

Angela Djupsjöbacka

Human Capital

Formation, Maintenance and Transmission



Åbo Akademi University Press
Tavastgatan 13, FI-20500 Åbo, Finland
Tel. +358 (0)2 215 4793
E-mail: forlaget@abo.fi

Sales and distribution:
Åbo Akademi University Library
Domkyrkogatan 2–4, FI-20500 Åbo, Finland
Tel. +358 (0)2 -215 4190
E-mail: publikationer@abo.fi

HUMAN CAPITAL



Human Capital

Formation, Maintenance and Transmission

Angela Djupsjöbacka

Åbo Akademis förlag | Åbo Akademi University Press
Åbo, Finland, 2020

CIP Cataloguing in Publication

Djupsjöbacka, Angela.

Human capital : formation, maintenance
and transmission / Angela Djupsjöbacka. -
Åbo : Åbo Akademi University Press,
2020.

Diss.: Åbo Akademi University.

ISBN 978-951-765-959-8

ISBN 978-951-765-959-8
ISBN 978-951-765-960-4 (digital)
Painosalama Oy
Åbo 2020

Contents

<i>Abstract, Sammanfattning & Acknowledgements</i>	iii
CHAPTER 1: The Mystery	1
References_	11
Appendix	12
CHAPTER 2: A Simple and Consistent Alternative for Estimating Ordered Response (and Regressor) Models	15
2.1 Introduction	16
2.2 Adjusted-POLS	18
2.3 A latent independent variable	20
2.4 Simulations: Adjusted-POLS vs Ordered probit	22
2.5 Interpretation	28
2.6 Conclusions	30
References	30
Appendix	31
CHAPTER 3: Higher Education and the Gains in Cognitive Abilities	48
3.1 Introduction	49
3.2 Background and previous research	51
3.3 Data and descriptive statistics	54
3.4 The instrumental variables approach	57
3.5 A country-level analysis	67
3.6 The trend in cognitive skills among university students	73
Discussion	76
3.8 Conclusions	78
References	79
Appendix	81

CHAPTER 4: Some Evidence for a Cognitive	
Decline from Leaving School in a Recession	88
4.1 Introduction	89
4.2 Background and previous research	91
4.3 Empirical strategies	94
4.4 Data and sample selection	98
4.5 Results	102
4.6 Validity	112
4.7 Conclusions	116
References	117
Appendix	122
CHAPTER 5: The Effect of Retiring on Cognitive	
Functioning and Subjective Health	131
5.1 Introduction	132
5.2 Background an previous research	133
5.3 Data and descriptive statistics	136
5.4 Instrumenting retirement and retirement duration	142
5.5 Conclusions	156
References	158
Appendix	161
CHAPTER 6: Socioeconomic Background and	
College Application Strategies	167
6.1 Introduction	168
6.2 Modeling portfolio choices	169
6.3 Data description	176
6.4 Evidence on differences in applicant behavior by	
neighborhood education level	181
6.5 Conclusions	192
References	193
Appendix	194

Abstract

Human capital, as measured by your text-based and numerical problem-solving skills, varies between countries as well as between cohorts of individuals from the same country. The forces driving these patterns are poorly understood, however. This thesis deals with topics related to the development and maintenance of human capital in different stages of life. The research questions include: To what extent does schooling contribute to the development in students' problem-solving skills (chapter 3)? Does graduating in a slow economy affect your future cognitive performance (chapter 4)? Does retirement influence the age-trajectory in cognitive functioning among the elderly (chapter 5)? This thesis also addresses a somewhat different but related topic, by analyzing college application behaviors of high school graduates, and how these behaviors differ depending on socioeconomic background (chapter 6). Chapter 2 is the odd one out, where I present the adjusted-POLS estimator. This is a variant on the POLS-estimator, suggested by Praag & Ferreri-i-Carbonell, and it is used for estimating ordered response models. In addition, I also suggest an estimator for a marginal effect, when controlling for a latent independent variable.

In order to answer these questions, I mainly rely on survey data from the The Programme for the International Assessment of Adult Competencies (PIAAC) which covers tens of thousands of individuals from different OECD-countries (including Russia). The identification regularly relies on exploiting different instruments as exogenous stimuli in the variable of interest. Based on the data, I argue for the following main conclusions: 1) Post-primary schooling has meaningful effects on cognitive performance later in life. The performance gap between university students and 'others' is, however, best explained by self-selection. 2) Graduating in a bad economy may have negative impacts on your future cognitive performance. This effect, however, is likely to be economically small. 3) Retiring may slow down the age-related decline in cognitive performance among men. For women, the evidence points towards a positive effect on self-perceived health. 4) When applying to college, students gravitate towards colleges located nearby, and this tendency is somewhat stronger for those from poorly educated neighborhoods. Women from poorly educated neighborhoods apply in a way that is consistent with being comparatively wage risk averse. 5) Adjusted-POLS can be a useful alternative for estimating models with ordered responses in situations where ordered probit or logit are inappropriate or inconvenient.

Sammanfattning

Humankapital – mätt som förmågan att lösa textbaserade och numeriska problem – varierar mellan länder och mellan kohorter av individer från samma land. Orsakerna bakom dessa mönster är fortfarande oklara. Den här avhandlingen behandlar humankapital, dess utveckling och bevarande under livets olika skeden. Till forskningsfrågorna hör: Till vilken grad bidrar utbildning till att utveckla individens problemlösningsförmåga (kapitel 3)? Hur påverkas individens kognitiva prestationsförmåga av att ta examen under en lågkonjunktur (kapitel 4)? Kan pensionering påverka ålderstrenden i kognitiv prestationsförmåga bland äldre (kapitel 5)? Denna avhandling behandlar också ett annat relaterat ämne genom att analysera ansökningsstrategier till högre utbildning, och hur dessa strategier skiljer sig beroende på socioekonomisk bakgrund (kapitel 6). Kapitel 2 avviker från de övriga; här presenterar jag estimatorn 'justerade-POLS' som används för att estimerar modeller där utfallet mäts på ordinalnivå. 'Justerade-POLS' är en variant av POLS, som föreslagits av Praag & Ferreri-i-Carbonell. Vidare ger jag också förslag på en estimator som kan användas för att mäta en marginaleffekt då vi kontrollerar för en latent oberoende variabel.

För att besvara dessa frågor använder jag främst enkätdata från 'The Programme for the International Assessment of Adult Competencies' (PIAAC). Detta datamaterial täcker tusentals individer från olika OECD-länder inklusive Ryssland. Identifikationen bygger ofta på instrumentvariabeltekniken, dvs. jag använder instrument i syfte att utnyttja exogen variation i den oberoende variabeln av intresse. Baserat på data drar jag följande slutsatser: 1) Eftergrundskoleutbildning har betydelsefulla effekter på kognitiv förmåga senare under livet. Skillnaden i prestationer mellan högskolestuderande och övriga förklaras dock bäst av självselektion. 2) Det kan finnas negativa kognitiva effekter av att ta ut examen i en lågkonjunktur. En sådan effekt är dock sannolikt ekonomiskt liten. 3) Pensionering kan potentiellt motverka den negativa ålderstrenden i kognitiv förmåga bland män; bland kvinnor pekar data på att pensionering gynnar självupplevd hälsa. 4) Individen har en tendens att söka sig till utbildningar som ligger nära hemmet. Denna tendens är något starkare bland dem som kommer från lågutbildade områden. Kvinnor från lågutbildade områden har en ansökningsstrategi som överensstämmer med en relativt hög löneriskaversion. 5) 'Justerade-POLS' kan vara ett användbart alternativ till 'ordinal probit' och logit i situationer där dessa estimatorer är olämpliga eller opraktiska.

Acknowledgements

Thanks to the Academy of Finland for financial support and to everyone who has guided me along the way, not least my colleagues at the Economics department at Åbo Akademi University (both present and former ones). A special thanks to Jonas Lagerström, Eva Österbacka, Roope Uusitalo and Gunilla Widén. I also want to thank my family and friends for their relentless support.

CHAPTER 1

The Mystery

Our cognitive functioning is of personal interest to most of us. To economists, however, human capital formation is important also for other reasons, for example due to its central role as a driver of economic growth. In the standard Solow growth model, labor-augmented technological change is the sole cause of long run economic growth. Hence, there should be few questions in economics that triumph over this one in importance: What, then, are the drivers of human capital growth?¹

In order to present some initial hypotheses, I assume a simple idea in the spirit of Becker: Your cognitive problem-solving skills (cog) are a function of practice, innate ability and other factors (such as age):

$$\text{cog} = f(\text{practice, innate ability, other factors})$$

Hence, your problem-solving skills are partly determined by how many problems you have already tried to solve; the more you practice, the better you become. It is reasonable to assume diminishing returns on practicing however, i.e. each additional hour contributes less than the previous one. Also, it seems reasonable to assume that innate abilities also contribute to the variation in cognitive skills. If we, for the sake of simplicity, assume a constant marginal cost of time spent practicing, then the individual chooses an optimal level where the marginal return meets this cost.

