 2007; Meagher, Lin, & Stellato, 2006; Noble & Sawyer, 2004). So why not use admission tests and keep L-type agents out of the university?

One of the arguments against admission tests is that they do not measure the actual productivity of the potential student in the labour market after graduation, so, it is argued, individuals would be better off if they were allowed to show their higher ability status by getting higher education (Arrow, 1973). However, as was already mentioned, this would justify everyone going to university, and this is exactly what has been happening in Europe. Moreover, especially if higher education is publicly financed, it is much more costly for the society to test the ability of individuals by allowing them to study several years at university as compared to requiring them to pass an appropriate admission test. One of the enhancements of the model would be to allow for the possibility of dropping out of the university, so that, e.g., some L-type agents understand that they do not really want to get higher education. It is, though, still less efficient (in monetary terms) than running an admission test, although it persuades the agent better because it is their own experience, not the result of some test.

Any restriction on admission – whether mechanistic or employing a selection mechanism – is likely to work against the goal of increasing tertiary attainment rates, a goal stated in the EU labour market strategy Europe 2020. This is a complex policy problem and should, hence, be managed as a complex problem. It is not in the interests of the EU to have a highly educated workforce working in low-qualified positions. Neither is it in its interests to have too few high-qualified citizens, as it would impair its competitive position. What is in its interests is to create coordinated policies in education *and* labour markets that would maximise the appropriate (i.e., without overeducation) employability of graduates by moving more resources (increasing incentives for universities to open more studentship positions and for businesses to open more job positions) to selected target areas that are key to improving its competitive position.

I believe that no one is against everyone going to university and not only learning to perform some new tasks that are useful in the labour market, but also learning to take a more general view of the picture and "engage in informed and thoughtful discussion of values and norms" (World Economic Forum, 2014, p. 35). The problem is that (nearly) all these graduates should be able to work in appropriate positions instead of feeling that they wasted time and resources on useless education. Mitigating this problem is a very important policy goal, and it will remain a focus of policy-makers in the foreseeable future.

5.4 Summary

This chapter presented three advanced agent-based models. The first incorporated job satisfaction into a labour-market model, where job satisfaction is based on monetary benefits, social support, job variety and career opportunities. The second model considered how fields of study are chosen and the problems faced by agents when they are unable to correctly choose their best field. The third was concerned with the decision to enter studies at the university and policy responses to the problem of overeducation, where the *laissez-faire* principle was compared with restricting access to education market.

All three models feature elaborate social network components, which were found to have a sizeable influence on model outcomes. For instance, the first model had important macro-level effects on the labour market. The second model showed the influence of social networks on the choice of field of study and emigration decisions. The third model showed that social network effects on the decision to enter the university cannot be removed by implementing a general restriction on entry into HEIs, but should rather be mitigated by admission tests.

Overall, all three models give plausible results about the dynamics in education and labour markets. These models are only a starting point, and more elaborate models would be welcome. In addition, all three models, currently modelling different aspects of the labour–education market system, can be combined in a larger model.

CONCLUSIONS AND RECOMMENDATIONS

The main results of the dissertation are:

- **Overeducation showed a stable increasing trend in 2002–2010:** Individuals who invested in formal education are increasingly failing to find jobs that match their education and, thus, are forced to occupy less demanding positions. This is the key struggle in the area of employability of graduates. (Ch. 1)
- **Overeducation risk is influenced by individual's demographics, personality, immigrant status, industry and labour market history:** Females, immigrants and students who study in parallel to work are at higher risk. Of personality factors, achievement orientation and openness to experience are buffers against, while social orientation a catalyst of, overeducation. Of immigrants, the highest risk premium was found for immigrants from CEE and Former Soviet Union countries. Across industries, the highest risk was found in finance, administrative activities, accommodation and public administration, but the influence from the traditional association of sexes with industries was also observed. Tenure helps decreasing overeducation, and this effect is consistent across industries, while lengthy unemployment spells and informal employment put individuals at increased risk. (Ch. 3)
- **The effects on overeducation risk from field of study differ markedly across country groups:** Of the four geographically constructed country groups (Northern, Southern, Eastern and Western Europe), only Western Europe shows signs of overproduction of economists (in terms of relative difficulties in finding a matched job). Eastern Europe is the only group where the overeducation risk of engineers is 75 per cent higher than that of economists. Education and health are the only fields whose graduates consistently show the lowest risk of overeducation. (Ch. 3)
- **The temporal dynamics of overeducation are well explained by macro-level variables:** The share of tertiary graduates, the share of occupations from ISCO major groups 1 through 3 (managers through associate professionals) and, to some extent, unemployment rate show strong performance in overeducation models. (Ch. 3)
- **Overeducation has detrimental effects on the psychological state of the individual in the labour market:** It decreases the job satisfaction of master's level graduates (but does not affect that of bachelor level graduates) and increases the propensity to quit the current job. Both effects are further connected to, in particular, the motivation

and productivity of individual and increase their uncertainty in the labour market. (Ch. 4)

- **Overeducation affects the motivation of continuing studies at doctoral level:** The study of doctoral students in Baltic higher education institutions showed that overeducated individuals are more inclined to get a doctoral degree for personal (e.g., own new achievements), rather than labour-market (e.g., better career prospects), reasons. (Ch. 4)
- **Vocational education is not necessarily a shield from overeducation and its effect on the propensity to quit the current job:** Individuals having vocational education at both secondary and tertiary levels are more likely to both become overeducated and to quit the current job in case of becoming overeducated than those with general secondary and academic tertiary education are. The analysis was done on Swiss data, which makes these findings especially important, because Switzerland is often used as an example of an excellent vocational education system. (Ch. 4)
- **Restricting access to higher education will keep overeducation low, but the use of admission tests is recommended:** The mechanism of restricting access on a random basis is effective in decreasing overeducation. Admission tests ensure that individuals with high profiles are not left out of the higher education system just because their place is occupied by low-profile individuals having been admitted by chance. (Ch. 5)
- **Job satisfaction is mostly affected by (in order of importance) job content, risks and compensation, although supporting environment is also important:** It was shown that job-related risks are more important for job satisfaction than compensation, which is often overlooked in studies of job satisfaction. Supporting environment (e.g., help of co-workers, ability to manage own time) has lower effect, but it was still found to be important, as it allows mitigating stressful situations. (Ch. 4)
- **Graduates of different education levels have different combinations of factors that affect job satisfaction:** Masters are more sensitive than bachelors are to career opportunities, job variety and whether learning is required in the job. Bachelors are more sensitive to the risk of moving to a less interesting job and monetary compensation. (Ch. 4)
- **Higher education institutions (HEIs) should understand both the goals prospective doctoral students pursue when deciding to get a doctoral degree and the factors that are important for them in choosing the HEI for doctoral studies:** The study of

the doctoral students of Baltic HEIs showed that individuals choose to get a doctoral degree mostly for personal reasons. The distribution of the importance of personal and labour-market goals depends heavily on the field of study and the country where the HEI is located, as well as individual's demographics, family and social circle, labour-market experience and previous educational paths. The most important characteristics of HEIs differ substantially across countries, but less so across the types of goal pursued. The most useful information sources are similar across both countries and types of goal, but important dependence on the type of goal was identified. (Ch. 4)

Based on these results, the following recommendations are made:

- **If available, the data on individual's personality should be included in the model of his or her overeducation status alongside the data approximating his or her ability:** Personality frequently performs in overeducation models better than ability approximated by current income. (Ch. 3)
- **The normative (ISCO-based) overeducation measure should be used alongside the subjective overskilling measure, as the former is sometimes more appropriate than the latter:** Each mismatch measure has its disadvantages, and some authors claim that the latter measure tends to be more appropriate than the former in the models of mismatch. Nevertheless, it was shown here that the former measure affects the propensity to quit the current job stronger than overskilling, although the probability of quitting is the highest for those who are both overeducated and overskilled. (Ch. 4)
- **Agent-based models should be used for modelling LEMS:** Three agent-based models presented in this work confirmed that agent-based modelling is an effective modelling technique for creating complex models with elaborate individual decision-making depending on social networks, which generate realistic results and whose output may be analysed to guide policy reforms. (Ch. 5)
- **Job satisfaction should be included in the agent-based models of the labour market as a unified mechanism guiding the behaviour of agents:** Basing agents' actions of searching or quitting job on job satisfaction dynamics, where job satisfaction is a complex concept depending not only on monetary factors, allows to generate realistic dynamics of LEMS. (Ch. 5)
- **LEMS models including multiple fields of study should define the mechanisms for choosing the field of study and the quality of match between field and occupation:**

In particular, the factors important for choosing the field and those responsible for the quality of match should be different. In the theoretical framework developed in this work, these are $\mathcal{H}$- and $\mathcal{F}$-characteristics of agents, respectively. (Ch. 5)

- **LEMS models should be checked for sensitivity to social-network effects:** Agent-based models are especially well suited to study the effects from social networks. The models developed in this work showed that including social networks in models could substantially distort theoretically expected results. In particular, they may affect the effectiveness and efficiency of a policy or the distribution of the demand for fields of study or explain smoother emigration patterns than expected in the absence of social networks. (Ch. 5)

The main practical recommendation to the ministries and agencies that develop labour market and education policies is that **the problem of increasing overeducation should be tackled synchronously on both sides of LEMS: in the labour market and in the education market**. Coordinated policies should be implemented in both markets, because the results are likely to be limited if only one side of LEMS is targeted. Of course, concrete recommendations should be given to countries based on their situation, and this dissertation does not aim at analysing the situation within every European country. Nevertheless, it is possible to give more specific ideas about the ways how the above recommendation could be realised. These include:

- **The selectivity of entrance to the higher education system should be increased, e.g., by using admission tests:** The expansion of higher education partly was allowed by the reduction in the selectivity of entrance to HEIs. Credential inflation that resulted from low selectivity adds to the rate of overeducation resulting purely from LEMS rigidities. One of the ways to keep overeducation low is to control the supply of tertiary graduates. It was shown here that a simple cap on the number of entrants into HEIs is able to decrease overeducation, but admission tests would be more appropriate to select the best HEI students from all secondary school graduates.
- **Government incentives to get higher education (in general or in a concrete field of study) should correspond to the volume of jobs that require that education:** Taking into account many negative consequences for an individual from being overeducated, secondary school graduates should not be motivated to enter higher education system if there is a lack of appropriate employment opportunities for them after graduation (but enough jobs exist for the secondary-educated). If policy-makers, nevertheless, choose to provide incentives to increase the number of HEI students,

they should ensure a corresponding number of jobs that match the qualifications of these students when they graduate. One of the ways to ensure it is to enforce the collaboration between HEIs and firms in understanding the future needs of the economy and provide incentives for firms to involve in the (partial) financing of tuition fees for students in return for a fixed-term contract with these students. That would ensure that the number of graduates “with high debts and mismatched skills” (World Economic Forum, 2014, p. 10) is not as high as it would otherwise be.

- **Specific labour market policies should be targeted on the groups of individuals who face a higher risk of overeducation:** Individuals who are in risk groups, as identified here, should be informed about that and should be provided with incentives to change their education or help in finding the jobs with a higher quality of match, perhaps outside their home country.
- **Both vocational and academic education should be promoted, clarifying to the public the benefits and drawbacks of each of them:** It was shown here that vocational education, although highly promoted in the recent years in some European countries, does not guarantee ending up in a job that matches the graduate’s education and, hence, is not a panacea against overeducation. Policy-makers should, thus, not fall into the trap of unconditionally favouring one type of education over the other. In this sense, this recommendation is in line with that of the World Economic Forum (2014, p. 10) to “adapt and integrate professional and academic education.”
- **Government incentives favouring higher education in the concrete fields of study should not exclude the possibility to get higher education in any other available field:** The points mentioned above do not mean that higher education should be *exclusively* granted in a list of selected fields – e.g., the fields that are important for maintaining or increasing the competitiveness of the country (or the EU as a whole). Policy-makers should find a balance between increasing the selectivity of access (in general and in selected fields) and keeping higher education open for students with limited financial resources. All fields are important for society, and the ultimate challenge of policy-makers is “not only to prepare the future workforce, but to do so in a way that preserves the role of higher education in focusing minds on the bigger picture” (World Economic Forum, 2014, p. 35).

WORKS CITED

Abdou, M., & Gilbert, N. (2009). Modelling the Emergence and Dynamic of Social and Workplace Segregation. *Mind & Society, 8*(2), 173-191. doi:10.1007/s11299-009-0056-3

Abramowitz, M., & Stegun, I. A. (Eds.). (1972). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables* (10th ed.). Washington, DC: National Bureau of Standards.

Acker, G. M. (2004). The Effect of Organizational Conditions (Role Conflict, Role Ambiguity, Opportunities for Professional Development, and Social Support) on Job Satisfaction and Intention to Leave Among Social Workers in Mental Health Care. *Community Mental Health Journal, 40*(1), 65-73. doi:10.1023/B:COMH.0000015218.12111.26

Adamic, L., & Adar, E. (2005). How to search a social network. *Social Networks, 27*(3), 187-203. doi:10.1016/j.socnet.2005.01.007

Addison, J. T., Centeno, M., & Portugal, P. (2004). *Reservation Wages, Search Duration, and Accepted Wages in Europe.* IZA Discussion Paper 1252, IZA.

Agut, S., Peiró, J. M., & Grau, R. (2009). The Effect of Overeducation on Job Content Innovation and Career-Enhancing Strategies Among Young Spanish Employees. *Journal of Career Development, 36*(2), 159-182. doi:10.1177/0894845308327738

Ahn, Y.-Y., Han, S., Kwak, H., Moon, S., & Jeong, H. (2007). Analysis of topological characteristics of huge online social networking services. *Proceedings of the 16th international conference on World Wide Web* (pp. 835-844). New York, NY: ACM. doi:10.1145/1242572.1242685

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*(6), 716-723. doi:10.1109/TAC.1974.1100705

Aleksynska, M., & Tritah, A. (2013). Occupation–education mismatch of immigrant workers in Europe: Context and policies. *Economics of Education Review, 36*, 229-244. doi:10.1016/j.econedurev.2013.06.001

Alexander, J. A., Lichtenstein, R., Oh, H. J., & Ullman, E. (1998). A Causal Model of Voluntary Turnover Among Nursing Personnel in Long-Term Psychiatric Settings. *Research in Nursing & Health, 21*(5), 415-427. doi:10.1002/(SICI)1098-240X(199810)21:5<415::AID-NUR5>3.0.CO;2-Q

Ali, A., & Ali, U. (2010). Predictability of engineering students' performance at the University of Engineering and Technology, Peshawar from admission test conducted by educational testing and evaluation agency (ETEA), NWFP, Pakistan. *Procedia - Social and Behavioral Sciences, 2*(2), 976–982. doi:10.1016/j.sbspro.2010.03.137

Amin, M. (2000). National Infrastructure as Complex Interactive Networks. In T. Samad, & J. Weyrauch, *Automation, Control and Complexity: An Integrated Approach* (pp. 263-286). Chichester: John Wiley & Sons Ltd.

Anderson, J. A. (1984). Regression and Ordered Categorical Variables. *Journal of the Royal Statistical Society. Series B, 46*(1), 1-30.

Anderson, P. W., Arrow, K. J., & Pines, D. (Eds.). (1988). *The Economy as an Evolving Complex System: The Proceedings of the Evolutionary Paths of the Global Economy Workshop, Held September, 1987 in Santa Fe, New Mexico.* Cambridge, MA: Perseus Books.

Arcalean, C., & Schiopu, I. (2010). Public versus private investment and growth in a hierarchical education system. *Journal of Economic Dynamics and Control, 34*(4), 604-622. doi:10.1016/j.jedc.2009.11.006

Argonne National Laboratory. (n.d.a). *Complex Adaptive Systems.* Retrieved Oct 27, 2011, from ANL Decision and Information Sciences: http://www.dis.anl.gov/exp/cas/index.html

Argonne National Laboratory. (n.d.b). *Electricity Market Complex Adaptive System (EMCAS).* Retrieved Nov 7, 2011, from ANL Decision and Information Sciences: http://www.dis.anl.gov/projects/emcas.html

Arrow, K. J. (1973). Higher education as a filter. *Journal of Public Economics, 2*(3), 193–216. doi:10.1016/0047-2727(73)90013-3

Association for the Advancement of Artificial Intelligence. (2010). *Complex Adaptive Systems: Resilience, Robustness, and Evolvability.* Retrieved Nov 2, 2011, from CAS AAAI 2010: http://sites.google.com/site/complexadaptivesystems2010/

Axtell, R. L. (2000). *Why Agents? On the Varied Motivations for Agent Computing in the Social Sciences.* Working Paper, The Brookings Institution, Center on Social and Economic Dynamics.

Axtell, R. L. (2007). What Economic Agents Do: How Cognition and Interaction Lead to Emergence and Complexity. *The Review of Austrian Economics, 20*(2-3), 105-122. doi:10.1007/s11138-007-0021-5

Baert, S., Cockx, B., & Verhaest, D. (2013). Overeducation at the start of the career: Stepping stone or trap? *Labour Economics, 25*, 123-140. doi:10.1016/j.labeco.2013.04.013

Bakker, A. B., & Demerouti, E. (2007). The Job Demands-Resources Model: State of the Art. *Journal of Managerial Psychology, 22*(3), 209-238. doi:10.1108/02683940710733115

Ballot, G. (2002). Modeling the labor market as an evolving institution: model ARTEMIS. *Journal of Economic Behavior & Organization, 49*(1), 51–77. doi:10.1016/S0167-2681(02)00058-6

Ballot, G., Kant, J.-D., & Goudet, O. (2013). Modeling both sides of the French labor market with adaptive agents under bounded rationality. *EAEPE Annual Conference.* Paris, France.

Bankes, S. C. (2002). Agent-Based Modeling: A Revolution? *Proceedings of the National Academy of Sciences, 99*(3), 7199-7200.

Barabási, A.-L. L., & Bonabeau, E. (2003). Scale-free networks. *Scientific American, 288*(5), 60-69.

Baranowska-Rataj, A., & Unt, M. (2012). Is it Worth Becoming an Engineer in Central and Eastern Europe? The Evidence from Poland and Estonia. *European Sociological Review, 28*(6), 717-728. doi:10.1093/esr/jcs054

Barone, C., & Ortiz, L. (2010). Overeducation among European University Graduates: a comparative analysis of its incidence and the importance of higher education differentiation. *Higher Education, 61*(3), 325-337. doi:10.1007/s10734-010-9380-0

Battisti, M. (2013). *Individual Wage Growth: The Role of Industry Experience.* Ifo Working Paper 152, Ifo Institute for Economic Research at the University of Munich.

Bauer, T. K. (2002). Educational mismatch and wages: a panel analysis. *Economics of Education Review, 21*(3), 221-229. doi:10.1016/S0272-7757(01)00004-8

Becker, G. S. (1993). *Human Capital: a theoretical and empirical analysis, with special reference to education* (3rd ed.). Chicago: The University of Chicago Press.

Beckhusen, J., Florax, R. J., Poot, J., & Waldorf, B. S. (2013). Attracting global talent and then what? Overeducated immigrants in the United States. *Journal of Regional Sciene, 53*(5), 834-854. doi:10.1111/jors.12030

Beinhocker, E. D. (2006). *The origin of wealth: evolution, complexity, and the radical remaking of economics.* Boston: Harvard Business School Press.

Betti, G., D'Agostino, A., & Neri, L. (2011). Educational Mismatch of Graduates: a Multidimensional and Fuzzy Indicator. *Social Indicators Research, 103*(3), 465-480. doi:10.1007/s11205-010-9712-6

Bewley, T. F. (1999). *Why Wages Don't Fall During a Recession.* Cambridge, MA: Harvard University Press.

Bianchi, C., Cirillo, P., Gallegati, M., & Vagliasindi, P. A. (2007). Validating and Calibrating Agent-Based Models: A Case Study. *Computational Economics, 30*(3), 245-264. doi:10.1007/s10614-007-9097-z

Biddle, J. C. (2013). Factors Affecting Student Choice of an Education Doctorate: A Case Study in the United States. *Work Based Learning, 3*(2), 35–53.

Birkin, M., & Wu, B. (2012). A Review of Microsimulation and Hybrid Agent-Based Approaches. In A. J. Heppenstall, A. T. Crooks, L. M. See, & M. Batty (Eds.), *Agent-Based Models of Geographical Systems* (pp. 51-68). Dordrecht: Springer. doi:10.1007/978-90-481-8927-4_3

Blau, P. M. (1964). *Exchange and Power in Social Life.* New York: John Wiley & Sons.

Blázquez, M., & Budría, S. (2012). Overeducation dynamics and personality. *Education Economics, 20*(3), 260-283. doi:10.1080/09645292.2012.679338

Blenkinsopp, J., & Scurry, T. (2007). “Hey GRINGO!”: the HR challenge of graduates in non-graduate occupations. *Personnel Review, 36*(4), 623-637. doi:10.1108/00483480710752830

Bogachev, V. I. (2007). *Measure Theory* (Vol. I). Berlin/Heidelberg: Springer.