Here, this simple idea is used as a foundation in order to form a hypothesis regarding one of the greatest mysteries in the data: Why are individuals from some countries, such as Japan and Finland, performing so much better on cognitive tests than individuals from some other OECD-countries? I study this question using survey data from The Programme for the International Assessment of Adult Competencies (PIAAC). This is a pooled cross-section covering roughly 200,000 individuals in the age range of 16-65 from 31 OECD-countries (including Russia). The survey was conducted during 2010–2015 using personal house interviews. The variables of key interest are literacy and numeracy, measuring your text-based and mathematical problem-solving skills. Below follows a description of these measures.

¹ Labor-augmented technological change may include other factors than human capital growth. From a policy perspective, human capital is arguably one of the most interesting, however, given its potentially endogenous character.

Literacy and numeracy

Literacy is defined as “the ability to understand, evaluate, use and engage with written texts to participate in society, to achieve one’s goals, and to develop one’s knowledge and potential.” Numeracy is defined as “the ability to access, use, interpret and communicate mathematical information and ideas in order to engage in and manage the mathematical demands of a range of situations of adult life.” (OECD, 2016a)

In comparison to conventional intelligence tests, the questions on literacy and numeracy are practical in nature and resemble the kinds of problems you face in everyday life (such as reading off a thermometer or data chart, transforming units of measurement, or interpreting written guidelines and rules).² The testing procedure is based on item response theory and multiple imputations; each individual answers a subset of questions and a distribution over test scores is constructed. I have access to ten draws or ‘plausible values’ from each individual distribution on literacy and numeracy. I use these, collectively, in all estimations and account for the added imputation variance.

The distributional properties of literacy and numeracy have been described in detail by OECD Skills studies (OECD, 2016b) and they have been found to be strong predictors of productivity and wages (Hanushek et al., 2015; OECD, 2016b).

Table 1 below describes average literacy and numeracy by country. Japan, Finland and the Netherlands are in the top for both measures; Italy, Spain, Turkey and Chile are in the bottom.³ All numbers on literacy and numeracy are standardized z-values. Example: A value at 0.1 means that you fall 10 percent of a standard deviation above the OECD average, i.e. the expected value for a randomly chosen individual from the PIAAC target population. The standard deviation is measured using within-country variation only. All estimates account for the country-specific survey designs.

² See OECD, 2016a for example questions.

³ Section A.1 in the Appendix presents the corresponding list for natives only. This has some effect on the ranking. For example, Sweden has now moved to 1st place on numeracy.

Table 1. Average numeracy and literacy by country

Country	NUMERACY		LITERACY	
	Rank	Average	Rank	Average
Japan	1	0.47	1	0.61
Finland	2	0.37	2	0.42
Netherlands	3	0.33	3	0.36
Belgium	4	0.32	9	0.16
Sweden	5	0.29	5	0.25
Denmark	6	0.29	15	0.08
Norway	7	0.29	6	0.24
Czech Republic	8	0.24	10	0.15
Austria	9	0.23	19	0.04
Slovak Republic	10	0.23	11	0.14
Estonia	11	0.19	7	0.20
Germany	12	0.16	16	0.06
New Zealand	13	0.15	4	0.28
Russia	14	0.13	8	0.17
Lithuania	15	0.08	21	-0.00
Canada	16	0.04	12	0.13
Cyprus	17	0.03	18	0.05
Korea	18	-0.00	13	0.12
United Kingdom	19	-0.02	14	0.11
Poland	20	-0.06	22	-0.00
Singapore	21	-0.10	24	-0.20
Slovenia	22	-0.11	25	-0.22
Ireland	23	-0.13	20	0.00
France	24	-0.18	23	-0.11
United States	25	-0.20	17	0.06
Greece	26	-0.21	27	-0.28
Israel	27	-0.24	26	-0.25
Italy	28	-0.31	29	-0.35
Spain	29	-0.33	28	-0.32
Turkey	30	-0.85	30	-0.88
Chile	31	-1.10	31	-1.00

Sample size by country: Japan 5,173; Finland 5,464; Netherlands 5,083; Belgium 4,984; Sweden 4,469; Denmark 7,286; Norway 4,947; Czech Republic 6,081; Austria 5,025; Slovak Republic 5,702; Estonia 7,586; Germany 5,379; New Zealand 6,074; Russia 3,892; Lithuania 5,051; Canada 26,683; Cyprus 4,392; Korea 6,651; United Kingdom 8,806; Poland 9,366; Singapore 5,393; Slovenia 5,293; Ireland 5,963; France 6,907; United States 4,898; Greece 4,916; Israel 5,344; Italy 4,589; Spain 5,971; Turkey 5,194; Chile 5,192

Table 1 reveals that the gap between different countries can be large indeed. Example: Only 8 percent of Chileans reach the Japanese average on numer-

acy.⁴ Similarly, if you perform at least one standard deviation above average, then it is roughly nine times more likely that you are Japanese than Chilean (here I assume that the countries are equally large).⁵

In the Appendix (see section A.2) I also present some other key statistics related to literacy and numeracy, including gender differences and age trends.

Why do some countries perform better than others do?

Sweden performs clearly better than the United Kingdom. On average, Swedes perform 0.48 standard deviations above the OECD average on numeracy; for the United Kingdom, this number is 0.05 standard deviations. In these samples, I only include natives.

So why does Sweden perform so well? There are several possible hypotheses. In order to sort them out, let us start by comparing the distributions: Is Sweden overrepresented by ‘militarily trained prodigies’? Or is Sweden evenly better across the scale? Or is Sweden, to the contrary, underrepresented by poorly performing individuals?

Data shows that the last alternative is closest to the truth: Sweden maintains a comparatively high level at the bottom. Sweden is probably a good country for those with weak innate abilities; otherwise, the country of residence is, perhaps, of less importance. This is shown in table 2a below. Sweden performs 0.26 standard deviations above the United Kingdom at the 90th percentile. This is a significant difference, but the largest difference is found at the bottom. If you, as a Swede, perform worse than 90 percent of your countrymen, then you score 0.65 standard deviations below the OECD average. In the United Kingdom, the corresponding number is 1.21 standard deviations. Hence, the gap at the bottom of the distributions is more than twice as large as the gap at the top.

⁴ For Chile: The 91.75th percentile is 0.47.

⁵ The probability of performing at least one standard deviation above average is 2.95 % in Chile and 26.99 % in Japan.

Table 2a. The numeracy distribution comparing Sweden to the UK

	Sweden	United Kingdom	Difference
5th percentile	-1.02	-1.61	0.59
10th percentile	-0.65	-1.21	0.56
Median	0.50	0.08	0.43
90th percentile	1.58	1.33	0.26
95th percentile	1.89	1.63	0.27

Notes: Natives only.

This pattern is not exclusive to Sweden and the United Kingdom, but can be seen in a wide country comparison: The main difference between highly and poorly performing countries can be found in the bottom of the distributions. This is shown in table 2b below. Here I divide the 31 countries into two groups: the top 16 performers and the bottom 15. The gap between these groups is the largest at the 5th and 10th percentiles; the gap is comparatively modest at the 90th and 95th percentiles.

Table 2b. The numeracy distribution comparing top and bottom countries

	Top 16 countries	Bottom 15 countries	Difference
5th percentile	-1.24	-2.03	0.79
10th percentile	-0.86	-1.55	0.69
Median	0.35	-0.16	0.51
90th percentile	1.40	1.02	0.39
95th percentile	1.69	1.32	0.37

Notes: Natives only.

Figure 1 below illustrates the same phenomena. The x-axis shows the percentile gap, i.e. the distance between the 90th and 10th percentile; the y-axis shows the average performance on numeracy. Countries with large percentile gaps also perform poorly on average, $r = -0.63$; $p < 0.01$.

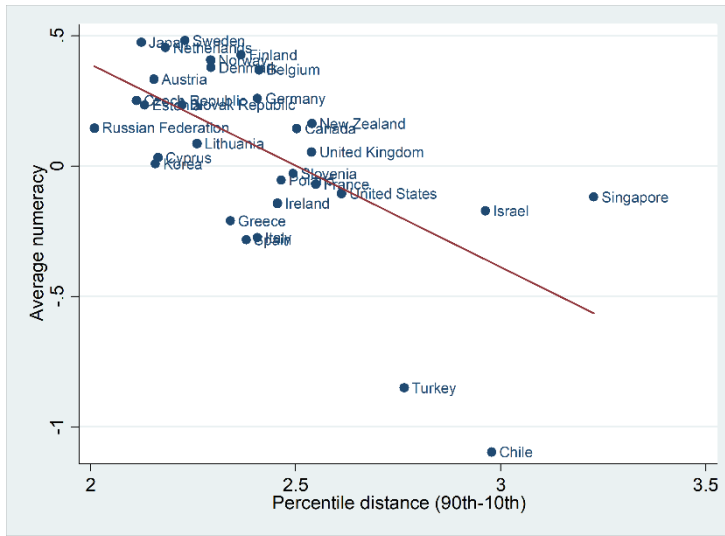


Figure 1. The relationship between the percentile distance and average numeracy

Notes: The percentile distance is the gap between the 90th and 10th percentiles on numeracy. Sample: Natives only.