Boudarbat, B. (2008). Field of study choice by community college students in Canada. *Economics of Education Review, 27*(1), 79-93. doi:10.1016/j.econedurev.2006.06.001

Boudarbat, B., & Montmarquette, C. (2009). Choice of fields of study of university Canadian graduates: the role of gender and their parents' education. *Education Economics, 17*(2), 185-213. doi:10.1080/09645290802133032

Boudon, R. (1974). *Education, Opportunity, and Social Inequality: Changing Prospects in Western Society.* New York: John Wiley & Sons.

Bourdieu, P. (1984). *Disctinction: A Social Critique of the Judgement of Taste.* (R. Nice, Trans.) Boston, MA: Harvard University Press.

Bourdieu, P., & Passeron, J.-C. (1990). *Reproduction in Education, Society and Culture* (2nd ed.). London: Sage Publications.

Bracke, P., Pattyn, E., & Von dem Knesebeck, O. (2013). Overeducation and depressive symptoms: diminishing mental health returns to education. *Sociology of Health & Illness, 35*(8), 1242-1259. doi:10.1111/1467-9566.12039

Brailsford, I. (2010). Motives and Aspirations for Doctoral Study: Career, Personal, and Inter-personal Factors in the Decision to Embark on a History PhD. *International Journal of Doctoral Studies, 5*, 15–27.

Bramoullé, Y., & Saint-Paul, G. (2010). Social networks and labor market transitions. *Labour Economics, 17*(1), 188-195. doi:10.1016/j.labeco.2009.09.011

Brant, R. (1990). Assessing Proportionality in the Proportional Odds Model for Ordinal Logistic Regression. *Biometrics, 46*, 1171–1178.

Breen, R., & Goldthorpe, J. H. (1997). Explaining Educational Differentials: Towards a Formal Rational Action Theory. *Rationality and Society, 9*(3), 275-305. doi:10.1177/104346397009003002

Brenner, T., & Werker, C. (2007). A Taxonomy of Inference in Simulation Models. *Computational Economics, 30*(3), 227-244. doi:10.1007/s10614-007-9102-6

Brough, P., & Frame, R. (2004). Predicting Police Job Satisfaction and Turnover Intentions: The role of social support and police organisational variables. *New Zealand Journal of Psychology, 33*(1), 8-16.

Brown, S., & Taylor, K. (2013). Reservation wages, expected wages and unemployment. *Economics Letters, 119*(3), 276–279. doi:10.1016/j.econlet.2013.02.035

Brynin, M., & Longhi, S. (2009). Overqualification: Major or minor mismatch? *Economics of Education Review, 28*(1), 114-121. doi:10.1016/j.econedurev.2008.01.003

Büchel, F. (2001). Overqualification: Reasons, measurement issues, and typological affinity to unemployment. In P. Descy, & M. Tessaring (Eds.), *Training and learning for competence. Second report on vocational training research in Europe: Synthesis report* (Vol. 2, pp. 453-560). Luxembourg: Office for Official Publications of the European Communities.

Büchel, F. (2002). The effects of overeducation on productivity in Germany – the firms' viewpoint. *Economics of Education Review, 21*(3), 263-275. doi:10.1016/S0272-7757(01)00020-6

Büchel, F., & Battu, H. (2003). The Theory of Differential Overqualification: Does it Work? *Scottish Journal of Political Economy, 50*(1), 1-16. doi:10.1111/1467-9485.00251

Büchel, F., & van Ham, M. (2003). Overeducation, regional labor markets, and spatial flexibility. *Journal of Urban Economics, 53*(3), 482-493. doi:10.1016/S0094-1190(03)00008-1

Budría, S. (2011). Are Educational Mismatches Responsible for the 'Inequality Increasing Effect' of Education? *Social Indicators Research, 102*(3), 409-437. doi:10.1007/s11205-010-9675-7

Button, L., Green, B., Tenhnah, C., Johansson, I., & Baker, C. (2005). The impact of international placements on nurses' personal and professional lives: literature review. *Journal of Advanced Nursing, 50*(3), 315-324. doi:10.1111/j.1365-2648.2005.03395.x

Byrne, M., Flood, B., Hassall, T., Joyce, J., Montaño, J. A., González, J., & Tourna-Germanou, E. (2012). Motivations, expectations and preparedness for higher education: A study of accounting students in Ireland, the UK, Spain and Greece. *Accounting Forum, 36*(2), 134-144. doi:10.1016/j.accfor.2011.12.001

Cai, Y. (2003). Choosing fields of graduate study among Chinese and Taiwanese students. *The Social Science Journal, 40*(3), 495-501. doi:10.1016/S0362-3319(03)00047-8

Cainarca, G., & Sgobbi, F. (2012). The return to education and skills in Italy. *International Journal of Manpower, 33*(2), 187-205. doi:10.1108/01437721211225444

Calvó-Armengol, A., & Jackson, M. O. (2004). The Effects of Social Networks on Employment and Inequality. *The American Economic Review, 94*(3), 426-454. doi:10.1257/0002828041464542

Campbell, D. P., Borgen, F. H., Eastes, S. H., Johansson, C. B., & Peterson, R. A. (1968). A set of basic interest scales for the Strong Vocational Interest Blank for men. *Journal of Applied Psychology, 52*(6), 1-54. doi:10.1037/h0026495

Cappellari, L., & Lucifora, C. (2009). The "Bologna Process" and college enrollment decisions. *Labour Economics, 16*(6), 638-647. doi:10.1016/j.labeco.2009.08.009

Carless, S. A., & Arnup, J. L. (2011). A longitudinal study of the determinants and outcomes of career change. *Journal of Vocational Behavior, 78*(1), 80-91. doi:10.1016/j.jvb.2010.09.002

Carroll, D., & Tani, M. (2013). Over-education of recent higher education graduates: New Australian panel evidence. *Economics of Education Review, 32*, 207-218. doi:10.1016/j.econedurev.2012.10.002

Cedefop. (2012). *Future skills supply and demand in Europe. Forecast 2012.* Research Paper No. 26, European Centre for the Development of Vocational Training, Luxembourg.

Chamberlain, G. (1980). Analysis of Covariance with Qualitative Data. *The Review of Economic Studies, 47*(1), 225-238.

Chen, Y.-M., & Johantgen, M. E. (2010). Magnet Hospital Attributes in European Hospitals: A multilevel model of job satisfaction. *International Journal of Nursing Studies, 47*(8), 1001-1012. doi:10.1016/j.ijnurstu.2009.12.016

Chevalier, A. (2003). Measuring Over-education. *Economica, 70*(279), 509-531. doi:10.1111/1468-0335.t01-1-00296

Chevalier, A., & Lindley, J. (2009). Overeducation and the skills of UK graduates. *Journal of the Royal Statistical Society: Series A, 172*(2), 307-337. doi:10.1111/j.1467-985X.2008.00578.x

Christou, C., & Haliassos, M. (2006). How do students finance human capital accumulation?: The choice between borrowing and work. *Journal of Policy Modeling, 28*(1), 39-51. doi:10.1016/j.jpolmod.2005.07.006

Chung, F., & Lu, L. (2006). *Complex Graphs and Networks (Regional Conference Series in Mathematics, No. 107).* Providence, Rhode Island: American Mathematical Society.

Colman, A. M. (2009). *A Dictionary of Psychology* (3rd ed.). Oxford University Press. Retrieved Aug 1, 2011, from http://www.oxfordreference.com/views/ENTRY.html?subview=Main&entry=t87.e686

Connolly, H., & Gottschalk, P. (2006). *Differences in Wage Growth by Education Level: Do Less-Educated Workers Gain Less from Work Experience?* IZA Discussion Paper 2331, IZA.

Convert, B. (2005). Europe and the Crisis in Scientific Vocations. *European Journal of Education, 40*(4), 361-366. doi:10.1111/j.1465-3435.2005.00233.x

Cortese, C. G., Colombo, L., & Ghislieri, C. (2010). Determinants of nurses' job satisfaction: the role of work–family conflict, job demand, emotional charge and social support. *Journal of Nursing Management, 18*(1), 35-43. doi:10.1111/j.1365-2834.2009.01064.x

Croce, G., & Ghignoni, E. (2012). Demand and Supply of Skilled Labour and Overeducation in Europe: A Country-level Analysis. *Comparative Economic Studies, 54*(2), 413-439. doi:10.1057/ces.2012.12

Cutillo, A., & Di Pietro, G. (2006). The effects of overeducation on wages in Italy: a bivariate selectivity approach. *International Journal of Manpower, 27*(2), 143-168. doi:10.1108/01437720610666182

Dar, Y., & Getz, S. (2007). Learning ability, socioeconomic status, and student placement for undergraduate studies in Israel. *Higher Education, 54*(1), 41-60. doi:10.1007/s10734-006-9045-1

Davies, R. (2008). The Bologna process: The quiet revolution in nursing higher education. *Nurse Education Today, 28*(8), 935-942. doi:10.1016/j.nedt.2008.05.008

Dawid, H., & Gemkow, S. (2014). How do social networks contribute to wage inequality? Insights from an agent-based analysis. *Industrial and Corporate Change, 23*(5), 1171--1200. doi:10.1093/icc/dtt049

Dawid, H., Gemkow, S., Harting, P., & Neugart, M. (2009). On the Effects of Skill Upgrading in the Presence of Spatial Labor Market Frictions: An Agent-Based Analysis of Spatial Policy Design. *Journal of Artificial Societies and Social Simulation, 12*(4).

Dawid, H., Gemkow, S., Harting, P., & Neugart, M. (2012). Labor market integration policies and the convergence of regions: the role of skills and technology diffusion. *Journal of Evolutionary Economics, 22*(3), 543–562. doi:10.1007/s00191-011-0245-1

Dawid, H., Gemkow, S., Harting, P., Kabus, K., Werching, K., & Neugart, M. (2008). Skills, Innovation, and Growth: An Agent-Based Policy Analysis. *Journal of Economics and Statistics, 228*(2–3), 251–275.

Dawid, H., Harting, P., & Neugart, M. (2013). *Spatial labor market frictions and economic convergence: policy implications from a heterogeneous agent model.* Working Paper in Economics and Management 09-2013, Bielefeld University.

Deckert, A., & Klein, R. (2010). Agentenbasierte Simulation zur Analyse und Lösung betriebswirtschaftlicher Entscheidungsprobleme. *Journal für Betriebswirtschaft, 60*(2), 89-125. doi:10.1007/s11301-010-0058-6

Deissenberg, C., van der Hoog, S., & Dawid, H. (2008). EURACE: A massively parallel agent-based model of the European economy. *Applied Mathematics and Computation, 204*(2), 541–552. doi:10.1016/j.amc.2008.05.116

Dekker, R., de Grip, A., & Heijke, H. (2002). The effects of training and overeducation on career mobility in a segmented labour market. *International Journal of Manpower, 23*(2), 106-125. doi:10.1108/01437720210428379

Del Carmen Calatrava Moreno, M., & Kollanus, S. (2013). On the motivations to enroll in doctoral studies in Computer Science – A comparison of PhD program models. *Information Technology Based Higher Education and Training (ITHET), 2013 International Conference on*, (pp. 1–8). doi:10.1109/ITHET.2013.6671028

DG Employment, Social Affairs and Inclusion. (2012). *Employment and Social Developments in Europe 2012.* Luxembourg: Publications Office of the European Union. Retrieved from http://ec.europa.eu/social/BlobServlet?docId=9604&langId=en

Di Pietro, G. (2002). Technological change, labor markets, and 'low-skill, low-technology traps'. *Technological Forecasting and Social Change, 69*(9), 885-895. doi:10.1016/S0040-1625(01)00182-2

Díaz-Serrano, L., & Cabral Vieira, J. (2005). *Low Pay, Higher Pay and Job Satisfaction within the European Union: Empirical Evidence from Fourteen Countries.* Discussion Paper No. 1558, IZA.

Diem, A., & Wolter, S. C. (2014). Overeducation among Swiss university graduates: determinants and consequences. *Journal for Labour Market Research, 47*(4), 313-328. doi:10.1007/s12651-014-0164-3

Docquier, F., & Marfouk, A. (2005). *International Migration by Educational Attainment (1990-2000) - Release 1.1.* Update to the World Bank Policy Research Working Paper No. 3382. Retrieved from http://perso.uclouvain.be/frederic.docquier/filePDF/DM_ozdenschiff.pdf

Ducharme, L. J., & Martin, J. K. (2000). Unrewarding Work, Coworker Support, and Job Satisfaction: A Test of the Buffering Hypothesis. *Work and Occupations, 27*(2), 223-243. doi:10.1177/0730888400027002005

Duncan, G. J., & Hoffman, S. D. (1981). The incidence and wage effects of overeducation. *Economics of Education Review, 1*(1), 75-86. doi:10.1016/0272-7757(81)90028-5

EHEA Ministerial Conference. (2012). *Bucharest Communiqué.* Retrieved from http://www.ehea.info/Uploads/%281%29/Bucharest%20Communique%202012%281%29.pdf

European Commission. (2012, Nov 16). *The EU Framework Programme for Research and Innovation.* Retrieved Jul 12, 2013, from Horizon 2020: http://ec.europa.eu/research/horizon2020/index_en.cfm?pg=h2020

European Commission. (2013, Apr 30). *ERA Communication.* Retrieved Jul 13, 2013, from European Research Area: http://ec.europa.eu/research/era/era_communication_en.htm

European Commission. (2013, Jan 21). *Europe 2020 Targets.* Retrieved Jul 5, 2013, from Europe 2020: http://ec.europa.eu/europe2020/europe-2020-in-a-nutshell/targets/index_en.htm

European Council. (2013, Feb 7/8). *Conclusions (Multiannual Financial Framework).* Retrieved Jul 12, 2013, from Council of the European Union: http://www.consilium.europa.eu/uedocs/cms_data/docs/pressdata/en/ec/135344.pdf

European Parliament & EU Council. (2006, Dec 30). Recommendation of the European Parliament and of the Council of 18 December 2006 on Key Competences for Lifelong Learning. *Official Journal of the European Union*, pp. 11-18. Retrieved from http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2006:394:0010:0018:EN:PDF

Eurydice. (2012). *National Student Fee and Support Systems 2011/12.* Retrieved from http://eacea.ec.europa.eu/education/eurydice/documents/facts_and_figures/fees_and_support.pdf

Fagiolo, G., Dosi, G., & Gabriele, R. (2004). Matching, bargaining, and wage setting in an evolutionary model of labor market and output dynamics. *Advances in Complex Systems, 7*(2), 157–186. doi:10.1142/S0219525904000135

Fagiolo, G., Moneta, A., & Windrum, P. (2007). A Critical Guide to Empirical Validation of Agent-Based Models in Economics: Methodologies, Procedures, and Open Problems. *Computational Economics, 30*(3), 195-226. doi:10.1007/s10614-007-9104-4

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters, 27*(8), 861-874. doi:10.1016/j.patrec.2005.10.010

Feldman, K. A., Ethington, C. A., & Smart, J. C. (2001). A Further Investigation of Major Field and Person-Environment Fit: Sociological versus Psychological Interpretations of Holland's Theory. *The Journal of Higher Education, 72*(6), 670-698.

Fernández, C., & Ortega, C. (2008). Labor market assimilation of immigrants in Spain: employment at the expense of bad job-matches? *Spanish Economic Review, 10*(2), 83-107. doi:10.1007/s10108-007-9032-4

Figueira, M. L., Caeiro, L., Ferro, A., Cordeiro, R., Duarte, P. M., Akiskal, H. S., & Akiskal, K. K. (2010). Temperament in Portuguese university students as measured by TEMPS-A: Implications for professional choice. *Journal of Affective Disorders, 123*(1), 30-35. doi:10.1016/j.jad.2009.10.005

Frank, R. H. (1978). Why Women Earn Less: The Theory and Estimation of Differential Overqualification. *The American Economic Review, 68*(3), 360-373.

Freeman, J. A., & Hirsch, B. T. (2008). College majors and the knowledge content of jobs. *Economics of Education Review, 27*(5), 517-535. doi:10.1016/j.econedurev.2007.07.004

Frei, C., & Sousa-Poza, A. (2012). Overqualification: permanent or transitory? *Applied Economics, 44*(14), 1834-1847. doi:10.1080/00036846.2011.554380

Frenette, M. (2004). The overqualified Canadian graduate: the role of the academic program in the incidence, persistence, and economic returns to overqualification. *Economics of Education Review, 23*(1), 29-45. doi:10.1016/S0272-7757(03)00043-8

Froot, K. A. (1989). Consistent Covariance Matrix Estimation with Cross-Sectional Dependence and Heteroskedasticity in Financial Data. *Journal of Financial and Quantitative Analysis, 24*(3), 333-355. doi:10.2307/2330815

Garcia-Garcia, M. J., Gonzales, C., & Argüelles, R. (2009). Methodological changes in technical teaching in order to the European higher education area comparison between countries: Italy and Spain. *Procedia - Social and Behavioral Sciences, 1*, 2701-2706. doi:10.1016/j.sbspro.2009.01.478

García-Mainar, I., & Montuenga-Gómez, V. M. (2005). Education returns of wage earners and self-employed workers: Portugal vs. Spain. *Economics of Education Review, 24*(2), 161--170. doi:10.1016/j.econedurev.2004.07.004

Gemkow, S., & Neugart, M. (2011). Referral hiring, endogenous social networks, and inequality: an agent-based analysis. *Journal of Evolutionary Economics, 21*(4), 1-17. doi:10.1007/s00191-011-0219-3

Ghignoni, E., & Verashchagina, A. (2014). Educational qualifications mismatch in Europe. Is it demand or supply driven? *Journal of Comparative Economics, 42*(3), 670-692. doi:10.1016/j.jce.2013.06.006

Goyette, K. A., & Mullen, A. L. (2006). Who Studies the Arts and Sciences? Social Background and the Choice and Consequences of Undergraduate Fields of Study. *The Journal of Higher Education, 77*(3), 497-538.

Granovetter, M. (2005). The impact of social structure on economic outcomes. *Journal of Economic Perspectives, 19*(1), 33-50. doi:10.1257/0895330053147958

Greene, W. H. (2008). *Econometric Analysis* (6th ed.). Upper Saddle River, New Jersey: Pearson Prentice Hall.

Groeneveld, S., & Hartog, J. (2004). Overeducation, wages and promotions within the firm. *Labour Economics, 11*(6), 701-714. doi:10.1016/j.labeco.2003.11.005

Groot, W., & van den Brink, H. M. (2000). Skill mismatches in the Dutch labor market. *International Journal of Manpower, 21*(8), 584-595. doi:10.1108/01437720010379493

Guimerà, R., Danon, L., Díez-Guilera, A., Giralt, F., & Arenas, A. (2003). Self-similar community structure in a network of human interactions. *Physical Review E, 68*(6), 065103(R). doi:10.1103/PhysRevE.68.065103

Guironnet, J.-P., & Peypoch, N. (2007). Human capital allocation and overeducation: A measure of French productivity (1987, 1999). *Economic Modelling, 24*(3), 398-410. doi:10.1016/j.econmod.2006.09.003

Gül, H., Gül, S. S., Kaya, E., & Alican, A. (2010). Main trends in the world of higher education, internationalization and institutional autonomy. *Procedia - Social and Behavioral Sciences, 9*, 1878-1884. doi:10.1016/j.sbspro.2010.12.417

Hackett, F., & Lent, R. W. (1992). Theoretical advances and current inquiry in career psychology. In S. D. Brown, & R. W. Lent (Eds.), *Handbook of counselig psychology* (pp. 419-451). Oxford: John Wiley & Sons.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). *Multivariate Data Analysis* (6th ed.). Upper Saddle River, NJ: Pearson Education.

Hand, D., Mannila, H., & Smyth, P. (2001). *Principles of Data Mining.* Cambridge, MA: The MIT Press.

Hansen, J. A., & Lehmann, M. (2008). Agents of change: universities as development hubs. *Journal of Cleaner Production, 14*(9-11), 820-829. doi:10.1016/j.jclepro.2005.11.048

Harris, J. I., Winskowski, A., & Engdahl, B. E. (2007). Types of workplace social support in the prediction of job satisfaction. *The Career Development Quarterly, 56*(2), 150-156.

Hartog, J. (2000). Over-education and earnings: where are we, where should we go? *Economics of Education Review, 19*(2), 131-147. doi:10.1016/S0272-7757(99)00050-3

Haug, S. (2008). Migration Networks and Migration Decision-Making. *Journal of Ethnic and Migration Studies, 34*(4), 585-605. doi:10.1080/13691830801961605

Hayes, L. J., O'Brien-Pallas, L., Duffield, C., Shamian, J., Buchan, J., Hughes, F., . . . North, N. (2012). Nurse turnover: A literature review – An update. *International Journal of Nursing Studies, 49*(7), 887–905. doi:10.1016/j.ijnurstu.2011.10.001

Hazans, M. (2011a). *Informal Workers Across Europe: Evidence from 30 Countries.* Discussion Paper No. 5871, IZA.

Hazans, M. (2011b). Labor market integration of ethnic minorities in Latvia. In M. Kahanec, & K. F. Zimmerman (Eds.), *Ethnic diversity in European labor markets: Challenges and solutions* (pp. 163-197). Cheltenham, UK - Northampton, MA, USA: Edward Elgar.

Hoff, K., & Stiglitz, J. E. (2001). Modern Economic Theory and Development. In G. M. Meier, & J. E. Stiglitz (Eds.), *Frontiers of Development Economics: The Future in Perspective* (pp. 389-459). New York: Oxford University Press.

Holland, J. H. (2006). Studying Complex Adaptive Systems. *Journal of Systems Science and Complexity, 19*(1), 1-8. doi:10.1007/s11424-006-0001-z

Holland, J. H., & Miller, J. H. (1991). Artificial Adaptive Agents in Economic Theory. *The American Economic Review, 81*(2), 365-370.

Holland, J. L. (1959). A Theory of Vocational Choice. *Journal of Counseling Psychology, 6*(1), 35-45. doi:10.1037/h0040767

Holland, J. L. (1997). *Making Vocational Choices: A Theory of Vocational Personalities and Work Environments* (3rd ed.). Lutz, FL: Psychological Assessment Resources.

Hu, N. B. (1996). *Effects of College Students' Perceptions of Labor Market Variables and Conditions on Their Choice of Academic Majors.* AIR 1996 Annual Forum Paper.

Initiative for Science in Europe. (2012, Oct 23). *A Petition for the Attention of the EU Heads of State or Government.* Retrieved Jul 12, 2013, from No Cuts on Research: http://www.no-cuts-on-research.eu/index.php?file=petition.htm

International Labour Organisation. (2013a). *Global Employment Trends for Youth 2013: A generation at risk.* Geneva: International Labour Office. Retrieved from http://www.ilo.org/wcmsp5/groups/public/---dgreports/---dcomm/documents/publication/wcms_212423.pdf

International Labour Organisation. (2013b). Key Indicators of the Labour Market, 8th Edition.

International Microsimulation Association. (n.d.). *What is Microsimulation*. Retrieved May 5, 2011, from IMA: http://www.microsimulation.org/IMA/What%20is%20microsimulation.htm

Jackson, M. O. (2008). *Social and Economic Networks.* Princeton, NJ: Princeton University Press.

Jacobs, B., & van der Ploeg, F. (2006). Guide to reform of higher education: a European perspective. *Economic Policy, 21*(47), 535-592. doi:10.1111/j.1468-0327.2006.00166.x

Jauhiainen, S. (2011). Overeducation in the Finnish regional labour markets. *Papers in Regional Science, 90*(3), 573-588. doi:10.1111/j.1435-5957.2010.00334.x

Jensen, U., Gartner, H., & Rässler, S. (2010). Estimating German overqualification with stochastic earnings frontiers. *AStA Advances in Statistical Analysis, 94*(1), 33-51. doi:10.1007/s10182-009-0121-6

Johnson, G. J., & Johnson, W. R. (2000). Perceived Overqualification and Dimensions of Job Satisfaction: A Longitudinal Analysis. *The Journal of Psychology: Interdisciplinary and Applied, 134*(5), 537-555. doi:10.1080/00223980009598235

Kalleberg, A. L. (1977). Work Values and Job Rewards: A Theory of Job Satisfaction. *American Sociological Review, 42*(1), 124-143.

Kalvans, R., & Salietis, V. (2009). *Factors determining choice of higher education in business administration and information technology study programs in Latvia.* Bachelor Thesis, Stockholm School of Economics in Riga.

Karakaya, G., Plasman, R., & Rycx, F. (2007). Overeducation on the Belgian labour market: evaluation and analysis of the explanatory factors through two types of approaches. *Compare: A Journal of Comparative and International Education, 37*(4), 513-532. doi:10.1080/03057920701366317

Kiersztyn, A. (2013). Stuck in a mismatch? The persistence of overeducation during twenty years of the post-communist transition in Poland. *Economics of Education Review, 32*, 78-91. doi:10.1016/j.econedurev.2012.09.009

Kim, A., & Kim, K.-W. (2003). *Returns to Tertiary Education in Germany and the UK: Effects of Fields of Study and Gender.* Working Paper No. 62, Mannheimer Zentrum für Europäische Sozialforschung.

Kirman, A. P. (1989). The Intrinsic Limits of Modern Economic Theory: The Emperor has No Clothes. *The Economic Journal, 99*(395), 126-139.

Kirman, A. P. (1992). Whom or What Does the Representative Individual Represent? *Journal of Economic Perspectives, 6*(2), 117-136.

Korpi, T., & Tåhlin, M. (2009). Educational mismatch, wages, and wage growth: Overeducation in Sweden, 1974–2000. *Labour Economics, 16*(2), 183-193. doi:10.1016/j.labeco.2008.08.004

Krauth, B. V. (2004). A dynamic model of job networking and social influences on employment. *Journal of Economic Dynamics and Control, 28*(6), 1185-1204. doi:10.1016/S0165-1889(03)00079-4

Kuddo, A. (2009). *Employment Services and Active Labor Market Programs in Eastern European and Central Asian Countries.* Washington DC: World Bank.

Kumar, R., Novak, J., & Tomkins, A. (2010). Structure and Evolution of Online Social Networks. In P. S. Yu, J. Han, & C. Faloutsos (Eds.), *Link Mining: Models, Algorithms, and Applications* (pp. 337-357). New York: Springer. doi:10.1007/978-1-4419-6515-8_13

Kuncel, N. N., Credé, M., & Thomas, L. L. (2007). A Meta-Analysis of the Predictive Validity of the Graduate Management Admission Test (GMAT) and Undergraduate Grade Point Average (UGPA) for Graduate Student Academic Performance. *Academy of Management Learning & Education, 6*(1), 51–68. doi:10.5465/AMLE.2007.24401702

Kwak, H., Lee, C., Park, H., & Moon, S. (2010). What is Twitter, a social network or a news media? *Proceedings of the 19th international conference on World wide web* (pp. 591-600). New York, NY: ACM. doi:10.1145/1772690.1772751

Lange, T. (2012). Job satisfaction and self-employment: autonomy or personality? *Small Business Economics, 38*(2), 165-177. doi:10.1007/s11187-009-9249-8

Leijonhufvud, A. (2009). Limits to the equilibrating capabilities of market systems. *Journal of Economic Interaction and Coordination, 4*(2), 173-182. doi:10.1007/s11403-009-0052-z

Leonard, D., Becker, R., & Coate, K. (2005). To prove myself at the highest level: The benefits of doctoral study. *Higher Education Research & Development, 24*(2), 135–149. doi:10.1080/07294360500062904

Leuven, E., & Oosterbeek, H. (2011). Overeducation and Mismatch in the Labor Market. In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook of the Economics of Education* (Vol. 4, pp. 283-326). Amsterdam: Elsevier.

Levin, S. A. (1998). Ecosystems and the Biosphere as Complex Adaptive Systems. *Ecosystems, 1*(5), 431-436. doi:10.1007/s100219900037

Lewkovicz, Z., Thiriot, S., & Caillou, P. (2011). How detailed should social networks be for labor market's models? *ocial Networks and Multiagent Systems Symposium.* York, UK.

Lianos, T. P., Asteriou, D., & Agiomirgianakis, G. M. (2004). Foreign university graduates in the Greek labour market: Employment, salaries and overeducation. *International Journal of Finance & Economics, 9*(2), 151-164. doi:10.1002/ijfe.238

Liu, Q., & Pierce, D. A. (1994). A note on Gauss-Hermite quadrature. *Biometrika, 81*(3), 624-629. doi:10.1093/biomet/81.3.624

Lorscheid, I., Meyer, M., & Hocke, S. (2013). Simulation Model and Data Analysis: Where are we and where should we go? *ESSA 2013: 9th Conference of the European Social Simulation Association.* Warsaw, Poland.

Luger, G. F. (2009). *Artificial Intelligence: Structures and Strategies for Complex Problem Solving* (6th ed.). Boston, MA: Pearson Education.

Martin, C. W., & Neugart, M. (2009). Shocks and Endogenous Institutions: An Agent-based Model of Labor Market Performance in Turbulent Times. *Computational Economics, 33*(1), 31–46. doi:10.1007/s10614-008-9149-z

Mavromaras, K., McGuinness, S., O'Leary, N., Sloane, P., & Fok, Y. K. (2010). The Problem of Overskilling in Australia and Britain. *The Manchester School, 78*(3), 219-241. doi:10.1111/j.1467-9957.2009.02136.x

Mayda, A. (2010). International migration: A panel data analysis of the determinants of bilateral flows. *Journal of Population Economics, 23*(4), 1249-1274. doi:10.1007/s00148-009-0251-x

McCrae, R. R., & John, O. P. (1992). An Introduction to the Five-Factor Model and Its Applications. *Journal of Personality, 60*(2), 175-215. doi:10.1111/j.1467-6494.1992.tb00970.x

McGuiness, S., & Bennett, J. (2007). Overeducation in the graduate labour market: A quantile regression approach. *Economics of Education Review, 26*(5), 521-531. doi:10.1016/j.econedurev.2005.12.003

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology, 27*(1), 415-444.

Meagher, D. G., Lin, A., & Stellato, C. P. (2006). A Predictive Validity Study of the Pharmacy College Admission Test. *American Journal of Pharmaceutical Education, 70*(3), 53.

Mendes de Oliveira, M., Santos, M. C., & Kiker, B. F. (2000). The role of human capital and technological change in overeducation. *Economics of Education Review, 19*(2), 199-206. doi:10.1016/S0272-7757(99)00020-5

Michaelides, M. (2011). The effect of local ties, wages, and housing costs on migration decisions. *Journal of Socio-Economics, 40*(2), 132-140. doi:10.1016/j.socec.2011.01.010

Millán, J. M., Hessels, J., Thurik, R., & Aguado, R. (2011). Determinants of job satisfaction: a European comparison of self-employed and paid employees. *Small Business Economics*, 1-20. doi:10.1007/s11187-011-9380-1

Ministerio de Educación, Cultura y Deporte. (2012). *Datos y cifras del sistema universitario español.* Retrieved from http://www.mecd.gob.es/dms/mecd/educacion-mecd/areas-educacion/universidades/estadisticas-informes/estadisticas-informes-documentum/datos-cifras/2012-2013-datos-y-cifras-sistema-universitario-espanol.pdf

Mitsos, A., Bonaccorsi, A., Caloghirou, Y., Allmendinger, J., Georghiou, L., Mancini, M., & Sachwald, F. (2012). *High Level Panel on the Socio-Economic Benefits of the ERA.* Luxembourg: Publications Office of the European Union. Retrieved from http://ec.europa.eu/research/era/pdf/high-level-panel-report_en.pdf

Moniarou-Papaconstantinou, V., Tsatsaroni, A., Katsis, A., & Koulaidis, V. (2010). LIS as a field of study: socio-cultural influences on students' decision-making. *Aslib Proceedings, 62*(3), 321-344. doi:10.1108/00012531011046934

Montgomery, J. D. (1991). Social networks and labor-market outcomes: Toward an economic analysis. *American Economic Review, 81*(5), 1408-1418.

Montmarquette, C., Cannings, K., & Mahseredjian, S. (2002). How do young people choose college majors? *Economics of Education Review, 21*(6), 543-556. doi:10.1016/S0272-7757(01)00054-1

Mora, J.-G., García-Aracil, A., & Vila, L. E. (2007). Job satisfaction among young European higher education graduates. *Higher Education, 53*(1), 29-59. doi:10.1007/s10734-005-2377-4

Mortensen, J., & Vilella-Vila, M. (2012). The future of employment supply and demand in social Europe. *Futures, 44*(7), 671-677. doi:10.1016/j.futures.2012.04.006

Munasinghe, L., Reif, T., & Henriques, A. (2008). Gender gap in wage returns to job tenure and experience. *Labour Economics, 15*(6), 1296–1316. doi:10.1016/j.labeco.2007.12.003

Murillo, I. P., Rahona-López, M., & del Mar Salinas-Jiménez, M. (2012). Effects of educational mismatch on private returns to education: An analysis of the Spanish case (1995–2006). *Journal of Policy Modeling, 34*(5), 646-659. doi:10.1016/j.jpolmod.2011.07.012

Myck, M., & Paull, G. (2001). *The role of employment experience in explaining the gender wage gap.* Working Paper 01/18, Institute for Fiscal Studies. doi:10.1920/wp.ifs.2001.0118

Myklebust, J. P. (2010, Jan 24). Denmark: Fees part of broader reforms. *University World News*(108). Retrieved from http://www.universityworldnews.com/article.php?story=20100123091852321

Myklebust, J. P. (2011, Jan 30). Sweden: Fees deter foreign applicants. *University World News*(156). Retrieved from http://www.universityworldnews.com/article.php?story=20110128224609573

Myklebust, J. P., & Dobson, I. R. (2013, Feb 2). Move towards tuition fees for non-European students. *University World News*(257). Retrieved from http://www.universityworldnews.com/article.php?story=2013012712220418

Nadeau, R. (2011, Oct 18). *Neoclassical Economic Theory.* Retrieved May 28, 2012, from The Encyclopedia of Earth: http://www.eoearth.org/article/Neoclassical_economic_theory

Naylor, J. C., & Smith, A. F. (1982). Applications of a Method for the Efficient Computation of Posterior Distributions. *Journal of the Royal Statistical Society C, 31*(3), 214-225.

Neugart, M. (2008). Labor market policy evaluation with ACE. *Journal of Economic Behavior & Organization, 67*(2), 418–430. doi:10.1016/j.jebo.2006.12.006

Ng, E. S., Carpenter, J. R., Goldstein, H., & Rasbash, J. (2006). Estimation in generalised linear mixed models with binary outcomes by simulated maximum likelihood. *Statistical Modelling, 6*(1), 23-42. doi:10.1191/1471082X06st106oa

Ng, T. S., & Kang, Y. C. (2013). IEEE Transactions on Systems, Man, and Cybernetics: Systems. *The Inquiry of Labor Market Dynamics in Knowledge Economies: An Agent-Based Approach, 43*(2), 416–427. doi:10.1109/TSMCA.2012.2195167

Nikolai, C., & Madey, G. (2009). Tools of the Trade: A Survey of Various Agent Based Modeling Platforms. *Journal of Artificial Societies and Social Simulation, 12*(2).

Noble, J. P., & Sawyer, R. L. (2004). Is High School GPA Better than Admission Test Scores for Predicting Academic Success in College? *College and University, 79*(4), 17–22.

North, M. J., & Macal, C. M. (2007). *Managing Business Complexity: Discovering Strategic Solutions with Agent-Based Modeling and Simulation.* Oxford: Oxford University Press.

Norwegian Social Science Data Services. (2002). *European Social Survey Round 1 Data. Data File Edition 6.3.*

Norwegian Social Science Data Services. (2004). *European Social Survey Round 2 Data. Data File Edition 3.3.*

Norwegian Social Science Data Services. (2006). *European Social Survey Round 3 Data. Data File Edition 3.4.*

Norwegian Social Science Data Services. (2008). *European Social Survey Round 4 Data. Data File Edition 4.1.*

Norwegian Social Science Data Services. (2010). *European Social Survey Round 5 Data. Data File Edition 3.0.*

Ortiz, L., & Kucel, A. (2008). Do Fields of Study Matter for Over-education? The Cases of Spain and Germany. *International Journal of Comparative Sociology, 49*(4-5), 305-327. doi:10.1177/0020715208093079

Panitsidou, E., Griva, E., & Chostelidou, D. (2012). European Union Policies on Lifelong Learning: In-between Competitiveness Enhancement and Social Stability Reinforcement. *Procedia - Social and Behavioral Sciences, 46*, 548-553. doi:10.1016/j.sbspro.2012.05.158

Parry, J. (2008). Intention to leave the profession: antecedents and role in nurse turnover. *Journal of Advanced Nursing, 64*(2), 157-167. doi:10.1111/j.1365-2648.2008.04771.x

Pedersen, P. J., Pytlikova, M., & Smith, N. (2008). Selection and network effects - Migration flows into OECD countries 1990-2000. *European Economic Review, 52*(7), 1160-1186. doi:10.1016/j.euroecorev.2007.12.002

Peiró, J. M., Agut, S., & Grau, R. (2010). The Relationship Between Overeducation and Job Satisfaction Among Young Spanish Workers: The Role of Salary, Contract of Employment, and Work Experience. *Journal of Applied Social Psychology, 40*(3), 666-689. doi:10.1111/j.1559-1816.2010.00592.x

Pepe, M. S. (2003). *The Statistical Evaluation of Medical Tests for Classification and Prediction.* New York: Oxford University Press.

Pike, G. R. (2006). Students' Personality Types, Intended Majors, and College Expectations: Further Evidence Concerning Psychological and Sociological Interpretations of Holland's Theory. *Research in Higher Education, 47*(7), 801-822. doi:10.1007/s11162-006-9016-5

Pinheiro, J. C., & Bates, D. M. (1995). Approximations to the Log-Likelihood Function in the Nonlinear Mixed-Effects Model. *Journal of Computational and Graphical Statistics, 4*(1), 12-35. doi:10.1080/10618600.1995.10474663

Pinheiro, J. C., & Chao, E. C. (2006). Efficient Laplacian and Adaptive Gaussian Quadrature Algorithms for Multilevel Generalized Linear Mixed Models. *Journal of Computational and Graphical Statics, 15*(1), 58-81. doi:10.1198/106186006X96962

Piracha, M., Tani, M., & Vadean, F. (2012). Immigrant over- and under-education: the role of home country labour market experience. *IZA Journal of Migration, 1*(3). doi:10.1186/2193-9039-1-3

Poggi, A. (2010). Job satisfaction, working conditions and aspirations. *Journal of Economic Psychology, 31*(6), 936-949. doi:10.1016/j.joep.2010.08.003

Pomazan, V., Mihalaşcu, D., Petcu, L. C., & Gîrtu, M. A. (2010). Enrolling in Science and Engineering Academic Programs—Motivating and Deterring Factors. *AIP Conference Proceedings* (pp. 1372-1377). American Institute of Physics. doi:10.1063/1.3322374

Quintini, G. (2011). *Over-Qualified or Under-Skilled: A Review of Existing Literature.* OECD Social, Employment and Migration Working Papers, No. 121, OECD Publishing. doi:10.1787/5kg58j9d7b6d-en

Rabe-Hesketh, S., & Skrondal, A. (2012). *Multilevel and Longitudinal Modeling Using Stata* (3rd ed., Vol. II). College Station, Texas: Stata Press.

Railsback, S. F., Lytinen, S. L., & Jackson, S. K. (2006). Agent-based Simulation Platforms: Review and Development Recommendations. *SIMULATION, 82*(9), 609–623. doi:10.1177/0037549706073695

Ralston, C. A., Borgen, F. H., Rottinghaus, P. J., & Donnay, D. A. (2004). Specificity in interest measurement: Basic Interest Scales and major field of study. *Journal of Vocational Behavior, 66*(2), 203-216. doi:10.1016/S0001-8791(03)00096-4

Ramos, R., & Sanromá, E. (2013). Overeducation and Local Labour Markets in Spain. *Tijdschrift voor economische en sociale geografie, 104*(3), 278-291. doi:10.1111/j.1467-9663.2012.00752.x

Rectors' Conference of the Swiss Universities. (2013, Jun 14). *University Education.* Retrieved Jul 12, 2013, from Study in Switzerland: http://www.crus.ch/information-programme/study-in-switzerland.html?L=2#8_Costs

Reimer, D., & Pollak, R. (2010). Educational Expansion and Its Consequences for Vertical and Horizontal Inequalities in Access to Higher Education in West Germany. *European Sociological Review, 26*(4), 415-430. doi:10.1093/esr/jcp029

Richiardi, M. (2006). Toward a Non-Equilibrium Unemployment Theory. *Computational Economics, 27*(1), 135–160. doi:10.1007/s10614-005-9019-x

Richiardi, M., Leombruni, R., Saam, N., & Sonnessa, M. (2006). A Common Protocol for Agent-Based Social Simulation. *Journal of Artificial Societies and Social Simulation, 9*(1), 15.

Rieckmann, M. (2012). Future-oriented higher education: Which key competencies should be fostered through university teaching and learning? *Futures, 44*(2), 127-135. doi:10.1016/j.futures.2011.09.005

Robbins, S. P., Judge, T. A., & Campbell, T. (2010). *Organizational Behaviour.* Harlow, England: Pearson Education Limited.