Hence, an equal country is also performing well on average. Why should this be the case? One possible explanation is diminishing returns to training: The more you practice on increasing your problem-solving skills the better you become, but each additional hour has a lower return than the previous one. In order to attain a high average performance level, a country can either increase the overall training time of its citizens, or then redistribute this training time from those with the lowest marginal returns to those with the highest. In practice, this would probably mean a relatively even distribution of training time.

This idea fits well with the data. Here, I estimate average numeracy (*num*) as a function of average schooling (*mean*, measured in years) and the standard deviation for schooling (*sd*):

$$\widehat{num} = -0.62 + 0.13mean - 0.34sd \quad R^2 = 0.48$$

Both effects are significant ($p < 0.01$) and both variables have similar predictive power.⁶ In other words, schooling is a strong predictor of numeracy, but equally important is an equal *distribution* of schooling.

⁶ I test this by keeping one of the independent variables constant while making predictions using the other one. For both variables, the correlation between the predictions and the outcome is 0.56.

One may ask why such a ‘redistribution of training time’ is not selfregulating. Suppose there is a constant marginal cost related to training, and that individuals train until it no longer pays off, i.e. until the marginal return meets this cost. Depending on innate ability, this optimal level would certainly be achieved sooner for some individuals than others. This kind of inequality in training time is still efficient, however. The inefficient inequality occurs when we introduce ‘artificial’ variation in training time, i.e. variation that is not explainable by differences in ability. Driving factors could include differences in upbringing, school segregation or financial restrictions on higher education.

Are there any empirical signs for such structures? In order to analyze this question, I measure the degree of *intergenerational mobility* for each country. Example: 4.3 percent of the variation in schooling is explainable by parental education levels for Sweden: $R^2 = 4.3\%$. Here, the degree of intergenerational mobility is defined as 100 minus this coefficient: $100 - 4.3 = 95.7$. In the United Kingdom, the corresponding coefficient of determination is 9.0 percent and the intergenerational mobility is 91.0 percent.⁷ Hence, the measure for intergenerational mobility can take on values between 0 and 100, where 0 describes ‘no mobility’ and 100 is ‘total mobility’. Now, with a low intergenerational mobility, your training time is, presumably, relatively strongly restricted by your socioeconomic background. Naturally, one would expect some correlation in education between parents and their children in any case – for example due to their shared genetics – but the variation that arises *between* countries is likely to reflect institutional differences.

The scatterplot in figure 2 below illustrates the relationship between the intergenerational mobility and the average performance on numeracy. The correlation is striking; countries where the intergenerational mobility is high are also the countries where average performance is high, $r = 0.72$; $p < 0.01$.

⁷ The coefficient of determination is measured in the following way: First, I measure parental education levels in six categories: (1) Both parents have a basic/unknown education, (2) One parent has a secondary degree, the other one has a basic/unknown education, (3) Both parents have a secondary degree, (4) One parent has tertiary degree, the other one has a basic/unknown education, (5) One parent has a tertiary degree, the other one has a secondary degree, and (6) Both parents have a tertiary degree. The coefficient of determination is retrieved from a regression where ‘years of schooling’ is predicted by parental education levels included using a set of dummy variables. (I do this regression separately for each country.) Note that countries with a low variation in parental education levels also get lower coefficients of determination (everything else equal). In a second step, I construct a coefficient of determination that corrects for this, i.e. I calculate R^2 as if each parental education level were equally common. This makes little practical difference, and the results are omitted here.

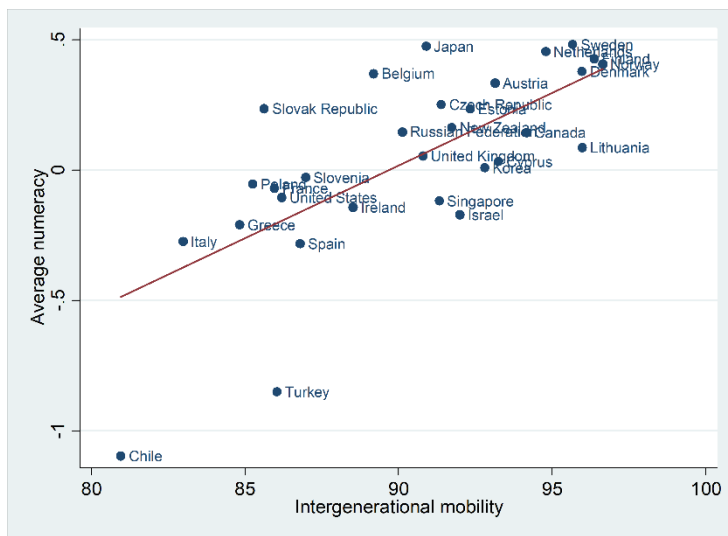


Figure 2. The relationship between intergenerational mobility and average numeracy

Notes: Intergenerational mobility is a measure that, theoretically, can take on values between 0 and 100 with 100 describing total mobility; 0 describing no mobility. Sample: Natives only.

This correlation is *not* explainable by ‘schooling’ functioning as a confounding factor. When comparing two countries – A and B – both with an equal amount of average schooling, but where A has ‘10 percentage points higher intergenerational mobility’, then the citizens of country A are predicted to perform half of a standard deviation above those in country B:

$$\widehat{num} = -5.60 + 0.10mean + 0.05mob \quad R^2 = 0.60$$

where *mean* is average schooling and *mob* is intergenerational mobility. Both of these effects are significant ($p < 0.05$ & $p < 0.01$) but the intergenerational mobility has stronger predictive power.⁸

Hence, one natural interpretation of the data is that ‘equal opportunities’ is an important part of the equation. It is worth noting, however, that the analysis presented in this chapter is exploratory in nature. In the remainder of this thesis, I dig deeper into the causal mechanisms underlying human capital formation and its maintenance. I start by looking into the role of higher education (chapter 3); continue onto the role of labor market conditions (chapter 4) and end with retirement (chapter 5). In chapter 6, I

⁸ I test this by keeping one of the independent variables constant while making predictions using the other one. For intergenerational mobility, the correlation between the predictions and the outcome is 0.72. For average schooling, the correlation is 0.56.

study a related – but somewhat different issue – college application strategies. Chapter 2 is the odd one out, where I evaluate the adjusted-POLS estimator, which I also apply in a subsequent chapter. Below follows a short summary of the main conclusions from each chapter. In these descriptions, I often substitute ‘literacy’ and/or ‘numeracy’ for cognitive performance, for short, although it is worth being aware of the distinction.

A simple and consistent alternative for estimating ordered response (and regressor) models

This chapter presents the adjusted-POLS estimator, which is a variant of the POLS-estimator suggested by Praag & Ferreri-i-Carbonell. Adjusted-POLS is a useful alternative for estimating models with ordered responses, such as self-perceived ‘health’, ‘job satisfaction’ or ‘happiness’. I show that adjusted-POLS is consistent, and that it generally performs on par with ordered probit in a cross-sectional setting where the distributional assumptions of both estimators are met. In comparison to probit or logit, however, adjusted-POLS is simple to combine with other linear estimation techniques – such as fixed effects or two-stage least squares – and estimates are easy to interpret. In addition, I also present a consistent estimator for a marginal effect, when controlling for a latent *independent* variable.

Higher education and the gains in cognitive abilities

In this chapter, I argue for the following two main conclusions: 1) Post-primary schooling has positive effects on cognitive performance later in life. 2) The *immediate* cognitive gains of university studies are likely to be rather modest, at least for individuals who choose to study at this level. I build my argument on different pieces of empirical evidence. Firstly, I exploit yearly variations in cohort sizes as an instrument for schooling. Example: Some countries experienced a rising number of births in the early 1960s, but as these individuals reached adulthood, the number of college and university spots seems to have lagged behind. I find that individuals born into ‘unpopular’ cohorts have longer education – and significantly higher cognitive performance – later in life. I interpret this cognitive effect as the result of education, although I do admit that other interpretations are possible. Secondly, I measure the trends in cognitive performance in the years following graduation from upper-secondary school, comparing university students to others. I find that the trend for those who enroll in a university program is similar to the trend for those who move into the workforce or other activities. This suggests that university studies have small (if any) effects on cognitive performance as measured by literacy and numeracy. The

data is a cross-section, however, meaning that I cannot exclude the possibility of confounding cohort effects.

Some evidence for a cognitive decline from leaving school in a recession

Several recent studies have shown that graduating in a bad economy has large and persistent adverse labor market effects. In this chapter, I argue that graduating in a bad economy may also affect your future cognitive performance. This effect, however, is likely to be economically small. I build my argument on a couple of complementary pieces of evidence. Firstly, students who graduate in a slow economy perform marginally (but significantly) worse later on. This effect persists after instrumenting the unemployment rate at graduation. Also, the unemployment rate in the year(s) following graduation is successful in predicting your future cognitive performance, conditional on the unemployment rate at graduation.

The effect of retiring on cognitive functioning and subjective health

Official retirement ages are now increasing in several OECD-countries. The effects on the cognitive and physical health of the elderly is unclear, however. In this chapter, I argue that retiring may slow down the age-related decline in cognitive performance among men. For women, the evidence points towards a positive effect on self-perceived health. I base these conclusions on the following data patterns: Firstly, I observe a significant trend-break in literacy as men reach the official retirement age of their country. I interpret this trend-break as the result of retirement. This conclusion is strengthened by the fact that I observe no such trend-break when creating 'fake' retirement ages by moving the real ones five years into the past. Similarly, for women I observe a significant discontinuity in self-perceived health: those who have reached the retirement age score their health significantly above that of their younger counterparts (controlling for age, among other factors).

Socioeconomic background and college application strategies (in cooperation with Roope Uusitalo)

In this chapter, we analyze college application behaviors of Finnish high school graduates, and how these behaviors differ depending on socioeconomic background as measured by the education level in your local neighborhood. By estimating a discrete choice model for portfolio choices, we make the following main arguments: 1) Students gravitate towards colleges located nearby, and this tendency is somewhat stronger for those

from poorly educated neighborhoods. 2) Women from poorly educated neighborhoods apply in a way that is consistent with being comparatively wage risk averse. We find little support, however, for the notion that these differences in application strategies would cause any large wage gaps between the student groups, i.e. those from poorly and highly educated neighborhoods.

My contribution: The research question, identifying its importance and choosing the methodology – a random expected utility approach within the framework of a conditional logit model – is not my own. My contribution lies mainly in the practical execution of this idea.

References

- Hanushek, E. A., Schwerdt, G., Wiederhold, S., & Woessmann, L. (2015). Returns to skills around the world: Evidence from PIAAC. *European Economic Review*, 73, 103-130.
- OECD (2016a). *The Survey of Adult Skills: Reader's Companion, Second Edition*, OECD Publishing, Paris.
- OECD (2016b). *Skills Matter: Further Results from the Survey of Adult Skills*, OECD Skills Studies, OECD Publishing, Paris

Appendix

A.1 A country comparison using natives only

Table A1. Average numeracy and literacy by country, natives only

Country	NUMERACY		LITERACY	
	Ranking	Average	Ranking	Average
Sweden	1	0.48	4	0.45
Japan	2	0.48	1	0.61
Netherlands	3	0.45	3	0.48
Finland	4	0.43	2	0.49
Norway	5	0.41	5	0.35
Denmark	6	0.38	12	0.17
Belgium	7	0.37	9	0.22
Austria	8	0.33	18	0.13
Germany	9	0.26	14	0.16
Czech Republic	10	0.25	15	0.16
Slovak Republic	11	0.24	16	0.15
Estonia	12	0.23	8	0.26
New Zealand	13	0.16	6	0.33
Russia	14	0.15	10	0.18
Canada	15	0.14	7	0.26
Lithuania	16	0.09	21	0.01
United Kingdom	17	0.05	11	0.18
Cyprus	18	0.03	19	0.07
Korea	19	0.01	17	0.13
Slovenia	20	-0.03	24	-0.16
Poland	21	-0.05	22	-0.00
France	22	-0.07	23	-0.01
United States	23	-0.10	13	0.17
Singapore	24	-0.12	26	-0.19
Ireland	25	-0.14	20	0.02
Israel	26	-0.17	25	-0.16
Greece	27	-0.21	28	-0.26
Italy	28	-0.27	29	-0.30
Spain	29	-0.28	27	-0.25
Turkey	30	-0.85	30	-0.87
Chile	31	-1.10	31	-1.01

A.2 Literacy and numeracy – descriptive statistics

The overall distribution

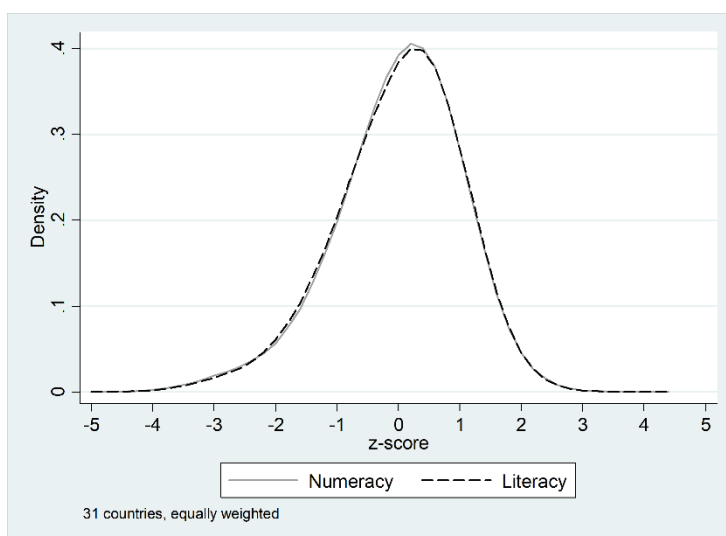


Figure A1. Distributions for literacy and numeracy

These distributions differ to some extent between the sexes, with women scoring 0.21 standard deviations below men on numeracy. This gender difference is greater at the top than at the bottom. At the 10th percentile, women score 0.19 standard deviations below men; at the 90th percentile, the corresponding gap is 0.26 standard deviations. There is no gender difference in literacy worth mentioning, however.

Age trends

Figure A2 below describes the trend in literacy and numeracy over the ages (16-65). From the age of 16 and above, performance grows rather rapidly until the age of ~30 after which performance starts declining. Note, however, that this is a cross-section and not a longitudinal dataset. Hence, there may be relevant differences between the cohorts besides age alone (e.g. education).

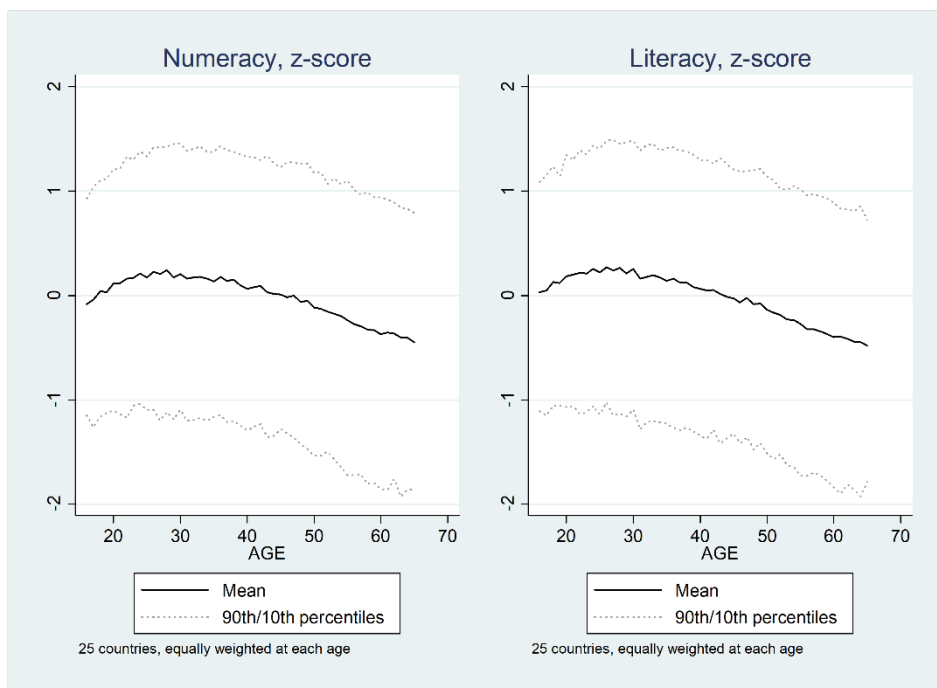


Figure A2. Age-trends in numeracy and literacy

CHAPTER 2

A Simple and Consistent Alternative for Estimating Ordered Response (and Regressor) Models

Abstract

The POLS-estimator – suggested by Praag & Ferreri-i-Carbonell – is a simple way of estimating a linear regression model with a latent outcome variable, assuming we observe a collapsed ordinal-scale version of this variable. The estimator can be described in two steps: 1) Estimate the expected standardized score for each observation, conditional on the ordinal score: $E(\text{Standardized latent variable} | \text{ordinal score})$. 2) Estimate your model using this expectation as the outcome. Here, I also add a third step: 3) Divide the obtained coefficients by the variance for the expected standardized scores. I call this estimator adjusted-POLS and show that it is consistent. Furthermore, my simulations suggest that adjusted-POLS generally performs on par with ordered probit in a cross-sectional setting where the distributional assumptions of both estimators are met. Most importantly, however, adjusted-POLS is easily combined with other linear estimation techniques, such as fixed effects or two-stage least squares. In addition, this chapter also presents a consistent estimator for a marginal effect, when controlling for a latent *independent* variable.

2.1 INTRODUCTION

Working with ordinal response variables can be frustrating. Firstly, the parameters of ordered probit and logit models are difficult to interpret in terms of effect sizes. Also, it can be challenging to combine these estimators with other statistical techniques, such as fixed effects, instrumental variables or multivariate techniques.