Robst, J. (2008). Overeducation and College Major: Expanding the Definition of Mismatch between Schooling and Jobs. *The Manchester School, 76*(4), 349-368. doi:10.1111/j.1467-9957.2008.01064.x

Rogers, W. H. (1993). sg17: Regression standard errors in clustered samples. *Stata Technical Bulletin*(13), 19-23. Retrieved from http://www.stata.com/products/stb/journals/stb13.pdf

Romanian Bologna Secretariat. (n.d.). *History of EHEA.* Retrieved Jul 5, 2013, from European Higher Education Area: http://www.ehea.info/article-details.aspx?ArticleId=3

Roodman, D. (2011). Estimating fully observed recursive mixed-process models with cmp. *Stata Journal, 11*(2), 159-206.

Roxburgh, S. (1999). Exploring the Work and Family Relationship: Gender Differences in the Influence of Parenthood and Social Support on Job Satisfaction. *Journal of Family Issues, 20*(6), 771-788. doi:10.1177/019251399020006003

Rubb, S. (2003a). Overeducation in the labor market: a comment and re-analysis of a meta-analysis. *Economics of Education Review, 22*(6), 621-629. doi:10.1016/S0272-7757(02)00077-8

Rubb, S. (2003b). Overeducation: a short or long run phenomenon for individuals? *Economics of Education Review, 22*(4), 389-394. doi:10.1016/S0272-7757(02)00052-3

Rubinstein, G. (2005). The big five among male and female students of different faculties. *Personality and Individual Differences, 38*(7), 1495-1503. doi:10.1016/j.paid.2004.09.012

Sakoda, J. M. (1971). The checkerboard model of social interaction. *The Journal of Mathematical Sociology, 1*(1), 119–132. doi:10.1080/0022250X.1971.9989791

Sánchez-Sánchez, N., & McGuiness, S. (2015). Decomposing the Impacts of Overeducation and Overskilling on Earnings and Job Satisfaction: An Analysis Using REFLEX data. *Education Economics, 23*(4), 419-432. doi:10.1080/09645292.2013.846297

Sattinger, M. (1993). Assignment Models of the Distribution of Earnings. *Journal of Economic Literature, 31*(2), 831-880.

Schelling, T. C. (1971). Dynamic models of segregation. *The Journal of Mathematical Sociology, 1*(2), 143–186. doi:10.1080/0022250X.1971.9989794

Schwarz, G. (1978). Estimating the Dimension of a Model. *Annals of Statistics, 6*(2), 461-464. doi:10.1214/aos/1176344136

Sherwood, D. (2002). *Seeing the Forest for the Trees: A Manager's Guide to Applying Systems Thinking.* London: Nicholas Brealey Publishing.

Sicherman, N. (1991). "Overeducation" in the Labor Market. *Journal of Labor Economics, 9*(2), 101-122.

Silva, S. T., Valente, J. M., & Teixeira, A. A. (2012). An evolutionary model of industry dynamics and firms' institutional behavior with job search, bargaining and matching. *Journal of Economic Interaction and Coordination, 7*(1), 23–61. doi:10.1007/s11403-011-0085-y

Simões, C., & Soares, A. (2010). Applying to higher education: information sources and choice factors. *Studies in Higher Education, 35*(4), 371-389. doi:10.1080/03075070903096490

Simpson, P. V. (2011, Apr 1). Foreign students turn their backs on Sweden. *The Local*. Retrieved from http://www.thelocal.se/32954/20110401/

Skalli, A., Theodossiou, I., & Vasileiou, E. (2008). Jobs as Lancaster goods: Facets of job satisfaction and overall job satisfaction. *Journal of Socio-Economics, 37*(5), 1906-1920. doi:10.1016/j.socec.2008.04.003

Smart, J. C., Feldman, K. A., & Ethington, C. A. (2000). *Academic Discplines: Holland's Theory and the Study of College Students and Faculty.* Nashville, TN: Vanderbilt University Press.

Spence, M. (1973). Job Market Signaling. *Quarterly Journal of Economics, 87*(3), 355-374. doi:10.2307/1882010

StataCorp. (2011a). *Stata Base Reference Manual. Release 12.* College Station, Texas: Stata Press.

StataCorp. (2011b). *Stata Longitudinal-Data/Panel-Data Reference Manual. Release 12.* College Station, Texas: Stata Press.

StataCorp. (2011c). *Stata User's Guide. Release 12.* College Station, Texas: Stata Press.

Stigler, G. J. (1961). The Economics of Information. *Journal of Political Economy, 69*(3), 213-225.

Stigler, G. J. (1962). Information in the Labor Market. *Journal of Political Economy, 70*(5), 94-105.

Støren, L. A., & Wiers-Jenssen, J. (2010). Foreign Diploma Versus Immigrant Background: Determinants of Labour Market Success or Failure? *Journal of Studies in International Education, 14*(1), 29-49. doi:10.1177/1028315308327951

Sursock, A., & Smidt, H. (2010). *Trends 2010: A decade of change in European Higher Education.* Brussels: European University Association. Retrieved from http://www.eua.be/eua-work-and-policy-area/building-the-european-higher-education-area/trends-in-european-higher-education/trends-vi.aspx

Sutherland, J. (2012). Qualifications mismatch and skills mismatch. *Education + Training, 54*(7), 619-632. doi:10.1108/00400911211265666

System Dynamics Society. (n.d.). *What is SD.* Retrieved Jul 18, 2013, from http://www.systemdynamics.org/what-is-s/

Takács, K., Squazzoni, F., & Bravo, G. (2012). The Network Antidote: An Agent-Based Model of Discrimination in Labor Markets. *MKE 2012 Conference.* Budapest.

Tamtam, A. G., Gallagher, F., Olabi, A. G., & Naher, S. (2012). A Comparative Study of the Implementation of EMI in Europe, Asia and Africa. *Procedia - Social and Behavioral Sciences, 47*, 1417-1425. doi:10.1016/j.sbspro.2012.06.836

Tani, M. (2012). Does immigration policy affect the education-occupation mismatch? Evidence from Australia. *Australian Bulletin of Labour, 38*(2), 111-140.

Tarvid, A. (2011). Job Satisfaction Modelling in Agent-Based Simulations. In A. Bruzzone, M. A. Piera, F. Longo, P. Elfrey, M. Affenzeller, & O. Balci (Ed.), *The 23rd European Modeling & Simulation Symposium* (pp. 158-165). Rende, Italy: DIPTEM Università di Genova.

Tarvid, A. (2014). Motivation to Study for PhD Degree: Case of Latvia. *Procedia Economics and Finance, 14*, 585-594. doi:10.1016/S2212-5671(14)00747-3

Tassier, T., & Menczer, F. (2001). Emerging small-world referral networks in evolutionary labor markets. *IEEE Transactions on Evolutionary Computation, 5*(5), 482-492. doi:10.1109/4235.956712

Tassier, T., & Menczer, F. (2008). Social network structure, segregation, and equality in a labor market with referral hiring. *Journal of Economic Behavior & Organization, 66*(3-4), 514-528. doi:10.1016/j.jebo.2006.07.003

Tesfatsion, L. (2001). Structure, behavior, and market power in an evolutionary labor market with adaptive search. *Journal of Economic Dynamics and Control, 25*(3-4), 419-457. doi:10.1016/S0165-1889(00)00032-4

The Brookings Institution. (2011, Dec 8). *Agent-based Models.* Retrieved Dec 8, 2011, from Brookings: http://www.brookings.edu/topics/agent-based-models.aspx

The Economist. (2010, Jul 22). *Agents of change.* Retrieved Dec 8, 2011, from The Economist: Markets & Data: http://www.economist.com/node/16636121

Thiriot, S., Lewkovicz, Z., Caillou, P., & Kant, J.-D. (2011). Referral Hiring and Labor Markets: A Computational Study. In S. Osinga, G. J. Hofstede, & T. Verwaart (Eds.), *Emergent Results of Artificial Economics* (pp. 15–25). Berlin: Springer. doi:10.1007/978-3-642-21108-9_2

Thurow, L. C. (1975). *Generating Inequality.* New York: Basic Books.

Tierney, L., & Kadane, J. B. (1986). Accurate Approximations for Posterior Moments and Marginal Densities. *Journal of the American Statistical Association, 81*(393), 82-86. doi:10.1080/01621459.1986.10478240

Tinbergen, J. (1956). On the Theory of Income Distribution. *Weltwirtschaftliches Archiv, 77*, 155-175.

Tobin, J. (1958). Estimation of Relationships for Limited Dependent Variables. *Econometrica, 26*(1), 24-36.

Toivonen, R., Kovanen, L., Kivelä, M., Onnela, J.-P., Saramäki, J., & Kaski, K. (2009). A comparative study of social network models: Network evolution models and nodal attribute models. *Social Networks, 31*(4), 240-254. doi:10.1016/j.socnet.2009.06.004

Tolsma, J., Need, A., & De Jong, U. (2010). Explaining Participation Differentials in Dutch Higher Education: The Impact of Subjective Success Probabilities on Level Choice and Field Choice. *European Sociological Review, 26*(2), 235-252. doi:10.1093/esr/jcp061

Tsai, Y. (2010). Returns to overeducation: A longitudinal analysis of the U.S. labor market. *Economics of Education Review, 29*(4), 606-617. doi:10.1016/j.econedurev.2010.01.001

Tuijnman, A. C. (2003). A "Nordic model" of adult education: What might be its defining parameters? *International Journal of Educational Research, 39*(3), 283-291. doi:10.1016/j.ijer.2004.04.008

UCLA Academic Technology Services. (2011, Oct 20). *FAQ: What are pseudo R-squareds?* Retrieved Aug 15, 2012, from http://www.ats.ucla.edu/stat/mult_pkg/faq/general/Psuedo_RSquareds.htm

Valtonen, M., & Lampinen, M. (2013, Jan 7). *SAMOK and SYL: Finland moves to the era of tuition fees – shuts doors for international professionals.* Retrieved from SAMOK: http://samok.fi/en/2013/01/07/kannanotto-syl-ja-samok-suomi-lukukausimaksujen-aikaan-ovet-kiinni-kansainvalisille-osaajille/

Van de Werfhorst, H. G., De Graaf, N. D., & Kraaykamp, G. (2001). Intergenerational Resemblance in Field of Study in the Netherlands. *European Sociological Review, 17*(3), 275-293. doi:10.1093/esr/17.3.275

Van de Werfhorst, H. G., Sullivan, A., & Cheung, S. Y. (2003). Social Class, Ability and Choice of Subject in Secondary and Tertiary Education in Britain. *British Educational Research Journal, 29*(1), 41-62. doi:10.1080/0141192032000057366

Van Deuren, R., & Santema, S. C. (2012). *How to choose your minor? Decision making variables used in the selection of minor by undergraduate students from a Dutch university of applied sciences.* Working Paper No. 2012/06, Maastricht School of Management.

Veiga, A., Amaral, A., & Mendes, A. (2008). Implementing Bologna in Southern European countries: Comparative analysis of some research findings. *Education for Chemical Engineers, 3*(1), e47-e56. doi:10.1016/j.ece.2008.01.004

Verdugo, R. R., & Verdugo, N. T. (1989). The Impact of Surplus Schooling on Earnings: Some Additional Findings. *The Journal of Human Resources, 24*(4), 629-643.

Verhaest, D., & Omey, E. (2010). The determinants of overeducation: different measures, different outcomes? *International Journal of Manpower, 31*(6), 608-625. doi:10.1108/01437721011073337

Verhaest, D., & Omey, E. (2012). Overeducation, Undereducation and Earnings: Further Evidence on the Importance of Ability and Measurement Error Bias. *Journal of Labor Research, 33*(1), 76-90. doi:10.1007/s12122-011-9125-6

Verhaest, D., & Van der Velden, R. (2013). Cross-country Differences in Graduate Overeducation. *European Sociological Review, 29*(3), 642-653. doi:10.1093/esr/jcs044

Vila, L. E., García-Aracil, A., & Mora, J.-G. (2007). The Distribution of Job Satisfaction among Young European Graduates: Does the Choice of Study Field Matter? *The Journal of Higher Education, 78*(1), 97-118.

Vossensteyn, J. J. (2005). *Perceptions of Student Price-Responsiveness: a behavioural economics exploration of the relationships between socio-economic status, perceptions of financial incentives and student choice.* University of Twente. Retrieved from http://www.utwente.nl/cheps/documenten/thesisvossensteynnl.pdf

Wald, S. (2005). The impact of overqualification on job search. *International Journal of Manpower, 26*(2), 140-156. doi:10.1108/01437720510597649

Weimer, L. (2011, Apr 11). *Thematic Week: The Introduction of Tuition Fees in the Nordic Countries.* Retrieved from Higher Education Development Association: http://uv-net.uio.no/wpmu/hedda/2011/04/11/thematic-week-the-introduction-of-tuition-fees-in-the-nordic-countries/

Wildberger, A. M. (1997). Complex adaptive systems: concepts and power industry applications. *IEEE Control Systems Magazine, 17*(6), 77-88. doi:10.1109/37.642976

Williams, N. (2009). Seniority, experience, and wages in the UK. *Labour Economics, 16*(3), 272-283. doi:10.1016/j.labeco.2008.09.003

Williams, R. (2006). Generalized ordered logit/partial proportional odds models for ordinal dependent variables. *Stata Journal, 6*(1), 58-82.

Williams, R. L. (2000). A Note on Robust Variance Estimation for Cluster-Correlated Data. *Biometrics, 56*(2), 645-646. doi:10.1111/j.0006-341X.2000.00645.x

Wilson, C., Boe, B., Sala, A., Puttaswamy, K. P., & Zhao, B. Y. (2009). User interactions in social networks and their implications. *Proceedings of the 4th ACM European conference on Computer systems* (pp. 205-218). New York, NY: ACM. doi:10.1145/1519065.1519089

Wirz, A., & Atukeren, E. (2005). An analysis of perceived overqualification in the Swiss labor market. *Economics Bulletin, 9*(2), 1-10.

Wolfe, R., & Gould, W. (1998). sg76: An approximate likelihood-ratio test for ordinal response models. *Stata Technical Bulletin*, 24-27.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data* (2nd ed.). Cambridge, MA: The MIT Press.

World Economic Forum. (2014). *Global Risks 2014: Ninth Edition.* Insight Report, Geneva.

Zafar, B. (2012). Double Majors: One for Me, One for the Parents? *Economic Inquiry, 50*(2), 287-308. doi:10.1111/j.1465-7295.2011.00403.x

Zheng, H., Son, Y.-J., Chiu, Y.-C., Head, L., Feng, Y., Xi, H., . . . Hickman, M. (2013). *A Primer for Agent-Based Simulation and Modeling in Transportation Applications.* U.S. Department of Transportation.

Zimmerman, R. D. (2008). Understanding the impact of personality traits on individuals' turnover decisions: A meta-analytic path model. *Personnel Psychology, 61*(2), 309-348. doi:10.1111/j.1744-6570.2008.00115.x

Колмогоров, А. Н., & Фомин, С. В. (1976). *Элементы теории функций и функционального анализа.* Москва: Наука.

APPENDIX A TRANSFORMATION OF EDUCATION VARIABLES IN THE ESS DATABASE

For the measures of mismatch used in this dissertation, it is crucial that education levels in a given country are defined consistently over time; i.e., it should not be the case that a narrowly defined level of education is assigned to secondary level in one round and to tertiary level in another. Otherwise, it is impossible to separate the true change in mismatch from the one induced by the change in the definition of an education level.

ESS data do face problems with the consistency of education variables. In preparing the data, I tried to minimise these problems by looking on changes in the shares of each of the three major education levels in the whole database (i.e., not including labour market status). These levels are primary (ISCED 1–2, excluding those without primary education), secondary and post-secondary non-tertiary (ISCED 3–4), and tertiary (ISCED 5–6, including 5A short-cycle and 5B programmes). There might be two reasons of large changes in these shares: (1) changing the share of holders of a particular degree in the sample, perhaps because they become more important for analysis than the holders of other degrees, and (2) re-assigning a narrow education level to a different broader level. In several cases, data for rounds 1–4 are available only in the aggregate form, with respondents assigned to quite broad levels, not indicating which minor levels form a given broad level, so it was impossible to find out which of the two reasons caused the change. In other cases, I was able to compare the mappings and change them if it was necessary.

In particular, I made the following changes to the default mappings from narrow education levels to the three major education levels. Those without primary education were removed from the analysis (originally, they have been mapped to “below secondary education”); this considerably decreased sample size in Spain, the UK and Portugal. In Estonia, rounds 2–4, professional education after secondary was re-mapped to secondary. In Spain, rounds 1–4, *F.P. 1* was re-mapped to primary level. In the UK, the current education level variables lead to a drastically different distribution of individuals across the three major levels in rounds 1, 3 and 4 (as compared to rounds 2 and 5); education variables from previous editions of ESS data[56] were instead taken for these three rounds, leading to a smoother distribution dynamics over time. In Greece, round 4, post-compulsory secondary was re-mapped to secondary. In Israel, rounds 4 and 5, secondary school or *yeshiva* without matriculation certificate was re-mapped to primary education. In the same country, round 5, part of ISCED 5B that was originally classified to ISCED 4A/4B was re-mapped to higher education. In Poland, rounds 1–4, basic vocational education (*zasadnicze zawodowe*) was re-mapped to primary. Finally, those who reported higher education but were aged under 20 were removed from the analysis. The resulting distribution of the three education levels by country and ESS round is shown in Table A-1.

[56] The variable ‘edulvl’ from previous editions was taken instead of the new ‘edulvla’ variable.

Table A-1 – Shares of Primary, Secondary and Tertiary Graduates in the ESS Database

	Round 1			Round 2			Round 3			Round 4			Round 5		
	Primary	Secondary	Tertiary	Primary	Secondary	Tertiary	Primary	Secondary	Tertiary	Primary	Secondary	Tertiary	Primary	Secondary	Tertiary
Austria	29.4	58.3	12.4	28.6	64.4	6.9	22.7	70.1	7.2	17.3	69.3	13.4			
Belgium	34.0	38.1	27.9	32.9	39.3	27.8	35.2	36.7	28.0	31.6	37.6	30.8	29.1	38.4	32.5
Bulgaria							31.3	48.2	20.5	30.5	50.1	19.4	30.0	48.1	21.9
Switzerl.	23.6	50.5	25.9	22.1	52.0	25.8	25.7	51.4	22.9	23.5	52.8	23.6	22.6	47.7	29.7
Cyprus							26.0	47.5	26.6	26.2	43.4	30.4	34.8	38.8	26.5
Czech Rep.	13.9	72.2	13.9	17.2	72.8	10.0				13.3	74.6	12.1	14.6	74.6	10.8
Germany	16.2	59.1	24.6	17.3	58.5	24.2	15.0	63.4	21.6	14.4	57.4	28.3	18.6	54.3	27.1
Denmark	24.7	46.7	28.6	23.7	36.9	39.4	20.9	36.4	42.7	21.4	38.5	40.2	29.3	32.5	38.2
Estonia				25.4	57.6	17.0	26.0	52.8	21.2	26.9	51.8	21.2	22.4	50.5	27.1
Spain	56.4	24.6	19.0	52.9	21.9	25.2	55.1	23.7	21.2	57.0	24.1	18.9	50.3	22.0	27.7
Finland	39.0	35.6	25.4	35.3	36.0	28.8	32.2	36.7	31.1	30.5	38.4	31.1	28.8	36.7	34.5
France	27.7	44.0	28.3	27.9	44.2	27.8	24.0	47.8	28.2	32.2	37.8	30.0	27.7	45.7	26.6
UK	55.1	22.2	22.7	44.0	25.0	31.0	47.4	22.7	29.9	42.9	25.4	31.7	40.6	28.9	30.5
Greece	52.4	32.5	15.0	50.0	31.3	18.7				37.3	46.4	16.2	34.4	47.2	18.4
Croatia										14.8	61.9	23.4	24.0	57.4	18.6
Hungary	37.5	48.9	13.6	24.0	53.8	22.2	33.9	50.8	15.3	33.1	52.5	14.3	23.7	58.7	17.7
Ireland	45.6	23.9	30.5	41.0	28.0	30.9	34.4	24.0	41.6	34.7	24.2	41.1	34.3	37.0	28.6
Israel	34.4	28.3	37.3							34.1	23.2	42.7	35.9	25.9	38.2
Iceland				34.6	31.8	33.6									
Italy	55.4	35.7	8.9	51.3	37.0	11.7									
Lithuania										26.9	30.8	42.2	20.2	40.7	39.1
Luxemb.	41.8	39.4	18.7	42.0	40.2	17.8									
Latvia							31.1	49.4	19.5	24.0	51.5	24.5			
Netherl.	44.7	33.6	21.7	43.7	32.4	23.9	40.7	35.0	24.3	38.4	35.5	26.2	45.1	24.7	30.2
Norway	14.8	57.1	28.2	22.0	43.0	35.1	18.2	45.7	36.1	14.8	48.5	36.7	21.4	38.4	40.1
Poland	55.6	33.2	11.2	54.3	34.7	11.0	56.0	32.6	11.5	48.0	34.7	17.3	43.5	35.3	21.1
Portugal	73.2	16.4	10.4	72.3	17.8	9.9	70.8	17.9	11.3	72.0	17.0	11.1	74.0	16.3	9.7
Romania							32.7	55.7	11.6	33.5	52.7	13.8			
Russia							19.2	33.2	47.6	15.7	32.1	52.2	12.5	35.1	52.5
Sweden	33.5	44.2	22.3	34.1	42.8	23.1	29.7	44.0	26.4	25.4	47.6	27.0	24.2	45.0	30.8
Slovenia	26.4	58.0	15.6	26.9	57.0	16.1	25.5	54.1	20.5	26.1	52.4	21.5	23.4	55.5	21.1
Slovakia				20.1	68.9	11.0	19.3	68.5	12.2	11.7	69.4	18.9	13.1	63.8	23.1
Turkey				70.7	21.8	7.5				70.9	21.1	8.0			
Ukraine				19.1	29.4	51.6	18.0	28.2	53.8	13.5	30.9	55.6	10.0	31.3	58.7

Those who did not finish primary education are removed. The table refers to all respondents holding a primary, secondary or tertiary education level, disregarding their status in the labour market.