One way to get around these problems is to quantify your ordinal variable, i.e. set the ‘appropriate’ distances between categories. But how would you do that? A couple of authors have made suggestions, which includes methods such as rident scoring⁹ (Bross, 1958), the conditional median¹⁰ (Brockett, 1981) and the conditional mean or the POLS-estimator. The last method has been suggested by several authors (see Brockett, 1981; Fielding 1997; Terza, 1987; and van Praag & Ferreri-i-Carbonell, 2006 who also named it POLS).¹¹ The idea behind POLS is the following: We start by assuming that there exists a latent version of our ordinal variable, and that this latent variable has a known distribution. Now we can calculate your expected value on that latent variable, given your ordinal score. At this point, we have ‘quantified’ the distances between categories and life goes back to normal, i.e. we can use ordinary linear estimation techniques. But there is one problem: the censoring will bias your estimate, potentially quite severely. The new ‘quantified variable’ is measured with error and this error is correlated with the independent variables, assuming they have predictive power. In section 2, I show that this bias can be eliminated by dividing the coefficients with the variance for the ‘quantified’ variable. Below, I call this estimator adjusted-POLS.

Now, if we assume that the latent variable is normally distributed, then adjusted-POLS is essentially estimating the same model as ordered probit (after appropriate rescaling).¹² This brings us to the question: How well does

⁹ The rident score is your expected percentile rank given your ordinal score: $R_j = 0,5(\pi_{j-1} + \pi_j)$ where R_j is the rident score corresponding to value j on the ordinal variable; π_j is a cumulative probability, i.e. the probability for this ordinal value, or a lower one.

¹⁰ For ordinal value j , the conditional median is given by the inverse distribution function, F , evaluated at rident score j : $F^{-1}(R_j)$ where F is an appropriate function as chosen by the researcher.

¹¹ Terza (1987) suggested this ‘quantification’ specifically for the case when the ordinal variable is a regressor; van Praag & Ferreri-i-Carbonell suggested it specifically for the case when the ordinal variable is an outcome. Hence, they named it POLS which is short for Probit OLS.

¹² The ordered probit model has a normally distributed error term; the adjusted-POLS model has a normally distributed latent variable. However, if the structural part and the error term are independently distributed, then adjusted-POLS also has a normally distributed error term (this is a result of Cramér’s decomposition theorem). The scaling of the model parameters differ, however: While ordered probit anchors the error variance, adjusted-POLS anchors the latent variable variance.

adjusted-POLS perform in comparison to ordered probit? The simulations presented in section 4.1 suggest that adjusted-POLS generally performs on par with ordered probit when the distributional assumptions of both estimators are met. You can find cases that slightly favors one estimator over the other, but any differences in performance are likely to be small in practice.

The main advantage of adjusted-POLS presents itself in situations where ordered probit is an inconvenient or inconsistent estimator. When I compare the estimators in a panel data setting this advantage becomes apparent (see section 4.2).¹³ Here, ordered probit is plagued by the so called ‘incidental parameters problem’ which becomes an issue when we would like to estimate fixed effects models (see Neyman and Scott, 1948). Adjusted-POLS, on the other hand, performs well also for short panel datasets. This is true when we describe the coefficients in their original unit of measurement, or when we use the estimated model to calculate average treatment effects, i.e. the average change in the probability of a specific outcome as x increases from one fixed number to another.¹⁴

Ordinal *predictors* are typically viewed as being less problematic than ordered outcomes. Usually, ordinal predictors are inserted into a regression using a dummy-specification, or then simply treated as a quantitative variable using the researcher’s prescribed numbering. Nevertheless, both strategies would generally lead to biased estimates. This can be troubling if the latent variable is judged to be an important control. Example: We want to estimate the effect of online lecturing on exam scores, controlling for the exam score from a previous course. Furthermore, assume we only observe if you passed or failed this course. In section 3, I present an alternative estimator that allows you to estimate the treatment effect *as if* the prior exam scores were indeed observable. Again, this requires an assumption regarding the distribution of the latent variable.

¹³ Riedl & Geishecker (2014) compares POLS to a range of alternative estimators for ordered response models in a panel data setting. They note that POLS performs well for estimating ratios between parameters (but is inconsistent for estimating single parameter values).

¹⁴ If, however, the goal is to estimate probabilities while fixing the value of the panel-specific intercept, then adjusted-POLS is also unreliable in this setting.

2.2 ADJUSTED-POLS

Let β_1 denote the parameter of interest in the latent variable model:

$$L = \beta_0 + \beta_1 x + z^T \beta_z + \varepsilon_L$$

where the latent variable (L) is measured on a standardized scale (mean 0, standard deviation 1); x is the independent variable of interest; z^T is a vector of controls and ε_L is the random error term. The estimator for β_1 is given by:

$$\widehat{\beta}_1 = \widehat{\alpha}_1 / \widehat{Var}(\mu)$$

where α_1 is the corresponding parameter in the ‘trimmed’ model:

$$\mu = \alpha_0 + \alpha_1 x + z^T \alpha_z + \varepsilon_\mu$$

Here, μ denotes your expected value on the latent variable (L) given your ordinal score. For a particular ordinal score (j) this expectancy is given by:

$$\mu_j = E[L | \text{ordinal score} = j] = \int_{\min_j}^{\max_j} \underline{L} f(\underline{L}) d\underline{L} / p_j$$

where \min_j and \max_j represent the smallest and largest possible values on the latent variable given this ordinal score¹⁵; p_j is the probability of having this score. \underline{L} denotes the values on the random variable (L) and $f(\underline{L})$ is its probability distribution. In practice, \min_j , \max_j and p_j are estimated from data, using some additional assumption regarding the distribution of L . In many applications, it would be natural to think of L as being normally distributed. In the remainder of this chapter, this is also my assumption.

Example: We measure ‘work satisfaction’ on a 5-point scale, and 4 percent of individuals are very dissatisfied with their jobs (work satisfaction = 1). Now, assuming that L is normally distributed, then the estimate for the expected standardized score for this group is -2.15:

$$\begin{aligned} \hat{\mu}_1 &= E[L | \text{ordinal score} = 1] = E[L | -\infty \leq \underline{L} \leq F^{-1}(0.04)] \\ &\approx E[L | -\infty \leq \underline{L} \leq -1.75] \approx -2.15 \end{aligned}$$

Appendix A.1 presents a simple way of computing this integral when L is normally distributed. Appendix A.2 presents the conditions for consistency.

It is worth noting the close relationship between adjusted-POLS and POLS, the only difference being that adjusted-POLS adds an ‘inflation factor’ – $1/\widehat{Var}(\mu)$ – offsetting the ‘attrition bias’ due to discretization. Hence, if the

¹⁵ In other words, $\max_j = F^{-1}(\pi_j)$ where π_j is the probability of having this ordinal score (j) or a smaller one. Similarly, $\min_j = F^{-1}(\pi_{j-1})$.

interest lies in estimating a ratio between parameters, then adjusted-POLS is identical to POLS. If, on the other hand, the interest lies in estimating single parameters, adjusted-POLS produces larger estimates in absolute terms than POLS, and sometimes distinctly so. For adjusted-POLS, estimates are independent of the degree of discretization: if, for example, one researcher uses a 3-point scale and another one a 10-point scale, estimates are still directly comparable.

2.3 A LATENT INDEPENDENT VARIABLE

In this model, our latent variable is on the right-hand side:

$$Y = \beta_0 + \beta_1 L + \beta_2 x + z^T \beta_z + \varepsilon$$

where β_1 and β_2 are the parameters of interest; z is a vector of controls and ε is the random error term.

The estimators for β_1 and β_2 are given by:

$$\hat{\beta}_1 = \frac{\hat{\beta}_{Y,L|z} - \hat{\beta}_{Y,x|z} \hat{\beta}_{x,L|z}}{1 - \hat{\beta}_{x,L|z} \hat{\beta}_{L,x|z}}$$

$$\hat{\beta}_2 = \frac{\hat{\beta}_{Y,x|z} - \hat{\beta}_{Y,L|z} \hat{\beta}_{L,x|z}}{1 - \hat{\beta}_{x,L|z} \hat{\beta}_{L,x|z}}$$

where $\beta_{Y,L|z}$, $\beta_{Y,x|z}$, $\beta_{x,L|z}$ and $\beta_{L,x|z}$ are parameters as defined by the following models:

$$Y = a + \beta_{Y,L|z} L + z^T \beta_{Y,z|L} + \epsilon$$

$$Y = b + \beta_{Y,x|z} x + z^T \beta_{Y,z|x} + u$$

$$x = c + \beta_{x,L|z} L + z^T \beta_{x,z|L} + v$$

$$L = d + \beta_{L,x|z} x + z^T \beta_{L,z|x} + w$$

These parameters can be estimated recursively. Starting with a model that includes no other independent variables than L , then β_1 is consistently estimated by replacing L with μ :

$$Y = \beta_0 + \beta_1 L + \varepsilon \leftrightarrow Y = \beta_0 + \beta_1 \mu + \varepsilon'$$

With two independent variables, the model is given by:

$$Y = \beta_0 + \beta_1 L + \beta_2 x + \varepsilon \quad (1)$$

and the parameters to be estimated are given by:

$$Y = a + \beta_{Y|L} L + \varepsilon \leftrightarrow Y = a + \beta_{Y|L} \mu + \varepsilon' \quad (1.1)$$

$$Y = b + \beta_{Y,x} x + u \quad (1.2)$$

$$x = c + \beta_{x,L} L + v \leftrightarrow x = c + \beta_{x,L} \mu + v' \quad (1.3)$$

$$L = d + \beta_{L,x} x + w \leftrightarrow \mu / \text{Var}(\mu) = d + \beta_{L,x} x + w' \quad (1.4)$$

Hence, the equations where L is on the right-hand side, (1.1) and (1.3), are estimated by replacing L with μ ; the equation where L is on the left-hand side

(1.4) is estimated by replacing L with μ divided by its variance (this is the adjusted-POLS estimator). In practical applications, the equations (1.1), (1.2) and (1.4) cannot be estimated directly, as μ is not observed. If, however, the distributional assumption regarding L is correct, then μ can be estimated consistently (as described in section 2).