APPENDIX B ADDITIONAL DESCRIPTIVE STATISTICS FOR MISMATCH

Table B-1 – Countries with Extreme Mismatch Incidence for Non-natives, Disabled and Living outside Big Cities, Total Sample

	Minority vs. Fully Native	Parent-immigrant vs. Fully Native	Both Parents Immigrants vs. Fully Native	CEE Immigrants vs. Fully Native	Former Soviet Union Immigrant vs. Fully Native	Latin America, Africa and Asia Immigrant vs. Fully Native	Non-CEE European Immigrant vs. Fully Native	Immigrant from Other Developed Countries vs. Fully Native	Immigrant, Any Origin Country, vs. Fully Native	Disabled vs. Not Disabled	Towns or Small Cities vs. Big Cities	Rural Areas vs. Big Cities
Austria				OU		U	U		U			
Belgium				O					O			
Bulgaria												
Croatia			U								U	
Cyprus				O			OU		OU			
Czech Rep.			U									U
Denmark				O		O			O			
Estonia	O			OU		U			OU			
Finland				O			U		OU			
France			U			OU			U			
Germany				O	O							
Greece				O	O		U		O			
Hungary				OU					U			
Iceland											U	
Ireland				O		OU	U		OU			
Israel				OU	O				O			
Italy		U										
Latvia												
Lithuania			O								O	O
Luxembourg				OU		OU	U		U			
Netherlands				U			OU		U			
Norway	O			OU		O	U	O	OU			
Poland									U			
Portugal	U	U		OU		U	OU		OU			
Romania						O			OU			
Russia			O		O				O			
Slovakia			U	OU					OU			
Slovenia	U											
Spain		U		OU		OU	U		OU			
Sweden			U	OU		OU	U		OU			
Switzerland				O								
Turkey										U	O	
UK		U		OU		OU	U		OU			
Ukraine			O	O	O				O			O

'O' marks the countries with too high overeducation incidence by the ISCO criterion. 'U' marks the countries with too low undereducation incidence by the ISCO criterion. Incidence in each category is compared with the respective reference category (indicated in column header). The incidence in a category in a given country is considered too high (low) if it is higher (lower) than the incidence in the reference category in that country for the age group under consideration by more than 0.05. For each country, data from all available rounds are pooled. Greyed cells denote cases where the number of observations is less than 15, so no meaningful conclusion could be made.

Table B-2 – Countries with Extreme Mismatch Incidence for Non-natives, Disabled and Living outside Big Cities, Youth Subsample (15–29)

	Minority vs. Fully Native	Parent-immigrant vs. Fully Native	Both Parents Immigrants vs. Fully Native	CEE Immigrants vs. Fully Native	Former Soviet Union Immigrant vs. Fully Native	Latin America, Africa and Asia Immigrant vs. Fully Native	Non-CEE European Immigrant vs. Fully Native	Immigrant from Other Developed Countries vs. Fully Native	Immigrant, Any Origin Country, vs. Fully Native	Disabled vs. Not Disabled	Towns or Small Cities vs. Big Cities	Rural Areas vs. Big Cities
Austria	U			U		U			U			
Belgium							OU					
Bulgaria												
Croatia											U	U
Cyprus									OU			
Czech Rep.										U	U	U
Denmark												
Estonia	U	O	U	U					U			
Finland							OU		O	O		
France			U			OU			OU		O	
Germany						U						
Greece				O			O					
Hungary										U		U
Iceland												
Ireland				OU		OU	U		OU	O		
Israel			O	OU	O	O			OU		O	
Italy												
Latvia		O	U							OU		
Lithuania	OU	U								U		
Luxembourg		U					U		U			
Netherlands							U		U			
Norway				O		O	OU		O			
Poland										O		
Portugal	U	U				OU	U		OU			
Romania												
Russia		U			O				U			
Slovakia										U		
Slovenia												U
Spain				OU		OU			OU		O	
Sweden			U	O					O		U	U
Switzerland			O				U			U		
Turkey										U		
UK	O			O		OU			OU			
Ukraine			O									O

'O' marks the countries with too high overeducation incidence by the ISCO criterion. 'U' marks the countries with too low undereducation incidence by the ISCO criterion. Incidence in each category is compared with the respective reference category (indicated in column header). The incidence in a category in a given country is considered too high (low) if it is higher (lower) than the incidence in the reference category in that country for the age group under consideration by more than 0.05. For each country, data from all available rounds are pooled. Greyed cells denote cases where the number of observations is less than 15, so no meaningful conclusion could be made.

Table B-3 – Countries with Extreme Mismatch Incidence for Non-natives, Disabled and Living outside Big Cities, Mature Subsample (30+)

	Minority vs. Fully Native	Parent-immigrant vs. Fully Native	Both Parents Immigrants vs. Fully Native	CEE Immigrants vs. Fully Native	Former Soviet Union Immigrant vs. Fully Native	Latin America, Africa and Asia Immigrant vs. Fully Native	Non-CEE European Immigrant vs. Fully Native	Immigrant from Other Developed Countries vs. Fully Native	Immigrant, Any Origin Country, vs. Fully Native	Disabled vs. Not Disabled	Towns or Small Cities vs. Big Cities	Rural Areas vs. Big Cities
Austria				OU		OU	U		U			
Belgium	O			O		O			O			
Bulgaria												
Croatia	O		U								U	
Cyprus				OU			OU		OU			
Czech Rep.												O
Denmark				OU		O			O			
Estonia	O			OU					OU			
Finland				O		U	U		OU			
France			U			OU	U		U			
Germany				OU	O				O			
Greece				OU	O		U		OU			
Hungary				OU					U			
Iceland											U	
Ireland		U		OU		OU	U	U	OU			
Israel				OU	O			U	O	O		
Italy									U			
Latvia									O			
Lithuania		O	OU	U							O	O
Luxembourg			U	U		OU	U		U			
Netherlands			U	U			OU		U			
Norway				OU		OU	U	OU	OU			
Poland									U			
Portugal	U			OU		U	OU		OU			
Romania						O			O			
Russia			O	O					O			
Slovakia			U	OU					OU			
Slovenia	U											
Spain		U		OU		OU	U		OU			
Sweden	U	U	U	OU		OU	U		OU			
Switzerland				O		U		U				
Turkey											O	
UK		U		U		U	U	U	U			
Ukraine		O	O	O	O				O			

'O' marks the countries with too high overeducation incidence by the ISCO criterion. 'U' marks the countries with too low undereducation incidence by the ISCO criterion. Incidence in each category is compared with the respective reference category (indicated in column header). The incidence in a category in a given country is considered too high (low) if it is higher (lower) than the incidence in the reference category in that country for the age group under consideration by more than 0.05. For each country, data from all available rounds are pooled. Greyed cells denote cases where the number of observations is less than 15, so no meaningful conclusion could be made.

APPENDIX C DISCRETE CHOICE MODELLING METHODOLOGY

This chapter discusses the methods used in Chapters 3 and 4. Throughout this section, I denote vectors with bold lower-case letters, matrices with bold upper-case letters and scalars in normal face. The dependent variable typically is y, and the vector of independent variables is $\mathbf{x}$. Models will have p parameters, typically indexed by j, using the dataset of N observations, typically indexed by i. Coefficients of the true model are denoted by β_j, while the estimated coefficients are denoted by $\hat{\beta}_j$. I use the prime notation (′) for vector/matrix transposition. As the empirical part of this work uses Stata® as statistical software, some peculiarities of its implementation of the described methods will be noted.

C.1 Maximum Likelihood Estimation

Because all models discussed in this section are estimated by maximum likelihood (ML), or variants thereof, I first provide a brief description of this estimation method. I base this description on Wooldridge (2010, pp. 473–475).

Denote by $f_0(\mathbf{y}|\mathbf{x})$ the true conditional density of $\mathbf{y} \in \mathcal{Y}$ given $\mathbf{x} \in \mathcal{X}$ with respect to a σ-finite measure[57] $\nu(d\mathbf{y})$. In the discrete case, it would be interpreted as the true probability of observing the data vector $\mathbf{y}$, given $\mathbf{x}$. It can be shown that for any function $f(\cdot\,|\mathbf{x})$ that is a conditional density function (i.e., $\int_{\mathcal{Y}} f(\mathbf{y}|\mathbf{x})\,\nu(d\mathbf{y}) = 1$ for all $\mathbf{x} \in \mathcal{X}$), the ***conditional Kullback–Leibler information inequality*** holds for all $\mathbf{x} \in \mathcal{X}$:

$$\mathcal{K}(f, \mathbf{x}) \equiv \int_{\mathcal{Y}} \ln\left[\frac{f_0(\mathbf{y}|\mathbf{x})}{f(\mathbf{y}|\mathbf{x})}\right] f_0(\mathbf{y}|\mathbf{x})\nu(d\mathbf{y}) \geq 0. \qquad \text{(C-1)}$$

Because $\mathcal{K}(f_0, \mathbf{x}) \equiv 0$, the integral above says that $\mathcal{K}(f, \mathbf{x})$ is minimised at $f = f_0$. For each observation i, by definition of conditional expectation, we can re-write (C-1) as

$$\mathrm{E}\{\ln f_0(\mathbf{y}_i|\mathbf{x}_i)\,|\mathbf{x}_i\} - \mathrm{E}\{\ln f(\mathbf{y}_i|\mathbf{x}_i)\,|\mathbf{x}_i\} \geq 0,$$

which yields

$$\mathrm{E}\{\ln f_0(\mathbf{y}_i|\mathbf{x}_i)\,|\mathbf{x}_i\} \geq \mathrm{E}\{\ln f(\mathbf{y}_i|\mathbf{x}_i)\,|\mathbf{x}_i\}.$$

Assume now that in addition to $\mathbf{x}$, the observed vector $\mathbf{y}$ is conditioned on parameters $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, so that $f(\cdot\,|\mathbf{x}, \boldsymbol{\theta})$ is also a conditional density function for all $\mathbf{x} \in \mathcal{X}, \boldsymbol{\theta} \in \boldsymbol{\Theta}$. We would like to find such parameters $\boldsymbol{\theta}_0$ that $f(\cdot\,|\mathbf{x}, \boldsymbol{\theta}_0) = f_0(\cdot\,|\mathbf{x})$ for all $\mathbf{x} \in \mathcal{X}$. If it is possible, the underlying parametric model is called a ***correctly specified model of the conditional density***. In this case, one can equivalently show that

$$\mathrm{E}\{\ln f(\mathbf{y}_i|\mathbf{x}_i, \boldsymbol{\theta}_0)\,|\mathbf{x}_i\} \geq \mathrm{E}\{\ln f(\mathbf{y}_i|\mathbf{x}_i, \boldsymbol{\theta})\,|\mathbf{x}_i\}.$$

The logarithm $\ln f(\mathbf{y}_i|\mathbf{x}_i, \boldsymbol{\theta})$ is denoted by $\ell_i(\boldsymbol{\theta})$ and called the ***(conditional) log-likelihood for observation i***.

One could then show that $\boldsymbol{\theta}_0$ solves $\max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \ell_i(\boldsymbol{\theta})$. Because observations are independent, for the whole dataset this would mean solving

[57] As it will be seen in the formulae that follow, the notion of a σ-finite measure is a generalisation of the notion of a probability measure. Technically, let $\mathcal{F}$ be a σ-algebra in space X. (A ***σ-algebra*** of a set is a system of its subsets such that (1) it contains an empty set and (2) it is closed with respect to a countable number of standard set operations.) Then $(X, \mathcal{F})$ forms a measurable space. Let $\mu: \mathcal{F} \to \mathbb{R}_+$ be a ***σ-additive measure***, i.e., a function that satisfies $\mu(\emptyset) = 0$ and $\mu(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$ for all pairwise disjoint sets $A_i \in \mathcal{F}$. This measure is called ***σ-finite*** if there is a countable number of subsets $\{A_i\}_{i=1}^{\infty} \subset \mathcal{F}$ such that $\mu(A_i) < \infty, i \in \mathbb{N}$ and $X = \cup_{i=1}^{\infty} A_i$. (Bogachev, 2007; Колмогоров & Фомин, 1976)

$$\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \mathcal{L}(\boldsymbol{\theta}) \equiv \max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sum_i l_i(\boldsymbol{\theta}),$$

where $\mathcal{L}(\boldsymbol{\theta})$ is the ***(conditional) log-likelihood function***. The estimator that maximizes $\mathcal{L}(\boldsymbol{\theta})$ is called the ***maximum likelihood estimator*** and is usually denoted as $\widehat{\boldsymbol{\theta}}$ or $\widehat{\boldsymbol{\theta}}_{ML}$.

Provided that several regularity conditions hold, the ML estimator is (Greene, 2008, p. 487):

- Consistent
- Asymptotically normally distributed
- Asymptotically efficient (and achieves the Cramér-Rao lower bound for consistent estimators)
- Invariant

Several optimisation methods for maximising the log-likelihood function are available (see, e.g., Wooldridge, 2010, pp. 431–435). Stata typically uses the modified Newton–Raphson method with step halving.

C.2 Logistic Regression

Logistic[58] regression, as probit regression, allows binary dependent variables, i.e., $y_i \in \{0,1\}$. In general, it is a special case of an ***index model*** (Wooldridge, 2010, p. 565)

$$\Pr(y = 1|\mathbf{x}) = G(\mathbf{x}\boldsymbol{\beta}), \tag{C-2}$$

where $0 < G(z) < 1$ for all $z \in \mathbb{R}$. The ***logit model*** is an index model with $G(z)$ being the cumulative distribution function (cdf) of the logistic distribution:

$$G(z) = \Lambda(z) \equiv \frac{\exp(z)}{1 + \exp(z)}.$$

Alternatively, one may want to have $\mathbf{x}\boldsymbol{\beta}$ instead of $G(\mathbf{x}\boldsymbol{\beta})$ on the right-hand side of the regression equation. This is easily done by applying the inverse function $G^{-1}(\cdot)$ to both sides of (C-2). In case of the logit model, this will lead to (Rabe-Hesketh & Skrondal, 2012, p. 502)

$$\operatorname{logit}\{\Pr(y = 1|\mathbf{x})\} \equiv \ln\left\{\frac{\Pr(y = 1|\mathbf{x})}{1 - \Pr(y = 1|\mathbf{x})}\right\} = \mathbf{x}\boldsymbol{\beta}.$$

The term in brackets is called the ***odds*** of y, given $\mathbf{x}$. In words, it is the expected number of successes per failure.

C.2.1 Estimation

Logistic regressions are estimated using ML estimators. For the logit model, the log-likelihood for observation i is (Wooldridge, 2010, p. 567)

$$\ell_i(\boldsymbol{\beta}) = y_i \ln \Lambda(\mathbf{x}_i\boldsymbol{\beta}) + (1 - y_i) \ln[1 - \Lambda(\mathbf{x}_i\boldsymbol{\beta})],$$

so that the log-likelihood of the whole sample is (StataCorp, 2011a, p. 980)

$$\mathcal{L}(\boldsymbol{\beta}) = \sum_{i\in S} \ln \Lambda(\mathbf{x}_i\boldsymbol{\beta}) + \sum_{i\notin S} \ln\big(1 - \Lambda(\mathbf{x}_i\boldsymbol{\beta})\big),$$

where S is the set of all observations for which the dependent variable $y_i = 1$.

C.2.2 Reporting

Typically, statistical packages output logistic regression results as coefficients. Their interpretation, though, is not as straightforward as in standard linear regression: for a one-unit change in x_j, the log odds of y will increase by $\hat{\beta}_j$.

[58] Throughout the whole work, I use the terms *logistic* and *logit* equivalently when referring to models.

Reporting results as odds ratios is commonly believed to be more natural for logit (and, generally, for logit-type) regressions. An ***odds ratio*** corresponding to $\hat{\beta}_j$ is $\psi_j = \exp(\hat{\beta}_j)$, and its standard error is $s_j^{\psi} = \psi_j s_j$, where s_j is the standard error of $\hat{\beta}_j$ (StataCorp, 2011a, p. 941). Many economists, however, find the interpretation of odds ratios difficult for understanding. It may be shown that (Rabe-Hesketh & Skrondal, 2012, p. 503)

$$\exp(\hat{\beta}_j) = \exp\left[\operatorname{logit}(y = 1|x_1, \dots, x_j + 1, \dots, x_p) - \operatorname{logit}(y = 1|\mathbf{x})\right] = \\ = \frac{\Pr(y = 1|x_1, \dots, x_j + 1, \dots, x_p)/\Pr(y = 0|x_1, \dots, x_j + 1, \dots, x_p)}{\Pr(y = 1|\mathbf{x})/\Pr(y = 0|\mathbf{x})}. \tag{C-3}$$

The odds ratio, thus, shows how many times the odds of y increases for a one-unit change in x_j. For example, assume that $y = 1$ if the respondent chooses to continue studies at university and x_j measures his average grade at secondary school. Then if $\psi_j = 1.2$, it means that if the average grade increases by one unit, the ratio of those who continue studies to those who do not increases 1.2 times or by 20%.

Another common way of reporting results is using ***marginal effects***. These are simply the partial derivatives[59] $\partial \Pr(y = 1|\mathbf{x})/\partial x_j$, whose interpretation is the same as for the standard linear regression. Denote

$$g(z) \equiv \frac{dG}{dz}(z).$$

Then consider two cases: a continuous explanatory variable and a discrete one. If variable x_j is continuous, applying the chain rule to (C-2), we get

$$\frac{\partial \Pr(y = 1|\mathbf{x})}{\partial x_j} = g(\mathbf{x}\boldsymbol{\beta})\beta_j. \tag{C-4}$$

Alternatively, if x_j is discrete then the effect on the probability of success from x_j increasing from c to $c + 1$ is

$$G\left[\sum_{k \neq j} \beta_k x_k + \beta_j(c + 1)\right] - G\left[\sum_{k \neq j} \beta_k x_k + \beta_j c\right].$$

Both expressions depend on $\mathbf{x}$ through $g(\mathbf{x}\hat{\boldsymbol{\beta}})$, meaning that to compute them, we should plug in some values for each explanatory variable. Nevertheless, this is what could be expected, since one is trying to measure the slope of a *nonlinear* function, which, generally, should be different at different points. For instance, one could evaluate the effect of higher education on the risk of unemployment for 20-year-old married individuals with one child. Typically, if not specified otherwise, marginal effects are computed at sample average, i.e., at $\bar{\mathbf{x}}$.

Return now to the interpretation of the coefficients and odds ratios. Given that for the logit cdf, $g(z) > 0$ for all $z \in \mathbb{R}$, (C-4) yields that the direction of the effect on the probability of success from a change in x_j is the same as the sign of β_j. For odds ratios, positive (negative) effects are those for which they are greater (smaller) than 1, simply because $\exp(0) = 1$.

C.3 Fixed-Effects Logistic Regression

In principle, fixed-effects logistic regression (Chamberlain, 1980) was developed for panel data analysis, i.e., when the same individuals are observed during several time periods. Conventionally, variables are indexed by i for independent groups and t for observations. The notation is typical for panel data: groups denote individuals, and observations denote time points

[59] In the case of a discrete variable, the marginal effect is the effect of a discrete change in its value.

at which individuals are observed. While the same method can be applied when groups and observations mean something completely different, I adhere to traditional notation here.