With three independent variables, the model is given by:

$$Y = \beta_0 + \beta_1 L + \beta_2 x + \beta_3 z + \varepsilon \quad (2)$$

and the parameters to be estimated are given by:

$$Y = a + \beta_{Y,L|Z} L + \beta_{Y,Z|L} Z + \epsilon \quad (2.1)$$

$$Y = b + \beta_{Y,x|Z} x + \beta_{Y,z|x} z + u \quad (2.2)$$

$$x = c + \beta_{x,L|Z} L + \beta_{x,Z|L} Z + v \quad (2.3)$$

$$L = d + \beta_{L,x|Z} x + \beta_{L,z|x} z + w \quad (2.4)$$

Note that (2.1) and (2.3) can be estimated in the same way as equation (1); (2.4) is estimated using adjusted-POLS. In a similar fashion, we can estimate a model with any number of independent variables. Note, however, that the number of equations increases fast with the number of independent variables: With three independent variables, we estimate eight equations in order to gain a consistent estimate of β_2 ; with ten independent variables, the corresponding number is 64 equations.¹⁶

In the Appendix (A.3: Latent independent variable – extensions) I show that this estimator is consistent. I also present a stepwise procedure that can aid in programming, and show a practical example using real data.

¹⁶ With k independent variables, we estimate $k^2 - \sum_{i=1}^{k-2} i$ equations.

2.4 SIMULATIONS: ADJUSTED-POLS vs ORDERED PROBIT

In this section, I simulate the sampling distribution of $\hat{\beta}$ under different assumptions regarding the distribution of the observed ordinal variable, and using different sample sizes and models. I find that adjusted-POLS performs on par with ordered probit in a cross-sectional setting, and can be particularly useful in a panel data setting.

2.4.1 The cross-sectional case

The benchmark latent variable model used here is:

$$L = 0.2x + z^T \delta + \varepsilon$$

where L follows a standard normal distribution; x is the independent variable of interest and z^T is a vector consisting of nine control variables. All independent variables are normally distributed with means at 0, standard deviations at 1 and pairwise covariances at 0.2. The vector δ is given by [0.2, 0.2, 0.2, 0.2, -0.2, -0.2, -0.2, -0.2, -0.2]. The independent variables and the error term (ε) are independently distributed, which implies that the error term is also normally distributed with mean 0 and standard deviation $\sqrt{0.68}$. In a second step, I experiment with a smaller error standard deviation at $\sqrt{0.318}$ which is the result of multiplying δ with 1.5. I denote these models (1) and (2), respectively.

Now, ordered probit uses a different scale than the one presented above. While both estimators measure the coefficients in units of standard deviations, adjusted-POLS uses the standard deviation of the latent variable while ordered probit uses that of the error term. In order to compare estimators, I convert the adjusted-POLS coefficients to the ordered probit scale and vice versa. For model (1), β_{prob} is equal to ~ 0.243 ; for model (2), β_{prob} is equal to ~ 0.355 .¹⁷ This conversion requires an estimate for the error variance (in the adjusted-POLS model) and an estimate for the latent variable

¹⁷ The adjusted-POLS estimates are converted to the ordered probit scale by dividing the coefficients with the error standard deviation:

$$\beta_{prob} = \beta / \sqrt{\text{Var}(\varepsilon)}$$

For model (1), $\beta_{prob} = 0.2 / \sqrt{0.68} \approx 0.243$ and for model (2), $\beta_{prob} = 0.2 / \sqrt{0.318} \approx 0.355$. In a similar fashion, the ordered probit coefficients are converted to the adjusted-POLS scale by dividing the coefficients with the standard deviation for the latent variable in the ordered probit model:

$$\beta = \beta_{prob} / \sqrt{\text{Var}(L_{prob})}$$

variance (in the corresponding ordered probit model). I describe the estimators for these variances in the Appendix, section A.4.

Table 1 below describes the characteristics of the simulated sampling distributions. Here I experiment with different distributions for the observed ordinal-scale variable and different sample sizes. As for the ordinal variable, I use the following five variants: (i) ‘Discrete normal’ distribution: the latent variable is rounded off to its closest integer and truncated as to fit a 7-point scale; (ii) Uniform (5 pts); (iii) Binomial(4, 0.75); (iv) U-shaped (5 pts): $f(1) = f(5) = 0.4$; $f(2) = f(4) = 0.08$ and $f(3) = 0.04$, and (v) Bernoulli ($p = 0.2$). With each ordinal variable distribution, the variance for μ decreases, starting at ~ 0.923 (for the ‘discrete normal’ distribution) and ending at ~ 0.490 (the Bernoulli distribution).¹⁸ In general, I set the sample size to 1000 but I also experiment with a small sample size at 100.

Adjusted-POLS performs well in all cases, producing no noteworthy biases. This is true independently of scale, i.e. whether the latent variable variance or error variance is set to one. Furthermore, there is little difference in performance between the estimators: In general, it comes down to the flip of a coin which estimator hits closer to target in a particular sample. You can find cases that slightly favors one estimator over the other, but any differences in performance are likely to be small in practice.

Now, the performance of adjusted-POLS hinges upon the assumption that the latent variable distribution is correctly modeled. If not, then adjusted-POLS can perform poorly, depending on the severity of the misspecification. Section A.5 in the Appendix provides some examples.

¹⁸ The variance for μ is equal to ~ 0.923 (for the ‘discrete normal’ ordinal variable distribution); ~ 0.897 (the Uniform distribution); ~ 0.852 (the skewed Binomial distribution); ~ 0.750 (the U-shaped distribution) and ~ 0.490 (the Bernoulli distribution).

Table 1. Characteristics of the simulated sampling distributions (10,000 repetitions). Multiple regression model, $n = 1000$ unless otherwise stated

	ADJUSTED-POLS						ORDERED PROBIT						
	Mean	Standard-deviation	Skewness	Kurtosis	$\widehat{Var}(\varepsilon)$ mean	MSD ($\times 100$)	Mean	Standard-deviation	Skewness	Kurtosis	$\widehat{Var}(L)$ mean	MSD ($\times 100$)	MSD _{pols} / MSD _{prob}
<i>Ordinal</i>	<i>Model (1): A-POLS scale: $\beta = 0.2$, $Var(\varepsilon) = 0.68$</i>						<i>Model (1): A-POLS scale: $\beta = 0.2$; $Var(L) = 1$</i>						
(i)	0.200	0.029	-0.02	3.00	0.680	0.843	0.200	0.029	-0.02	3.00	1	0.845	0.997
(ii)	0.200	0.030	-0.01	2.97	0.681	0.909	0.200	0.030	-0.00	2.98	1	0.903	1.007
(iii)	0.200	0.031	-0.05	3.08	0.681	0.939	0.200	0.031	-0.05	3.09	1	0.932	1.008
(iv)	0.200	0.034	-0.01	2.99	0.681	1.150	0.200	0.034	-0.01	2.97	1	1.129	1.019
(v)	0.201	0.044	0.00	2.93	0.681	1.972	0.201	0.044	-0.02	2.94	1	1.894	1.041
	<i>n = 100</i>						<i>n = 100</i>						
(iii)	0.201	0.104	-0.07	3.03	0.686	10.865	0.202	0.105	-0.06	3.03	1	11.070	0.981
	<i>Model (1): Probit scale: $\beta_{prob} \approx 0.243$; $Var(\varepsilon_{prob}) = 1$</i>						<i>Model (1): Probit scale: $\beta_{prob} \approx 0.243$; $Var(L_{prob}) \approx 1.47$</i>						
(i)	0.243	0.037	0.02	2.99	1	1.334	0.245	0.037	0.02	2.99	1.49	1.360	0.981
(ii)	0.242	0.038	0.04	2.99	1	1.441	0.244	0.038	0.05	3.00	1.49	1.451	0.993
(iii)	0.243	0.039	-0.00	3.09	1	1.489	0.245	0.039	0.00	3.10	1.49	1.502	0.991
(iv)	0.243	0.043	0.05	3.01	1	1.832	0.245	0.043	0.05	2.99	1.49	1.831	1.001
(v)	0.244	0.056	0.09	2.96	1	3.179	0.248	0.056	0.07	2.97	1.51	3.147	1.010
	<i>n = 100</i>						<i>n = 100</i>						
(iii)	0.246	0.132	0.10	3.16	1	17.563	0.267	0.144	0.12	3.18	1.73	21.454	0.819
	<i>Model (2): A-POLS scale: $\beta = 0.2$, $Var(\varepsilon) = 0.318$</i>						<i>Model (2): A-POLS scale, $\beta = 0.2$; $Var(L) = 1$</i>						
(iii)	0.203	0.024	-0.02	2.95	0.330	0.586	0.204	0.024	-0.01	2.95	1	0.568	1.032
	<i>Model (2): Probit scale: $\beta_{prob} \approx 0.355$, $Var(\varepsilon) = 1$</i>						<i>Model (2): Probit scale, $\beta_{prob} \approx 0.355$, $Var(L_{prob}) \approx 3.14$</i>						
(iii)	0.355	0.044	0.06	2.99	1	1.956	0.358	0.044	0.05	2.98	3.10	1.913	1.022