The fixed-effects logit model is written as

$$\Pr(y_{it} = 1|\mathbf{x}_{it}) = \Lambda(\alpha_i + \mathbf{x}_{it}\boldsymbol{\beta}). \tag{C-5}$$

The panel is allowed to be unbalanced, with $i = 1, \dots, n$ and $t = 1, \dots, T_i$. As in the standard logit, $y_{it} \in \{0,1\}$.

C.3.1 Estimation

Unfortunately, estimating the model (C-5) using ML estimator is impossible because of the ***incidental parameters problem*** (see Greene, 2008, p. 801). In short, the problem arises because the estimator depends on α_i, which causes it to be inconsistent not only for constant terms, but also for $\boldsymbol{\beta}$. The solution is to use ***conditional maximum likelihood*** estimator. Details below are taken from StataCorp (2011a, pp. 278–283).

Denote $\mathbf{y}_i = (y_{i1}, \dots, y_{iT_i})$ as the vector of outcomes for the ith group. Denote by $k_{1i} \equiv \sum_t y_{it}$ the number of ones in this vector and by $k_{2i} \equiv T_i - k_{1i}$ the number of zeros. Consider the probability that we observe vector $\mathbf{y}_i$, given that $\sum_t y_{it} = k_{1i}$:

$$\Pr\left(\mathbf{y}_i \middle| \textstyle\sum_{t=1}^{T_i} y_{it} = k_{1i}\right) = \frac{\exp\left(\sum_{t=1}^{T_i} y_{it}\mathbf{x}_{it}\boldsymbol{\beta}\right)}{\sum_{d_i \in S_i} \exp\left(\sum_{t=1}^{T_i} d_{it}\mathbf{x}_{it}\boldsymbol{\beta}\right)},$$

where $d_{it} \in \{0,1\}$, $\sum_t d_{it} = k_{1i}$ and S_i is the set of all possible combinations of k_{1i} ones and k_{2i} zeros. This formula does not depend on α_i and, hence, does not suffer from the incidental parameters problem. The denominator, denoted as $f_i(T_i, k_{1i})$, can be computed recursively:

$$f_i(T,k) = f_i(T-1,k) + f_i(T-1,k-1)\exp(\mathbf{x}_{iT}\boldsymbol{\beta}),$$

where $f_i(T,k) \equiv 0$ if $T < k$ and $f_i(T,0) \equiv 1$.

In this case, the conditional log-likelihood function is

$$\mathcal{L} = \sum_{i=1}^{n}\left\{\sum_{t=1}^{T_i} y_{it}\mathbf{x}_{it}\boldsymbol{\beta} - \ln f_i(T_i, k_{1i})\right\}.$$

C.3.2 Computation Time

Computation time is proportional to

$$p^2 \sum_{i=1}^{n} T_i \min(k_{1i}, k_{2i}).$$

C.4 Mixed-Effects Logistic Regression

If panel data or repeated cross-sections are used for analysis, the ordinary logistic regression model would fit *population-averaged* probabilities. Consider, for example, a dataset where there are repeated cross-sections of several countries and a model explaining overeducation. In that model, consider a binary explanatory variable indicating the disability of the respondent and assume that its estimated population-averaged odds ratio is 1.20. This would mean that the odds of being overeducated *among all individuals in all countries* is 20% higher for the disabled.

Another option might be to fit *subject-specific* (in the running example, country-specific) probabilities. This might be done using panel data methods, for instance, mixed effects logistic regression model with random intercepts at the level of countries. In particular, this would allow taking into account intra-country correlation. Continuing the example, if in a mixed effects logistic model, the estimated country-specific odds ratio for disability is 1.20, this would mean that the odds of being overeducated *for individuals in a given country* is 20% higher for the disabled.

Mixed-effects logistic model contains both fixed effects and random effects. The purpose is to allow for intra-cluster correlation of observations, in this case by assuming that they share common cluster-level random effects. I base the technical description of mixed-effects logit model on StataCorp (2011b, pp. 243–263).

In principle, this model allows several levels of nested clusters of random effects (also called hierarchical clustering). For instance, region clusters can be nested within country clusters. In this case, individuals comprise the first level, regions form the second and countries – the third.

For now, assume a two-level model with binary dependent variable, M independent clusters and a set of random effects $\mathbf{u}_k, k = \overline{1, M}$:

$$\Pr(y_{ik} = 1|\mathbf{u}_k) = \Lambda(\mathbf{x}_{ik}\boldsymbol{\beta} + \mathbf{z}_{ik}\mathbf{u}_k). \tag{C-6}$$

Here, $\mathbf{x}_{ik}$ are the covariates for the fixed effects (as in standard logit) with coefficients (fixed-effects) $\boldsymbol{\beta}$. The $1 \times q$ vector $\mathbf{z}_{ik}$ stores the covariates for the random effects, representing both random intercepts and random coefficients, as needed. The random effects $\mathbf{u}_k$ are M realisations from a multivariate normal distribution with mean $\mathbf{0}$ and $q \times q$ variance matrix $\boldsymbol{\Sigma}$. The random effects are not estimated directly, but instead are summarised according to the unique elements of $\boldsymbol{\Sigma}$, known as variance components. As before, $\Lambda(\cdot)$ denotes the logistic cdf.

C.4.1 Estimation

The two-level model (C-6) assumes Bernoulli data, which is a special case of binomial data. Hence, the methods to be discussed now are in terms of the more general two-level binomial mixed-effects model. See Pinheiro and Chao (2006) for the extension of the below methods to higher-level models.

Consider the response y_{ik} as the number of successes from a series of r_{ik} Bernoulli trials. For cluster $k, k = \overline{1, M}$, containing n_k elements, the conditional distribution of $\mathbf{y}_k = \left(y_{k1}, \dots, y_{kn_k}\right)'$, given a set of cluster-level random effects $\mathbf{u}_k$, is

$$f(\mathbf{y}_k|\mathbf{u}_k) = \prod_{i=1}^{n_k}\left[\binom{r_{ik}}{y_{ik}}\{\Lambda(\mathbf{x}_{ik}\boldsymbol{\beta} + \mathbf{z}_{ik}\mathbf{u}_k)\}^{y_{ik}}\{1 - \Lambda(\mathbf{x}_{ik}\boldsymbol{\beta} + \mathbf{z}_{ik}\mathbf{u}_k)\}^{r_{ik}-y_{ik}}\right] =$$

$$= \exp\left(\sum_{i=1}^{n_k}\left[y_{ik}(\mathbf{x}_{ik}\boldsymbol{\beta} + \mathbf{z}_{ik}\mathbf{u}_k) - r_{ik}\ln\{1 + \exp(\mathbf{x}_{ik}\boldsymbol{\beta} + \mathbf{z}_{ik}\mathbf{u}_k)\} + \ln\binom{r_{ik}}{y_{ik}}\right]\right).$$

Now define vector $\mathbf{r}_k \equiv \left(r_{k1}, \dots, r_{kn_k}\right)'$ and function

$$c(\mathbf{y}_k, \mathbf{r}_k) \equiv \sum_{i=1}^{n_k}\ln\binom{r_{ik}}{y_{ik}}.$$

Note that $c(\mathbf{y}_k, \mathbf{r}_k)$ does not depend on the model parameters.

We can now re-write the expression for $f(\mathbf{y}_k|\mathbf{u}_k)$ in matrix notation:

$$f(\mathbf{y}_k|\mathbf{u}_k) = \exp[\mathbf{y}_k'(\mathbf{X}_k\boldsymbol{\beta} + \mathbf{Z}_k\mathbf{u}_k) - \mathbf{r}_k'\ln\{1 + \exp(\mathbf{X}_k\boldsymbol{\beta} + \mathbf{Z}_k\mathbf{u}_k)\} + c(\mathbf{y}_k, \mathbf{r}_k)],$$

where $\mathbf{X}_k$ is the matrix of stacked row vectors $\mathbf{x}_{ik}$, $\mathbf{Z}_k$ is analogically formed by row vectors $\mathbf{z}_{ik}$ and the functions $\ln(\cdot)$ and $\exp(\cdot)$ are extended to be vector functions where necessary.

Because the prior distribution of $\mathbf{u}_k$ is multivariate normal with mean $\mathbf{0}$ and $q \times q$ variance matrix $\boldsymbol{\Sigma}$, the log-likelihood contribution for the kth cluster is obtained by integrating $\mathbf{u}_k$ out of the joint density $f(\mathbf{y}_k, \mathbf{u}_k)$:

$$\ell_k(\boldsymbol{\beta},\boldsymbol{\Sigma}) = (2\pi)^{-\frac{q}{2}}|\boldsymbol{\Sigma}|^{-\frac{1}{2}}\int f(\mathbf{y}_k|\mathbf{u}_k)\exp\left(-\frac{\mathbf{u}_k'\boldsymbol{\Sigma}^{-1}\mathbf{u}_k}{2}\right)d\mathbf{u}_k = \\ = \exp\{c(\mathbf{y}_k,\mathbf{r}_k)\}\,(2\pi)^{-\frac{q}{2}}|\boldsymbol{\Sigma}|^{-\frac{1}{2}}\int \exp\{g(\boldsymbol{\beta},\boldsymbol{\Sigma},\mathbf{u}_k)\}\,d\mathbf{u}_k, \tag{C-7}$$

where

$$g(\boldsymbol{\beta},\boldsymbol{\Sigma},\mathbf{u}_k) \equiv \mathbf{y}_k'(\mathbf{X}_k\boldsymbol{\beta}+\mathbf{Z}_k\mathbf{u}_k) - \mathbf{r}_k'\ln\{1+\exp(\mathbf{X}_k\boldsymbol{\beta}+\mathbf{Z}_k\mathbf{u}_k)\} - \frac{\mathbf{u}_k'\boldsymbol{\Sigma}^{-1}\mathbf{u}_k}{2}.$$

For convenience, the dependence on the observable data $(\mathbf{y}_k,\mathbf{r}_k,\mathbf{X}_k,\mathbf{Z}_k)$ is suppressed in the arguments of $g(\cdot)$.

The integral in (C-7) has no closed form and, thus, must be approximated. Of many proposed approximations (for the discussion of other available approaches, see Ng, Carpenter, Goldstein, & Rasbash, 2006), I here note two: Laplacian approximation and adaptive Gaussian quadrature.

The Laplacian approximation (Pinheiro & Bates, 1995; Tierney & Kadane, 1986) is based on a second-order Taylor expansion of $g(\boldsymbol{\beta},\boldsymbol{\Sigma},\mathbf{u}_k)$ about the value of $\mathbf{u}_k$ that maximises it. The maximiser of $g(\boldsymbol{\beta},\boldsymbol{\Sigma},\mathbf{u}_k)$ is $\widehat{\mathbf{u}}_k$ such that $g'(\boldsymbol{\beta},\boldsymbol{\Sigma},\widehat{\mathbf{u}}_k)=\mathbf{0}$. Suppressing the derivation, the Laplacian log-likelihood contribution of the kth cluster is

$$\ell_k^{Lap}(\boldsymbol{\beta},\boldsymbol{\Sigma}) = -\frac{1}{2}\ln|\boldsymbol{\Sigma}| - \ln|\mathbf{R}_k| + g(\boldsymbol{\beta},\boldsymbol{\Sigma},\widehat{\mathbf{u}}_k) + c(\mathbf{y}_k,\mathbf{r}_k),$$

where $\mathbf{R}_k$ is an upper-triangular matrix such that $-g''(\boldsymbol{\beta},\boldsymbol{\Sigma},\widehat{\mathbf{u}}_k)=\mathbf{R}_k\mathbf{R}_k'$. It was shown that $\widehat{\mathbf{u}}_k$ and $\mathbf{R}_k$ can be computed efficiently as the iterative solution to a least-squares problem (Pinheiro & Chao, 2006).

The idea of adaptive Gaussian quadrature (AGQ) (Liu & Pierce, 1994; Naylor & Smith, 1982) approximation for this case (as applied by Pinheiro & Bates, 1995) is based on re-writing the integral in question as

$$\int \exp\{g(\boldsymbol{\beta},\boldsymbol{\Sigma},\mathbf{u}_k)\}\,d\mathbf{u}_k = |\mathbf{R}_k|^{-1}\int \exp\{g(\boldsymbol{\beta},\boldsymbol{\Sigma},\widehat{\mathbf{u}}_k+\mathbf{R}_k^{-1}\mathbf{t})\}\,d\mathbf{t} = \\ = (2\pi)^{\frac{q}{2}}|\mathbf{R}_k|^{-1}\int \exp\left\{g(\boldsymbol{\beta},\boldsymbol{\Sigma},\widehat{\mathbf{u}}_k+\mathbf{R}_k^{-1}\mathbf{t}) + \frac{\mathbf{t}'\mathbf{t}}{2}\right\}\phi(\mathbf{t})\,d\mathbf{t}, \tag{C-8}$$

where $\phi(\cdot)$ is the standard multivariate normal density. As the integrand is now expressed as some function multiplied by a normal density, it can be estimated by applying the rules of standard Gauss–Hermite quadrature. In other words, we will replace the integral with a weighted sum of functions evaluated at a specific set of points. For a predetermined number of quadrature points N_Q, define $a=\sqrt{2}a^\star$ and $w=\pi^{-1/2}w^\star$, where $a^\star$ and $w^\star$ are a set of abscissas and weights for Gauss–Hermite quadrature approximations of $\int \exp(-x^2)\,f(x)\,dx$ (see Abramowitz & Stegun, 1972, p. 924). Define $\mathbf{a}=\left(a_{m_1},\dots,a_{m_q}\right)'$ to be a vector that spans the N_Q abscissas over the q dimensions of the random effects. Applying quadrature rules to (C-8) yields the AGQ approximation:

$$\int \exp\{g(\boldsymbol{\beta},\boldsymbol{\Sigma},\mathbf{u}_k)\}\,d\mathbf{u}_k \approx (2\pi)^{\frac{q}{2}}\widehat{G}_k(\boldsymbol{\beta},\boldsymbol{\Sigma}),$$

where

$$\widehat{G}_k(\boldsymbol{\beta},\boldsymbol{\Sigma}) \equiv |\mathbf{R}_k|^{-1}\sum_{m_1=1}^{N_Q}\dots\sum_{m_q=1}^{N_Q}\left[\exp\left\{g(\boldsymbol{\beta},\boldsymbol{\Sigma},\widehat{\mathbf{u}}_k+\mathbf{R}_k^{-1}\mathbf{a})+\frac{\mathbf{a}'\mathbf{a}}{2}\right\}\prod_{p=1}^{q}w_{m_p}\right].$$

From this, one can get the AGQ log-likelihood contribution of the kth cluster:

$$\ell_k^{AGQ}(\boldsymbol{\beta},\boldsymbol{\Sigma}) = -\frac{1}{2}\ln|\boldsymbol{\Sigma}| + \ln\left\{\widehat{G}_k(\boldsymbol{\beta},\boldsymbol{\Sigma})\right\} + c(\mathbf{y}_k,\mathbf{r}_k).$$

Table C-1 – Structures of the Random Effects Variance Matrix in Stata

Variance Structure	Variance	Covariance
Identity	equal for all random effects	all zero
Independent	unique for each random effect	all zero
Exchangeable	equal for all random effects	equal for all random effects
Unstructured	unique for each random effect	unique for each random effect

Source: compiled from StataCorp (2011b, p. 239)

AGQ is called adaptive, because it translates and rescales the integration variables in (C-8) by using $\hat{\mathbf{u}}_k$ and $\mathbf{R}_k^{-1}$, respectively, to better capture the features of the integrand. AGQ, as opposed to the Laplace method, does not critically depend on the quality of Taylor approximation: it is only used to redirect the quadrature abscissas, with the quality of AGQ approximation improving with N_Q. One can actually show that AGQ with $N_Q = 1$ is equivalent to Laplacian approximation (Pinheiro & Bates, 1995).

The log-likelihood function is simply the sum of the contributions of all clusters, i.e. $\mathcal{L}(\boldsymbol{\beta},\boldsymbol{\Sigma}) = \sum_{k=1}^{M} \ell_k^{Lap}(\boldsymbol{\beta},\boldsymbol{\Sigma})$ in case of Laplacian approximation or $\mathcal{L}(\boldsymbol{\beta},\boldsymbol{\Sigma}) = \sum_{k=1}^{M} \ell_k^{AGQ}(\boldsymbol{\beta},\boldsymbol{\Sigma})$ for AGQ approximation. $\mathcal{L}(\boldsymbol{\beta},\boldsymbol{\Sigma})$ is then maximised with respect to $(\boldsymbol{\beta},\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a vector of the unique elements of the matrix square root[60] of $\boldsymbol{\Sigma}$.

C.4.2 Specifying the Variance–Covariance Structure of Random Effects

In Stata, one is allowed to specify different structures of the variance matrix $\boldsymbol{\Sigma}$. The four available options are listed in Table C-1. Moreover, one can easily execute the model with the variance matrix for random effects combined from sub-matrices each having one of the four structures. For instance, it may take the form

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma & \sigma & 0 & 0 & 0 \\ \sigma & \sigma_1^2 & \sigma & 0 & 0 & 0 \\ \sigma & \sigma & \sigma_1^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_4^2 & \sigma_{45} & \sigma_{46} \\ 0 & 0 & 0 & \sigma_{45} & \sigma_5^2 & \sigma_{56} \\ 0 & 0 & 0 & \sigma_{46} & \sigma_{56} & \sigma_6^2 \end{bmatrix}.$$

In this case, the first three random effects have exchangeable variance structure, while the last three have unstructured variance structure.

C.4.3 Computation Time

The computation time of mixed-effects logistic regression is a function of

$$p^2 M\left[1 + \left(N_Q\right)^{q_t}\right],$$

where p is the number of estimable parameters and q_t is the total number of random intercepts and coefficients at all levels. Given the model, the computation time is a power function of the number of quadrature points, meaning that even with a few random parameters, the computation time can quickly become prohibitively long.

Laplacian approximation is a good alternative to AGQ, especially taking into account the quick computation time, as $N_Q = 1$ in this case. Based on their experience, StataCorp (2011b, p. 256) notes that

- Odds ratios and their standard errors are well approximated by the Laplacian method

[60] Alternatively, the matrix logarithm of $\boldsymbol{\Sigma}$ can be used.

- Estimates of variance components under Laplacian approximation exhibit bias, especially at lower levels
- The model log-likelihood and comparison likelihood ratio test[61] under Laplacian approximation are very close to AGQ results

Laplacian approximation, thus, can be used when the interest is mainly in fixed effects and during the model-building phase when comparing competing models. After choosing the best model, one can increase the number of quadrature points to get better estimates.

C.4.4 Reporting

As all logit-type regressions, mixed-effects logit outputs fixed effects as coefficients or odds ratios. Random effects appear as variances and covariances or as standard deviations and correlation coefficients.

Marginal effects cannot be computed after this type of models. The reason is that marginal effects depend on all covariates, including random effects, which is not allowed, since the latter are stochastic quantities (recall that they are simply draws from a certain distribution). The only available option is to compute marginal effects using predicted results based on fixed effects only. However, this means ignoring random effects and, hence, may lead to incorrect results.

C.5 Ordered Discrete Choice with Few Levels: (Generalised) Ordered Logit

The discrete dependent variable may have more than two values, in which case it is necessary to disentangle the cases when these values can be ordered from the cases when these values represent groups that cannot be ordered in any meaningful way. This and the following section deal with the former cases.

If the number of categories is sufficiently small, ordered logit models could be used. Assume y takes on $K+1$ possible values from $\{0,1,\dots,K\}$ .The standard ***ordered logit model*** is formulated as (Wooldridge, 2010, p. 655)

$$\begin{cases} y_i^\star = \mathbf{x}_i\boldsymbol{\beta} + \varepsilon_i \\ y_i = 0, \qquad y_i^\star \le \alpha_1 \\ y_i = 1, \qquad \alpha_1 < y_i^\star \le \alpha_2 \\ \dots \\ y_i = K, \qquad y_i^\star > \alpha_K, \end{cases}$$

where $\alpha_1 < \alpha_2 < \dots < \alpha_K$ are unknown ***cut points*** and ε_i is distributed logistically.

This model is also known as the ***proportional-odds ordered logit model*** after its key assumption, that the proportion of the odds $\Pr(y \le k)/\Pr(y > k)$ has the same value for all $k = \overline{1,K}$.[62] This need not be the case, of course, and indeed, some of its diagnostic tests test this assumption explicitly (Brant, 1990; Wolfe & Gould, 1998).

There are several options to proceed when the proportional-odds assumption fails to hold. One of them is using the ***generalised ordered logit model*** (Williams R. , 2006), which constrains the coefficients for which the assumption holds but relaxes the constraints on those variables whose odds vary by category of the dependent variable. This, thus, allows estimating a model that is more realistic than the standard ordered logit model but still more parsimonious than a multinomial logit model, which assumes that the alternatives cannot be ordered and estimates $K \times p$ parameters (as compared to p in case of proportional-odds ordered logit and somewhere between p and $K \times p$ in case of generalised ordered logit model).

---

[61] See Section C.8 for definition.

[62] This assumption is also known as the ***parallel-lines assumption***.

C.5.1 Estimation of the Proportional-Odds Ordered Logit Model

Both coefficients and cut points are estimated by ML. It can be shown that the probability of a given observation is (StataCorp, 2011a, p. 1396)

$$\Pr(y_i = k) = \Pr(\alpha_{k-1} < \mathbf{x}_i\boldsymbol{\beta} + \varepsilon_i \le \alpha_k) =$$
$$= (1 + \exp(-\alpha_k + \mathbf{x}_i\boldsymbol{\beta}))^{-1} - (1 + \exp(-\alpha_{k-1} + \mathbf{x}_i\boldsymbol{\beta}))^{-1},$$

where $\alpha_0 \equiv -\infty$ and $\alpha_K \equiv \infty$.

Let $I_j(y_i)$ be an indicator function that is equal to one only if $y_i = j$. Then the log-likelihood function is (StataCorp, 2011a, p. 1396)

$$\mathcal{L} = \sum_{i=1}^{N}\sum_{j=0}^{J} I_j(y_i)\ln(\Pr(y_i = j)).$$

C.5.2 Reporting of the Proportional-Odds Ordered Logit Model

When reporting the results of the standard ordered logit model, it is necessary to bear in mind that β_j unambiguously determines the direction of the effect of x_j on the probabilities $\Pr(y = 0)$ and $\Pr(y = K)$, but not on the intermediary probabilities $\Pr(y = k), 0 < k < K$ (Wooldridge, 2010). Thus, depending on the purpose of the study, it may be sufficient to report the coefficients (possibly, in form of odds ratios) or to report the marginal effects on the probability of each of K outcomes (relatively to the base outcome, $k = 0$).

C.6 Ordered Discrete Choice with Many Levels: Censored Linear Regression

Assume again that the dependent variable represents categories that can be ordered, but that the number of categories is sufficiently high to make the standard output of ordered logit in terms of marginal effects unreadable. For instance, such variable may measure the opinion of the respondent on a 0–10 scale. In this case, one could approximate the discrete variable by a *censored continuous* variable[63]. For instance, if the actual variable takes values over the set $\{0,1,\ldots,10\}$, we can assume that this is actually a continuous variable running over the interval $[0,10]$. The latter model can be fit by the two-limit tobit regression, which I discuss below. The original tobit regression (Tobin, 1958) assumed that data are left-censored (e.g., it can only be nonnegative), and the two-limit tobit model is its generalisation.

Let $a_1 < a_2$ be the two limit values of the dependent variable y. The ***two-limit tobit model*** in terms of an underlying latent variable is formulated as (Wooldridge, 2010, p. 704)

$$\begin{cases} y_i^\star = \mathbf{x}_i\boldsymbol{\beta} + \varepsilon_i \\ y_i = a_1, & y_i^\star \le a_1 \\ y_i = y_i^\star, & a_1 < y_i^\star < a_2 \\ y_i = a_2, & y_i^\star \ge a_2. \end{cases}$$

where $\varepsilon_i|\mathbf{x}_i$ is normally distributed with zero mean and variance σ^2. Here, the variable y is observed, and $y^\star$ is unobserved or latent.

C.6.1 Estimation

One can show that (Wooldridge, 2010, p. 704)

---

[63] Wooldridge (2010) prefers to call this variable a ***corner solution response*** or a ***corner solution outcome***, and the models for such variables as ***corner solution models*** as opposed to a more conventional, but also misleading in our case, term ***censored regression models***.

$$\Pr(y = a_1|\mathbf{x}) = \Phi\left(\frac{a_1 - \mathbf{x}\boldsymbol{\beta}}{\sigma}\right),$$
$$\Pr(y = a_2|\mathbf{x}) = \Phi\left(-\frac{a_2 - \mathbf{x}\boldsymbol{\beta}}{\sigma}\right),$$

where $\Phi(\cdot)$ is the cdf of the standard normal distribution.

Two-limit tobit is estimated by ML. Let L be the subset of left-censored observations (i.e., where $y_i^\star \le a_1$), R be the subset of right-censored observations and I be the subset of observations where $a_1 < y_i^\star < a_2$. Then the log-likelihood function for the sample is (StataCorp, 2011a, p. 788)

$$\mathcal{L} = \sum_{i\in L} \ln \Phi\left(\frac{a_1 - \mathbf{x}\boldsymbol{\beta}}{\sigma}\right) + \sum_{i\in R} \ln \Phi\left(-\frac{a_2 - \mathbf{x}\boldsymbol{\beta}}{\sigma}\right) + \sum_{i\in I} \ln \left\{\Phi\left(\frac{a_2 - \mathbf{x}\boldsymbol{\beta}}{\sigma}\right) - \Phi\left(\frac{a_1 - \mathbf{x}\boldsymbol{\beta}}{\sigma}\right)\right\}.$$

This function is then maximised with respect to $\boldsymbol{\beta}$ and σ.

C.7 Variance Estimation for Clustered Samples

Sometimes, observations form clusters in the dataset and observations in the same cluster can be correlated, while observations in different clusters remain uncorrelated. This violates a standard assumption of the independence of observations and has to be dealt with. In technical parlance, this is called *allowing for intra-cluster correlation*. Some of the panel-data methods applicable in such cases were discussed above, but there are cases when it makes more sense to stick to the standard cross-sectional methods.

This section covers the estimator that affects the standard errors and the variance–covariance matrix, but keeps the estimated coefficients unchanged. As noted by Williams (2000)[64], this is a special case of the robust variance estimator used when the explanatory variables and the error term are distributed independently but not necessarily identically; the latter estimator is also called Huber–White estimator (since it was independently discovered by these two scientists) or sandwich estimator (because of how its formula appears).

Assume that the model is estimated with ML. Denote the true coefficient vector as $\boldsymbol{\beta}$ and the number of observations as N. Under a technical assumption that $\boldsymbol{\beta}$ is in the interior of $\mathbf{B}$ and the log-likelihood function $\ell_i(\boldsymbol{\beta})$ is twice continuously differentiable on that interior for each observation i, define the ***Hessian*** for observation i as the matrix of second partial derivatives of $\ell_i(\boldsymbol{\beta})$:

$$\mathbf{H}_i(\boldsymbol{\beta}) \equiv \nabla^2_{\boldsymbol{\beta}}[\ell_i(\boldsymbol{\beta})].$$

The conventional estimator of variance is then[65] (Wooldridge, 2010, p. 479)

$$\widehat{\mathbf{V}}_{ML} = \left[-\sum_{i=1}^{N} \mathbf{H}_i(\widehat{\boldsymbol{\beta}})\right]^{-1}.$$

Denote by $\mathbf{u}_i$ the ***score*** of the log-likelihood for observation i, $\mathbf{u}_i \equiv \nabla_{\boldsymbol{\beta}} l_i(\boldsymbol{\beta})$. Then the robust estimator of variance is (StataCorp, 2011c, p. 295)

$$\widehat{\mathcal{V}}_{ML} = \widehat{\mathbf{V}}_{ML}\left(\sum_{i=1}^{N} \mathbf{u}_i'\mathbf{u}_i\right)\widehat{\mathbf{V}}_{ML}.$$

If now the observations are only independent across clusters $C_1, \dots, C_M$, the estimator becomes (StataCorp, 2011c, p. 295)

[64] But see also Froot (1989) as the seminal reference and Rogers (1993) on how this is estimated in Stata.

[65] Henceforth, I deliberately ignore the finite-sample corrections for simplicity.

$$\hat{\mathcal{V}}_{ML}^{CL} = \hat{\mathbf{V}}_{ML}\left(\sum_{k=1}^{M} \mathbf{u}_k^{(C)\prime}\mathbf{u}_k^{(C)}\right)\hat{\mathbf{V}}_{ML},$$

where $\mathbf{u}_k^{(C)}$ is now the score of the log-likelihood for kth cluster.

Indeed, it is easy to see that the clustered estimator simply uses a different decomposition of $\nabla_{\boldsymbol{\beta}}\mathcal{L}(\boldsymbol{\beta})$: $\mathbf{u}_k^{(C)}, k = \overline{1, M}$, instead of $\mathbf{u}_i, i = \overline{1, N}$.

C.8 Typical Goodness-of-Fit Statistics for Logit-Type Regressions

One of the most common model comparison tests for models estimated by methods equivalent to ML is the ***likelihood ratio test*** (LR test). Two models are compared: an unrestricted model (with the log-likelihood of $\mathcal{L}_U$) and a restricted model (with the log-likelihood of $\mathcal{L}_R$). The restricted model should be nested within the unrestricted one, meaning that the set of explanatory variables of the former should be a subset of the set of explanatory variables of the latter. The ***likelihood ratio*** is then defined as

$$LR = -2(\mathcal{L}_R - \mathcal{L}_U).$$

If the constrained model is true, LR is approximately χ^2 distributed with the number of degrees of freedom equal to the number of restrictions imposed (Greene, 2008, p. 500).

Another common method to compare models estimated by ML-equivalent methods is using information criteria. Two information criteria are usually used. The ***Akaike's*** (1974) ***information criterion*** (AIC) is

$$AIC = -2\mathcal{L} + 2k,$$

while Schwarz's (1978) ***Bayesian information criterion*** (BIC) is

$$BIC = -2\mathcal{L} + k \ln N.$$

Here, $\mathcal{L}$ is model's log-likelihood and k is the number of estimated parameters. The benefit of both information criteria is that models need not be nested to be compared. A better model has a lower value of AIC or BIC. Information criteria allow to balance model fit (determined by its log-likelihood) and its size (the second summand in AIC/BIC), thus, both allowing a good fit and preventing over-fitting the data at hand (see Hand, Mannila, & Smyth, 2001, Section 7.4, for an excellent discussion).

Analogously to the R^2 goodness-of-fit measure for models fit by OLS, several goodness-of-fit measures were proposed for models fit by ML. These are called ***pseudo $\boldsymbol{R^2}$*** measures, as they look like the traditional R^2 in that their range is $[0, 1]$ and a higher pseudo R^2 indicates a better fit, but they cannot be interpreted as the proportion of variance for the dependent variable explained by the model. Many different pseudo R^2s have been proposed (see UCLA Academic Technology Services, 2011; Greene, 2008, pp. 790–793, and references therein). One of the most frequent measures is ***McFadden's pseudo $\boldsymbol{R^2}$***:

$$R_{McF}^2 = 1 - \frac{\mathcal{L}}{\mathcal{L}_0},$$

where $\mathcal{L}$ is the log-likelihood of the model and $\mathcal{L}_0$ is the log-likelihood of the same model but only with a constant term (i.e., it measures the change in log-likelihood after adding all explanatory variables).

Finally, if the dependent variable is binary, the model could be viewed as a classifier and one could use ***receiver operating characteristic*** (ROC) ***analysis*** to summarise the accuracy of its prediction (see Fawcett, 2006, for an introduction; and Pepe, 2003, for more elaborate treatment of ROC analysis). The analysis uses the ***ROC curve***, a graph of sensitivity versus (1 − specificity), where ***sensitivity*** is the fraction of correctly-classified positive cases (true positives rate) and ***specificity*** is the fraction of correctly classified negative cases (true negatives rate). For each possible outcome of the model, sensitivity and specificity are computed and the corresponding point is plotted on the graph; these points are then connected by straight lines. The performance of a classifier can be summarised as the ***area under*** its ***ROC curve*** (ROC AUC statistic). Classifiers can then be compared by comparing their ROC AUCs: the larger the area, the better the quality of classification.

APPENDIX D AGENT-BASED MODELLING METHODOLOGY

The Agent-Based Computational Modelling technique […] is an effective public policy research tool for complex social systems.

(The Brookings Institution, 2011)

A hearing of the House of Representatives Committee of Science and Technology on July 20th [2010 that] targeted the dynamic stochastic general equilibrium models […] question[ed] the wisdom of relying for national economic policy on a single, specific model when alternatives are available. […] The Institute for New Economic Thinking at New York […] has attacked many of the assumptions […] that were clearly too simplistic. […] One of the most promising options was […] agent-based models of the economy.

(The Economist, 2010)

In this chapter, I describe agent-based modelling (ABM). I then argue that this is one of the best methods for modelling complex adaptive systems, one of the brightest examples of which is an economy. Of course, there are limitations to ABM, as to any other method; I discuss them in the end of the chapter.

D.1 What Is Agent-Based Modelling

The pioneers in using agent-based modelling to study social phenomena were James Sakoda and Thomas Schelling (for their early research, see Sakoda, 1971; Schelling, 1971). In the 1980s, after analysing biological and social models, research groups at the Massachusetts Institute of Technology started studying the possibilities of using multi-agent systems in problem solving. Typically, intelligent agents are software problem solvers with four characteristics (Luger, 2009):

1. The agent is ***situated***: it reacts to the signals from its environment, its feedback can change the environment
2. It is ***autonomous***: it controls its actions and internal state and is able to act without direct manipulation by other agents
3. It is ***flexible***: the agent not only responds to environmental signals, it also can plan its actions to possible future signals to reach its goal
4. It is ***social***: it can interact with other agents

Luger (2009, p. 19) thus defines an ***agent*** as "an element of a society that can perceive (often limited) aspects of its environment and affect that environment either directly or through cooperation with other agents."

In applications of ABM to economics and finance, these features remain unchanged. Without any global controller, autonomous agents representing various model entities like individuals, organisations or governments – depending on the purpose of the model – are making choices, performing actions and communicating with each other, trying to achieve some specific goal(s).

In artificial intelligence (AI) and other areas where ABM is applied, modelling is never done analytically. In AI, ABM was initially used as an alternative to approaches that use mathematical logic with assumptions like perfect rationality of system components to specify component behaviour. Instead, the goal of ABM is to observe and then interpret the dynamically evolving structures during simulation runs (Deckert & Klein, 2010). Thus, computer simulations are run, and simulation output is then interpreted by the modeller in the context of model purpose.

One of the unquestionable benefits of applying ABM to study the dynamics of some system is that one is not constrained to specifying system behaviour in the form that is mathematically soluble. Firstly, this allows building complex behaviour models using many if–then rules. This is important, because behavioural patterns are captured by flow charts much easier and more naturally than by systems of equations. Secondly, simulations can contain many random elements governed by different probability distributions. Applications in economics and finance, as opposed to some other fields ABM is applied in, nearly always require randomness in the specifications of agent or system behaviour. Mathematical modelling in such settings quickly becomes prohibitively difficult.

As in any model, there is a trade-off between the correspondence with reality and complexity in ABM. Each agent-based model depends on a set of parameters. Each parameter can be defined by a single value or a set of values[66] or drawn from a probability distribution. If there is no randomness in the model, it is sufficient to run it once for each combination of parameters. Otherwise, for each combination of non-random parameters, the model should be run several times, for obvious reasons (of course, more is better). Because, as it was already mentioned, in agent-based models in economics and finance, randomness is always included, one is forced to define parameters as precisely as possible to minimise the number of model runs. Moreover, if many parameters are not defined precisely using empirical data or theory, the simulation gives less reliable results. The reason is that the modeller has to run the simulation many times, and he does not really know, which parameter combination corresponds to the real world; thus, only results that appear consistent in all runs might be treated as reliable (Brenner & Werker, 2007). Nevertheless, it is always recommended to check the sensitivity of the model, i.e., the stability of its output with respect to small changes in parameter values.

D.2 How to Do Agent-Based Modelling

The typical process of building an agent-based model follows these steps:

1. Identify the goal of the model
2. Create the abstract structure of the model
3. Implement the model (e.g., using a specialised software package[67])
4. Set the values for model parameters
5. Validate the model
6. Run the model
7. Analyse the output of the model
8. Do sensitivity analysis

D.2.1 From Goal to Implementation

Every model has its purpose reflected in the research question. Four general types of research questions that may be pursued by using ABM are distinguished by Brennen and Werker (2007). Firstly, the aim might be to describe the characteristics of the modelled system and analyse the

---

[66] Even if a parameter is defined to belong to some interval of values, the modeller should choose several distinct values from it. Obviously, one cannot run simulations for the whole interval of continuous parameter values.

[67] Several packages are available on the market. Some of the most popular include NetLogo, Repast, MASON and Swarm. Reviewing the available packages is outside the scope of this text. An interested reader is referred to, e.g., Railsback, Lytinen and Jackson (2006), Zheng et al. (2013, Ch. 3) for a comparison of the most popular platforms and Nikolai and Madey (2009) for a more comprehensive review.

relationships between them. Secondly, the model might be built as a prediction tool. Thirdly, the modeller might aim at testing hypotheses about causal relationships, in which case the hypotheses are translated into model assumptions, which are then checked against empirically observable behaviour of the modelled system. Finally, the goal might be to categorise the modelled system into sub-classes based on its behaviour in different specifications.

Subject to the chosen research question, an abstract (mathematical) model is specified. It is then implemented in the general-purpose or specialised software package, depending on the modeller's programming skills and model complexity.

D.2.2 Calibration: Setting Parameters

Even if the general specification of agent behaviour corresponds to reality, the modeller should pay considerable attention to setting model parameters. For instance, if a LEMS model is created for a specific country – e.g., France – then it should reflect the particularities of LEMS in France and not in the US.

This is one of the longest steps in the model building process. There are two reasons behind that. Firstly, not all parameters of the model are readily available in macro- or micro-level statistical data and might have to be estimated or approximated. Secondly, in many cases, there will be parameters for which there are in principle no data available, and these will have to be set to the values at which the model generates the expected output. Searching for these "right" values might take a long time. Generally, the model should be run for every possible value of every such parameter. In practice, the model is run for only several of these values for a given parameter. The number of runs to be executed increases rapidly with the number of such unknown parameters, which is why very few unknown parameters are included in the model.

There are three main approaches to calibration: indirect calibration, Werker–Brenner approach and history-friendly approach (see Table D-1).[68] Indirect calibration works by setting parameter values so that the model generates a set of ***stylised facts***, the well-known facts that have been repeatedly proven empirically. The Werker–Brenner approach advocates the use of Bayesian inference procedures and expert opinions in choosing between competing models. Finally, the history-friendly approach aims at finding parameter values that replicate historical dynamics. Each approach has its drawbacks; these are also described in Table D-1.

D.2.3 Validation: Checking the Model

Validation[69] is the process of checking that the model is a correct representation of reality or, in other words, that the output of the simulation is comparable to that of the real system (Deckert & Klein, 2010).

In agent-based models, it is particularly important to validate model structure and behaviour (Richiardi, Leombruni, Saam, & Sonnessa, 2006).[70] ***Model structure validation*** includes checking:

---

[68] While in Fagiolo, Moneta and Windrum (2007), these are called validation approaches, their aim is "to reduce the number of model parameters and the space of possible 'worlds' that are explored by tying the model down to an observed empirical reality," (Fagiolo, Moneta, & Windrum, 2007, p. 206), which is closer to setting parameters (calibration) than to checking the model for correctness (validation).

[69] Frequently, validation is mentioned together with verification, which means checking that the computer programme executes correctly, i.e., the system implemented by the programme corresponds to the conceptual model the modeller intended to study. Informally, verification means checking that the modeller is "building the thing right." Verification is an important step, but it should be performed at the implementation stage.

[70] Richiardi, Leombruni, Saam and Sonnessa (2006) also consider validating the implementation, which is the same as verification.

Table D-1 – Comparison of Approaches to Building Simulations

Indirect Calibration:

1. Identify a set of stylised facts to be reproduced/explained by the model
2. Build the model, keeping micro behaviour close to empirical and experimental evidence
3. Use empirical evidence on stylised facts to restrict parameter space
4. Explore the causal mechanisms that underlie stylised facts or explore the emergence of fresh stylised facts

Problems:

- Micro and macro parameters not calibrated using their empirical counterparts
- Unclear how to interpret alternative parameter values in the sub-region of the parameter space that appears after step 3

Werker–Brenner Approach:

1. Use empirical knowledge to calibrate initial conditions and the ranges of parameters
2. Perform empirical validation of the outputs for each of the model specifications derived from step 1 (using Bayesian inference procedures)
3. Ask expert opinion

Problems:

- Assessing fitness among a class of models does not automatically help to identify a true underlying model
- Calibration tends to influence the models developed
- The quality of available empirical data might be poor
- Unclear whether the data generating process is ergodic*
- Unclear what should be the initial conditions
- The observed parameters might be time-dependent
- Unclear to what extent predictions take into account data outside the current regime

History-Friendly Approach:

1. Build models on various detailed data – from detailed empirical studies to anecdotal evidence to histories written about the industry under study
2. Compare model output to the “actual” history
3. Having identified “history-replicating” parameters, seek for “history-divergent” results

Problems:

- In practice, such modelling is based on the history of a few key players rather than of the entire industry
- Impossible to get all relevant data to build an empirically sound model
- Limited attention given to sensitivity analysis
- An individual simulated trace that resembles the actual history may or may not be typical of the model
- If several combinations of parameters produce an identical output trace, it is unclear which combination is correct for initial settings
- Difficult to construct counter-factual histories

Source: compiled from Fagiolo, Moneta and Windrum (2007)

* Informally, if the underlying stochastic process is ***ergodic***, we can take its single sufficiently long realisation and, by studying it, infer about all possible realisations of this process. If it is not, the generalisation of the properties of a single realisation of the process to all its realisations is impossible.