Notes: (i) The latent variable is rounded off to its closest integer and truncated as to fit a 7-point scale; the variance for μ is ~ 0.923 . (ii) The ordinal variable is uniformly distributed (5 pts); the variance for μ is ~ 0.897 . (iii) The ordinal variable is binomially distributed ($n = 4$; $p = 0.75$); the variance for μ is ~ 0.852 . (iv) The ordinal variable is u-shaped: $f(1) = f(5) = 0.4$; $f(2) = f(4) = 0.08$; $f(3) = 0.04$; the variance for μ is ~ 0.750 . (v) The ordinal variable is Bernoulli distributed ($p = 0.2$); the variance for μ is ~ 0.490 . MSD is the mean square deviation: $\text{avg}(\beta - \hat{\beta})^2$.

2.4.2 Adjusted-POLS in a panel data setting

In this section, I compare adjusted-POLS to ordered probit in a panel data setting:

$$L_{it} = 0.2x_{it} + \alpha_i + \varepsilon_{it}$$

$$x_{it} = \mu_{xi} + u_{it} \quad u_{it} = 0.5u_{i(t-1)} + v_{it}$$

I sample 250 panels ($i = 1, 2, \dots, 250$) and make four measurements for each ($t = 1, 2, 3$ and 4). The panel-specific intercept (α_i) and x_{it} are both normally distributed with means at 0 and a correlation at 0.5. x_{it} is modeled as an autoregressive process with values fluctuating around a panel-specific mean, μ_{xi} . The standard deviation for this mean is $\sqrt{0.5}$ and the standard deviation for u_{it} is also set to $\sqrt{0.5}$. Hence, x_{it} has a standard deviation at 1. As for α_i , I experiment with two different standard deviations; firstly $\sqrt{0.5}$ and secondly $\sqrt{0.7}$. I label these models (3) and (4), respectively. The error term, ε_{it} , is independently and normally distributed, with a mean at 0 and a standard deviation at $\sim\sqrt{0.319}$ for model (3) and $\sim\sqrt{0.093}$ for model (4).¹⁹ For both estimators, the panel-specific intercepts are captured using a set of dummies.

The result is presented in table 2, separately for three different ordinal variable distributions: (i) ‘discrete normal’ (7 points), (iii) skewed binomial ($n = 4$; $p = 0.75$) and (v) Bernoulli ($p = 0.2$). In order to compare the estimators, I convert the adjusted-POLS coefficients to the ordered probit scale and vice versa. For model (3), β_{prob} is equal to ~ 0.354 ; for model (4), β_{prob} is equal to ~ 0.657 .²⁰ This conversion requires an estimate for the error variance (in the adjusted-POLS model) and an estimate for the latent variable variance (in the corresponding ordered probit model). For details, see section A.4 in the Appendix.

Adjusted-POLS performs well in all cases when I estimate β on its original scale ($\beta = 0.2$). The estimate for the error variance is shaky, however. This is especially so when this error variance is low as in model (4) or when $\text{Var}(\mu)$ is low as for the Bernoulli-distributed ordinal variable. Hence, the

¹⁹ The error variance is given by $1 - (0.5 + 0.2^2 + 0.02^{0.5}) \approx 0.319$ for model (3) and by $1 - (0.7 + 0.2^2 + 0.028^{0.5}) \approx 0.093$ for model (4). More generally:

$$\text{Var}(\varepsilon_{it}) = 1 - [\text{Var}(\alpha_i) + \text{Var}(0.2x_{it}) + 2\text{Cov}(\alpha_i, 0.2x_{it})]$$

²⁰ The adjusted-POLS estimates are converted to the ordered probit scale by dividing the coefficients with the error standard deviation. For model (3),

$$\beta_{prob} = \beta / \sqrt{\text{Var}(\varepsilon_{it})} = 0.2 / \sqrt{0.3185 \dots} \approx 0.354;$$

for model (4)

$$\beta_{prob} = \beta / \sqrt{\text{Var}(\varepsilon_{it})} = 0.2 / \sqrt{0.0926 \dots} \approx 0.657.$$

conversion to the ordered probit scale can produce large biases indeed. Ordered probit, on the other hand, is generally unreliable independently of scale.

To further test the performance of adjusted-POLS, I experiment with a small sample consisting of 200 observations ($i = 1, 2, \dots, 100$ and $t = 1, 2$); an autocorrelated error term and lastly, several covariates combined with distinct time-trends. Adjusted-POLS – as measured on its original scale – still performs well in these cases (not included in table 2).²¹

The lack of a valid error variance estimator prohibits us from estimating individual-specific probabilities using adjusted-POLS. However, adjusted-POLS can still be used to estimate average treatment effects, i.e. the average change in the probability of a specific outcome as x increases from one fixed number to another. In the Appendix (see section A.6) I show how these probabilities are estimated and I also provide simulations that compare the theoretical probabilities to those produced by adjusted-POLS, ordered probit and – in the case of a binary response – OLS. These simulations suggest that adjusted-POLS can serve as a useful alternative for estimating such probabilities in a panel data setting: In general, the adjusted-POLS probabilities – and the corresponding average treatment effects – match well with the theoretical predictions.

²¹ In the first two cases, I estimate model (4) using the skewed binomially distributed ordinal variable and I make 10,000 repetitions for each sampling distribution. For a sample size at 200, the estimates average at 0.203 (SD 0.098). In the second case, I slightly modify model (4) by letting the error term follow: $\varepsilon_{it} = 0.5\varepsilon_{it-1} + w_{it}$ (the sample size is now set to 1000). Here, the estimates average at 0.201 (SD 0.031). In the third case, I estimate a model with four covariates as well as distinct time trends, combined with an IID error. The time-specific intercepts are uncorrelated with the other covariates and I control for them in the regressions. The model:

$$\begin{aligned} L_{it} &= 0.2x_{1it} + 0.2x_{2it} - 0.2x_{3it} - 0.2x_{4it} + \alpha_i + \tau_t + \varepsilon_{it} \\ \alpha_i &= 0.3\mu_{1it} + 0.3\mu_{2it} - 0.3\mu_{3it} - 0.3\mu_{4it} + e_i \quad \text{Var}(e_i) = 0.462 \\ \tau_t &= -0.4, 0, 0.1, 0.3 \end{aligned}$$

where μ_j is the expected value for the j :th covariate; the covariates are normally distributed and autocorrelated over time, with variances at unity and pairwise covariances at 0.2, but no correlation within panels. The error variance is 0.093. Here, the estimates for the effect of x_1 average at 0.200 (SD 0.033).

Table 2. Characteristics of the simulated sampling distributions (10,000 repetitions) in a panel data setting: $n = 1000$ ($i = 1, 2, \dots, 250$; $t = 1, 2, 3, 4$)

	ADJUSTED-POLS						ORDERED PROBIT						
	Mean	Standard-deviation	Skew-ness	Kurtosis	$\widehat{Var}(\varepsilon)$ mean	MSD ($\times 100$)	Mean	Standard-deviation	Skew-ness	Kurtosis	$\widehat{Var}(L)$ Mean	MSD ($\times 100$)	MSD _{pols} / MSD _{probit}
<i>Ordinal</i>	<i>Model (3): $\beta = 0.2$, $Var(\varepsilon) \propto 0.319$</i>						<i>Model (3): $\beta = 0.2$; $Var(L) = 1$</i>						
(i)	0.201	0.041	0.03	2.95	0.322	0.166	0.198	0.041	0.03	2.92	1	0.167	0.996
(iii)	0.200	0.044	0.01	3.01	0.304	0.195	0.151	0.034	0.05	3.07	1	0.357	0.545
(v)	0.202	0.071	-0.01	2.91	0.118	0.498	0.139	0.048	0.04	3.02	1	0.609	0.818
	<i>Model (3): $\beta_{prob} \approx 0.354$; $Var(\varepsilon_{prob}) = 1$</i>						<i>Model (3): $\beta_{prob} \approx 0.354$; $Var(L_{prob}) \propto 3.14$</i>						
(i)	0.354	0.073	0.03	2.92	1	0.528	0.431	0.089	0.03	2.92	4.76	1.368	0.386
(iii)	0.364	0.081	0.01	3.01	1	0.672	0.442	0.099	0.06	3.08	8.68	1.747	0.385
(v)	0.697 ^a	0.763	21.0	750.9	1	69.98	0.518	0.187	0.13	3.09	13.86	6.150	11.38
	<i>Model (4): $\beta = 0.2$, $Var(\varepsilon) \propto 0.093$</i>						<i>Model (4): $\beta = 0.2$; $Var(L) = 1$</i>						
(i)	0.202	0.029	0.05	2.99	0.093	0.082	0.149	0.037	0.24	2.57	1	0.400	0.206
(iii)	0.201	0.032	0.06	3.10	0.051	0.105	0.145	0.038	0.26	2.44	1	0.444	0.235
(v)	0.202	0.056	0.10	2.99	-0.399	0.316	0.218	0.059	0.28	3.41	1	0.383	0.824
	<i>Model (4): $\beta_{prob} \approx 0.657$, $Var(\varepsilon_{prob}) = 1$</i>						<i>Model (4): $\beta_{prob} \approx 0.657$, $Var(L_{prob}) \propto 10.79$</i>						
(i)	0.664	0.096	0.11	3.03	1	0.919	0.910	0.139	0.20	3.18	43.42	8.339	0.110
(iii)	0.950 ^b	0.322	5.53	72.99	1	18.99	0.930	0.159	0.26	3.34	47.69	9.969	1.905
(v)	-	-	-	-	-	-	0.985	0.290	0.44	3.73	20.38	19.19	-