- whether the model structure is consistent with the relevant descriptive knowledge of the system (***structure verification***)[71]
- whether model behaviour makes sense even when parameters take on extreme values (***extreme condition verification***)
- whether the important concepts for addressing the problem are endogenous (***boundary adequacy verification***)

Model behaviour validation includes checking:

- whether the model generates the symptoms of the problem under study, behaviour modes, phasing, frequencies and other characteristics of behaviour of the real system (***behaviour reproduction***)

[71] ***Input*** or ***ex ante validation*** (Bianchi, Cirillo, Gallegati, & Vagliasindi, 2007) is also concerned with this subtype of model structure validation, but the authors also include validating parameter values through analysing empirical data, for which there is no need if calibration was properly performed.

- whether anomalous behaviour arises if an assumption of the model is deleted (***behaviour anomaly***)
- whether the model can reproduce the behaviour of other examples of systems in the same class as the model (***family reproduction***)
- whether the model behaves properly when subjected to extreme policies (***extreme policy verification***)

Besides reproducing historical behaviour, some researchers argue for using ***predictive*** (as opposed to ***descriptive***) ***validation***, i.e., checking the ability of the model to match future, yet unknown, behaviour to make "exhaustive analysis of a model meant to reproduce reality" (Bianchi, Cirillo, Gallegati, & Vagliasindi, 2007, p. 247).

D.2.4 Running the Model

After the model has been implemented and its parameters set, it is ready to run. If there is absolutely no randomness in the model, which is nearly never the case in practice, it is sufficient to run the model once and study its output. Typically, there is at least one random parameter – e.g., random change in the sales of a specific firm or selection of a random agent with whom to communicate from the list of agents. In this case, it is necessary to run the model multiple times to see how it performs with different sequences of random numbers.

The question is, of course, how many times to run the model. There are two possibilities. One is to select a sufficient from the statistical point of view number of runs – e.g., 30, 50 or 100 – depending on how many resources one run of the model uses, and execute these runs. The second option is to choose a set of key output variables and run the model until they stabilise (Lorscheid, Meyer, & Hocke, 2013). This option might be more attractive than the first option: there might be no reason to run the model for 500 times if output stabilises already after 70 runs. No one guarantees that, though, and the modeller could easily have to run the model more times than (s)he expected to do. Of course, following this approach requires additional effort from the researcher (checking how the distributions of the key variables change with the number of runs and executing additional runs if necessary), but the result of that is certainty that running the model more times will not result in substantial changes.

D.2.5 Analysing Output and Its Sensitivity

Methods used in analysing output depend on the research question and range from the description of the dynamics of some characteristic of the model to heavy statistical analysis.

An additional analysis required after analysing output at basic assumptions and parameter values is ***sensitivity analysis***. It amounts to checking how the output of the model changes if its assumptions or parameters are altered. This has two purposes: to understand which assumptions and/or parameters influence model's results and to identify the region of parameter values where the central result of the model holds (Deckert & Klein, 2010).

D.3 The Motivation behind Using Agent-Based Modelling

D.3.1 Complex Adaptive Systems

There are several definitions of the term ***complex adaptive system*** (CAS), some of which are given in Table D-2. Before providing a deeper conceptual understanding of what a CAS is, I note some examples of this class of systems. Naturally, any group of living organisms could be considered a CAS (Levin, 1998), including any social group of humans. Indeed, studies on *complex*

Table D-2 – Definitions of a Complex Adaptive System

Definition	Author
A "system that involves many components that adapt or learn as they interact"	Holland (2006)
A "fluidly changing collection of distributed interacting components that react to both their environments and to one another"	Argonne National Laboratory (n.d.a)
A system "having a large number of self-similar agents that utilize one or more levels of feedback, exhibit emergent properties and self-organization, and produce non-linear dynamic behaviour"	Association for the Advancement of Artificial Intelligence (2010)
A system whose macro-level properties "emerge from interactions among components and may feed back to influence the subsequent development of those interactions"	Levin (1998)
A system consisting of a network of interacting adaptive agents that exhibits a dynamic aggregate behaviour that emerges from the individual activities of the agents but that can be described without a detailed knowledge of the behaviour of the individual agents; agents are adaptive in the sense that their actions in the environment can be assigned a value that, due to agents' behaviour, increases over time	Holland & Miller (1991)

social networks are a subclass of social group studies that analyse these from the complex systems point of view. One of the richest examples is, of course, the economy[72]. As noted by Amin (2000), a "CAS model is particularly appropriate for any industry made up of many geographically dispersed components that can exhibit rapid global change as a result of local actions." For instance, energy markets are nowadays frequently modelled as CASs, combining the behavioural simulation of both energy companies and their clients (Argonne National Laboratory, n.d.b; Wildberger, 1997).

I will now rephrase the above-given definitions to highlight several distinguishing features of CASs and illustrate them by the example of a firm.

Firstly, such systems consist of a set of elements, usually called *agents*. Normally, the number of agents in a system is considerable but finite. Depending on the problem, this number can be fixed or can change. In the firm example, agents are workers, their number is, obviously, finite and it can change due to employee turnover and other typical workforce size changes.

An agent itself is a collection of properties and behaviours. The fundamental characteristic of agents is the heterogeneity of their properties within a given CAS. For instance, workers are typically of different age, they have different education, experience and abilities, which makes a particular worker more or less suitable for performing particular tasks. Agent's properties do not have to be fixed over time: while worker's ability level might be considered innate, their age and experience change during their life.

Agents are connected by networks. For instance, workers are connected by friendship or acquaintance networks. Networks normally can evolve over time, with some edges being added and some deleted.

Both agent properties and networks to which it belongs shape its behavioural response to changes in the environment. The behaviour of a CAS agent adaptively changes over time, in the

[72] The first notable modern cross-disciplinary discussion on what class of systems the economy belongs to and what methods, thus, would be appropriate for its analysis was held during the workshop "Evolutionary Paths of the Global Economy" in the Santa Fe Institute in 1987. As stated in the proceedings of the workshop (Anderson, Arrow, & Pines, 1988), its purpose was "to explore the potential usefulness of a broadly transdisciplinary research program on the dynamics of the global economic system, by bringing together a group of economists and a group of natural scientists who have developed techniques for studying nonlinear dynamical systems and adaptive paths in evolutionary systems." An informal description of the event can be found in Beinhocker (2006).

sense that it constantly evaluates the gains from previous behaviours and seeks ways to increase them in the future. Nowhere is it claimed that the agent *optimises* its utility, fitness or some other measure of gains. In the running example, suppose that an unemployed agent is seeking a job. He has two channels to get the information on the job: formal (newspaper advertisements, specialised jobseeker portals) and informal (ask friends or relatives). In the latter case, he utilises his social network connections. In principle, he can allocate his time resources to formal search and to informal search. Given that total resources are constant, he can then try different allocations, moving more resources to the method that, historically, provided better results. It is, however, unreasonable to assume that he finds the *optimal* allocation of time resources.

Finally, the crucial property of a CAS is ***emergence***, described by Axtell (2007, p. 111) as the phenomenon when "the interaction of autonomous or quasi-autonomous components of the system […] leads to higher level functionality that is not present in any of the individual components." In other words, agents in a CAS are able to create macro-level structures and dynamics that could not be forecasted using only knowledge about micro-level characteristics and behaviour of agents.

Even a quick glance on the description of ABM and CAS leads to a conclusion that the former is a natural paradigm for modelling the latter. Because economic systems are a subclass of complex adaptive systems, ABM is a natural paradigm for modelling economy.

D.3.2 Traditional Economic Modelling vs. Agent-Based Modelling

For historical reasons, economics has used analytical modelling based on the physics of the 19th century. Consequently, traditional models sacrifice their ties with reality in favour of being mathematically soluble. There is a vast critical literature on neoclassical economics, which is beyond the scope of this work. An interested reader is referred to Beinhocker (2006) for an excellent review of historical peculiarities that were crucial to the development of neoclassical economics (but see also Nadeau, 2011). See Kirman (1989; 1992), Hoff and Stiglitz (2001) and Leijonhufvud (2009) for a more rigorous treatment of the resulting problems. The main flaws of traditional economic modelling are summarised and contrasted with ABM in Table D-3.

D.3.3 Agent-Based Modelling vs. Other Individual-Level Simulation Techniques

Some of the above-mentioned characteristics of ABM in its comparison with traditional mathematical modelling of economic systems could also be applied to other simulation techniques. It is, thus, necessary to compare and contrast ABM with some of the most popular other individual-level simulation approaches used in social and economic modelling. In individual-level simulations, researchers implement the so-called "bottom-up" modelling approach by first specifying individual-level behaviour and then observing macro-level results. Besides ABM, this class includes *microsimulation models* and *cellular automata models* (International Microsimulation Association, n.d.).

In pure ***cellular automata models***, the modelled entities operate on a grid and have a single attribute whose values depend deterministically on the values of this attribute of the neighbouring entities. Thus, spatial location of the entity plays an important role in determining its attribute values. While the dependence of attributes on grid neighbours might be considered as a model of network connections effects, it must be noted that typically, the neighbourhood that affects the agent is fixed and, thus, one cannot analyse evolving networks with cellular automata models.

Table D-3 – Comparison of Traditional Economic Modelling with Agent-Based Modelling

	Traditional Economic Modelling			**Agent-Based Modelling**
	Description	**Problem**		**Description**
Rationality	• Agents can figure out the future of the world, at least on average • They act to maximise their welfare	• Humans have a bounded capacity to figure out how to behave • Finding Walrasian/Nash equilibria is too computationally difficult for an ordinary human • The assumed rational behaviour lacks procedural/algorithmic bases • Human behaviour systematically departs from rationality requirements	**Bounded Rationality**	• Agents inspect local, not global, environment for possible utility gains • They take actions that they believe to lead to satisfactory, not necessarily optimal, outcomes, at least with high probability • They have little global information, and the information they have may be significantly out-of-date • They may be able to acquire up-do-date global information, but it can be costly • They have limited knowledge of their own preferences and understand them after trying different alternatives
Agent Homogeneity	• Agents are homogeneous or with very limited heterogeneity • This significantly improves analytical tractability, because one can now analyse the *representative agent*	• No plausible formal justification exists for the assumption that the aggregate of individuals acts itself like an individual maximizer • Representative agent's reaction to policy changes need not reflect those of individuals, and its preferences may be opposed to those of society • To comply with empirical tests showing complicated dynamics, the representative agent should possess very unnatural characteristics	**Agent Heterogeneity**	• Each agent is modelled individually: there is no representative agent • Agents have access to different resources and information • The behaviour of each agent is different from the average behaviour • Each agent acts in its unique environment

Table D-3 (cont.)

	Traditional Economic Modelling		Agent-Based Modelling	
	Description	Problem		Description
Lack of Interaction	• Agents interact indirectly, through aggregate economic variables like price vector or unemployment level • Agents do not set the levels of these aggregate variables	• This does not square with the way real-world individuals interact	**Interaction through Networks**	• Permits truly direct non-anonymous interactions • Makes the information local, hence agents: o Have locally purposive behaviour that can be either reinforced or not by the global environment o Cannot optimise over all states of the world • Allows us to use our knowledge on human behaviour, motivation and relationships in building these models
Agent-Level Equilibrium	• Only individual-level fixed-point equilibria are of interest • Typically, any fluctuations in the economy are assumed to be exogenous	• Such systems are *thoroughly static* in the von Neumann–Morgenstern sense: no agent has any incentive to unilaterally change • Thus, there is no mechanism for endogenously creating novelty • The source of economic change is sought in non-economic phenomena	**Emergence**	• Macro patterns emerge from agent micro-level behaviours and interactions • This opens the way for the evolutionary process to create novelty in the system • The systems under consideration are, thus, open and dynamic, not closed and static
Simplistic Assumptions	• Many simplifying assumptions are introduced for analytical tractability • This allows to easily summarise results through a comparison of fixed-point end states • Typical assumptions: linearity, homogeneity, normality, stationarity	• While these assumptions allow to solve the problems analytically, many of them fail empirical tests • Making wrong assumptions, even for mathematical tractability, leads to solving the wrong problem	**Complexity**	• The use of simulations allows to add more real-world complexity into the model without concerning the detrimental effects on analytical tractability

Source: compiled from Axtell (2000; 2007), Bankes (2002), Beinhocker (2006), Deckert and Klein (2010), Kirman (1992)

In pure ***microsimulation models*** (MSMs), entities have several attributes that change either deterministically or stochastically over time. By varying the distributions guiding transition probabilities of the attributes, one can model the effects of different policy changes. Network effects, however, are completely ignored in traditional MSMs. These models are also not useful when one needs to implement evolutionary changes in behaviour, when not only parameters of a fixed behaviour change (such as product preferences as a function of prices) but a completely new behaviour emerges. Moreover, as MSMs are traditionally used for policy evaluation, they focus on the effects of policies on individuals and ignore the feedback. Nevertheless, they have proved to be quite successful in replicating complex policy structures.

To sum up, ABM allows for a more detailed modelling of behaviour, adding rich interaction effects at the micro-level through networks and allowing for a feedback from the micro- to the macro-level. Overall, though, there is a trend towards uniting all three modelling approaches, thus, combining the strengths of each. See, in particular, a discussion in Birkin and Wu (2012).

D.3.4 Agent-Based Modelling vs. System Dynamics

System dynamics (SD) is another alternative to agent-based modelling. It is a computer simulation technique actively used for policy analysis and design in the context of social, managerial, economic or ecological systems (System Dynamics Society, n.d.).

SD models the system as consisting of stock and flow variables, the relationships among them, and time delays. ***Stock variables*** represent aspects of the system that accumulate (or deplete) over time, while ***flow variables*** represent changes in stock variables. Relationships in SD models are represented by ***causal loop diagrams***, which may be divided into positive (or reinforcement) feedback loops and negative (or balancing) feedback loops. In a ***positive feedback loop***, an increase in one variable ultimately, through the loop, leads to a further increase in it; in a ***negative feedback loop***, it leads to a decrease in it. The incorporation of feedback loops and delays in the propagation of information through the loops is what makes SD models able to capture the highly nonlinear dynamics of complex systems. Then external variables, which do not participate in any feedback loops, are added and connected to the relevant internal variables. The exact change in a variable resulting from the change in the variable it is connected to is given by a formula, which is derived through precise knowledge or the assumptions about how the system's elements function. (Sherwood, 2002)

Nevertheless, system dynamics has problems in simulating models with (North & Macal, 2007, p. 71):

- Strong spatial or geographical components
- Dynamically interacting networks of agents
- Discrete decision variables
- Constraints on decision variables

ABM, on the contrary, is able to incorporate these aspects successfully.

D.4 The Problems of Agent-Based Modelling

That ABM is a natural way to model CASs does not, however, mean that ABM is a panacea. Below, I mention its most cited drawbacks.

The main drawback of ABM, which is also typical for any simulations, is the so-called "curse of dimensionality" (Axtell, 2000). Recall that ideally, we have to run the model for all values of parameters in the parameter space. The time required for this, however, increases exponentially from the number of parameters.

Moreover, while agent-based models allow us to express the "enormous amount of data and knowledge about the behaviour, motivations and relationships of social agents" (Bankes, 2002, p. 7199), there are concerns that we do not understand the mechanisms of human behaviour well enough to be sure that the models we build are correct (Deckert & Klein, 2010).

Finally, ABM is still a very young field of science. Consequently, there is currently no standard way to construct and describe agent-based models, as well as to analyse the data stemming from simulation runs. Furthermore, models are rarely comparable, as they entail highly heterogeneous theoretical content, explain different phenomena and no formal tests are usually conducted to measure the relative performance of different models of the same phenomenon. (Fagiolo, Moneta, & Windrum, 2007; Richiardi, Leombruni, Saam, & Sonnessa, 2006)

APPENDIX E COMPUTATIONS

Proposition E.1 (Setting Time-Dependent Probability to Find Job] A temporal structure in the data is given by a set of shares of respondents with different unemployment spells, $\{f_t | 0 \le f_t \le 1 \wedge \sum_t f_t = 1\}, t = \overline{1, T}$. Given these, the probability to exit unemployment after spending t periods there is

$$p_t = \frac{f_t}{1 - \sum_{k=1}^{t-1} f_i}.$$

If the last share, f_T, refers to an open interval (i.e., the share of respondents who found a job after T or more periods) and it is assumed that $p_T = p_{T+1} = p_{T+2} = \cdots$ then $p_t \equiv 1/2$ for all $t \ge T$.

Proof. Assume that in period 0, the person lost job. In period 1, the person can find a job (exit unemployment) with probability p_1 or not find it (continue searching in period 2) with probability $(1 - p_1)$, and similarly in periods 2, 3 and so on. Empirically, we observe f_1, the share of persons with unemployment of length 1 period, f_2, the share of persons who found job after 2 periods, etc. The key to relating $\{p_i\}$ with $\{f_i\}$ is to understand that f_i is actually the probability that the person did not find a job *before* period i but found it *in* period i:

$$f_i = (1 - p_1)(1 - p_2) \dots (1 - p_{i-1}) p_i.$$

The expressions for p_i are then obtained inductively:

$$\begin{aligned} p_1 &= f_1 \\ p_2 &= \frac{f_2}{1 - p_1} = \frac{f_2}{1 - f_1} \\ p_3 &= \frac{f_3}{(1 - p_1)(1 - p_2)} = \frac{f_3}{1 - f_1 - f_2} \\ &\dots \end{aligned}$$

Now assume that the last empirical observation, f_T, is the share of those exiting unemployment after T or more periods and $p_T = p_{T+1} = p_{T+2} = \cdots = p$. Denote the probability to stay unemployed for $(T - 1)$ consecutive periods by P. Then, assuming infinite life of a person, we have the following relationship:

$$\begin{aligned} f_T &= Pp + P(1 - p)p + P(1 - p)^2 p + \cdots \\ &= P \sum_{i=0}^{\infty} (1 - p)^i p \\ &= P \frac{p}{1 - p}, \end{aligned}$$

from where we get

$$p = \frac{f_T}{f_T + P}. \tag{E-1}$$

It is easy to see that this expression is equal to $1/2$. Indeed, consider an event that the agent finds a job in period $T + 1$, $T + 2$ or later. The probability that this event happens is f_T. But the fact that this event happens is equivalent to the event that in the first 12 periods, the agent was unemployed. The probability of this event is P. Hence, $P \equiv f_T$ and (E-1) evaluates to $1/2$. ■

Proposition E.2 (Setting Time-Independent Probability to Lose Job Given Time-Dependent Probabilities to Find Job) Given a set of time-dependent probabilities to find job, $\{p_i\}$, and the unemployment rate, π_U, the time-independent probability to lose job is

$$p = \left(1 - \sqrt[k]{(1-p_1)(1-p_2)\dots(1-p_k)}\right)\frac{\pi_U}{1-\pi_U},$$

where k is such that $(1-p_1)(1-p_2)\dots(1-p_k) \approx 50\%$.

Proof. Consider a two-state Markov chain where state E corresponds to employment and state U to unemployment. Denote the probability of moving from E to U by q, remaining in E by $(1-q)$, remaining in U by $(1-p)$ and moving from U to E by p. From the steady-state equations describing the probabilities of being in state U, π_U, and in state E, $\pi_E = 1 - \pi_U$, one could find q, given p, by the following equation:

$$q = p\frac{\pi_U}{1-\pi_U}. \quad \text{(E-2)}$$

The value of π_U is the observed unemployment rate (among those active in the labour market). The problem is that we do not have a single probability of exiting unemployment, but instead we have a set of probabilities depending on the length of staying in unemployment. One approach to tackle this problem is to take an *average* probability of leaving unemployment to represent p. This average probability is computed as follows. Given the set of p_i (computed as shown in Proposition E.1), I get the number of periods until the cumulative probability of continuing being unemployed is around 50%. In other words, I compute k such that

$$(1-p_1)(1-p_2)\dots(1-p_k) \approx 50\% .$$

This is the average number of periods spent in unemployment. Then the average probability of *not* leaving unemployment is the average probability of not leaving unemployment in these first k periods, or

$$1 - p = \sqrt[k]{(1-p_1)(1-p_2)\dots(1-p_k)},$$

from where it is easy to get p. It remains to input p and π_U into (E-2) to get q. ■