Notes: (i) 'Discrete normal' distribution (7 pts); (iii) Binomial(4, 0.75); (v) Bernoulli ($p = 0.2$). MSD is the mean square deviation: $\text{avg}(\beta - \hat{\beta})^2$. ^aSee model (3v): For the 9088 cases with positive error variance estimates, $\hat{\beta}_{prob}$ averaged at 0.697. ^bSee model (4iii): For the 9967 cases with positive error variance estimates, $\hat{\beta}_{prob}$ averaged at 0.950.

2.5 INTERPRETATION

It is well known that the coefficients from ordered probit models are difficult to interpret. “Indeed, without a fair amount of extra calculation, it is quite unclear how the coefficients in the ordered probit model should be interpreted.” (Greene, 2012, p. 830). The unit of measurement – error standard deviations – does nothing to ease in interpretation. Adjusted-POLS, on the other hand, gives the latent variable a standard deviation of one. Nevertheless, there is still room for saying a couple of words on interpretation. I will do this using an example.

Example: We want to measure how body mass index (*BMI*) varies with *age* and *gender*:

$$\ln(BMI) = \beta_0 + \beta_1 age + \beta_2 female + \varepsilon$$

where $\ln(BMI)$ is the natural logarithm of BMI. Here, $\ln(BMI)$ is also our latent variable, and we observe it measured in five categories (BMI<18.5: Underweight; 18.5-25: Normal weight; 25-30: Overweight; 30-35: Moderately obese; >35: Severely obese). The distribution for this variable is illustrated in figure 1 below. Here, I exploit ‘nhanes2’ which is part of the Stata data library.²²

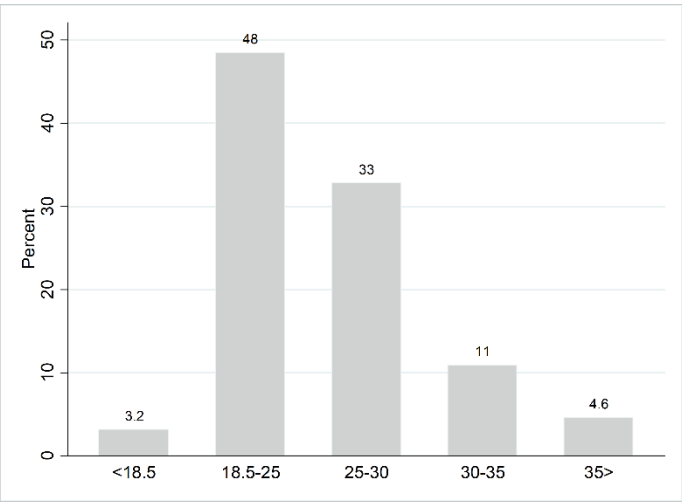


Figure 1. Body mass index

Table 3 below presents the estimates when using (1) POLS, (2) Adjusted-POLS, (3) OLS on BMI (standardized) and (4) OLS on BMI measured on a logarithmic scale (and standardized). Hence, we would only be able to

²² For the sake of simplicity, I ignore the sampling weights.

observe the POLS- and adjusted-POLS-coefficients. The estimates from adjusted-POLS correspond more closely to (4) than (3) which can be explained by the fact that BMI is approximately log normally distributed. Now, without much background knowledge, we would not generally know this (although in this case, the distribution for the ordinal variable would give us a hint). Hence, we would conclude that aging by a decade is associated with an increase in (normalized) BMI by 0.105 standard deviations, controlling for gender.

Table 3. The effect of aging on BMI (POLS, Adjusted POLS & OLS)

	(1) POLS	(2) Adj.-POLS	(3) OLS, BMI	(4) OLS, ln(BMI)
Age (years)	0.00886*** (0.00052)	0.0105*** (0.00061)	0.00994*** (0.00056)	0.0107*** (0.00056)
Female	-0.0275 (0.018)	-0.0325 (0.021)	0.00774 (0.019)	-0.0451** (0.019)
Observations	10,351	10,351	10,351	10,351

Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Is 0.105 standard deviations a small or large movement up the BMI-distribution? One way to assess this is to convert the estimate into an ‘average percentile rank’ effect as follows:

$$APE = E \left[\frac{\partial F(\underline{L})}{\partial x} \right] = \frac{\beta}{2\sqrt{\pi}}$$

where $F(\underline{L})$ is the standard normal distribution function and $\frac{\partial F(\underline{L})}{\partial x}$ is the marginal effect of x on $F(\underline{L})$; β is the marginal effect of x on \underline{L} . See the Appendix (A.7) for the intermediate steps.

Example continued: On average, and conditional on gender, aging by a decade is predicted to move you three percentile ranks up the BMI-distribution:

$$\frac{10\hat{\beta}}{2\sqrt{\pi}} = \frac{0.105}{2\sqrt{\pi}} \approx 0.030$$

2.6 CONCLUSIONS

Adjusted-POLS is a simple and consistent estimator for estimating ordinal response models. The estimator generally performs on par with ordered probit when the distributional assumptions of both estimators are. Most importantly, however, adjusted-POLS is easily combined with other linear estimation techniques, such as fixed effects or instrumental variables estimation. On the downside, adjusted-POLS relies on a rather strong assumption: If the structural part of the model and the error term are independently distributed, then adjusted-POLS requires that both of these are also normally distributed; ordered probit only requires normality of the error term.

In addition to adjusted-POLS, this chapter also presents a consistent estimator for a marginal effect, when controlling for a latent *independent* variable. This is useful when this variable is an important control, i.e. when the bias induced by controlling for the observed ordinal-scale variable is judged unacceptable.

References

- Brockett, P.L. (1981). A note on the numerical assignment of scores to ranked categorical data. *Journal of Mathematical Sociology*, 8, 91-101.
- Bross, I. D. (1958). How to use riddit analysis. *Biometrics*, 18-38.
- Fielding, A. (1997). On scoring ordered classifications. *British Journal of Mathematical and Statistical Psychology*, 50(2), 285-307.
- Greene, W. H. (2012). *Econometric analysis* (International edition).
- Maddala, G. S. (1983). *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge University Press.
- Neyman, J., & Scott, E. (1948). Consistent Estimates Based on Partially Consistent Observations. *Econometrica*, 16(1), 1-32.
- Riedl, M., & Geishecker, I. (2014). Keep it simple: estimation strategies for ordered response models with fixed effects. *Journal of Applied Statistics*, 41(11), 2358-2374.
- Terza, J. V. (1987). Estimating linear models with ordinal qualitative regressors. *Journal of Econometrics*, 34(3), 275-291.
- Van Praag, B., & Ferrer-i-Carbonell, A. (2006). *An almost integration-free approach to ordered response models* (No. 06-047/3). Tinbergen Institute discussion paper.

Appendix

A.1 Calculating the μ :s

Here, the latent variable, L , is taken to follow a standard normal distribution. For a particular ordinal score (j) the expected value of L can be calculated as:

$$\mu_j = \frac{f(\min_j) - f(\max_j)}{p_j}$$

where \min_j and \max_j are the smallest and largest possible values on L for individuals belonging to this category (j); $f(\cdot)$ is the standard normal density function and p_j is the probability of belonging to this category (see, for example, Maddala, 1983, p. 366).

A.2 Consistency of adjusted-POLS

First note that the latent variable (L) is the sum of your expected score given your ordinal value (μ) and an error (u):

$$L = \mu + u$$

Now, using the simple linear regression model as a starting point:

$$L = \beta_0 + \beta_1 x + \varepsilon$$

we write the *inverse* latent variable model as:

$$x = \gamma_0 + \gamma_1 L + \epsilon$$

$$x = \gamma_0 + \gamma_1(\mu + u) + \epsilon$$

$$x = \gamma_0 + \gamma_1\mu + v, \text{ where } v = \epsilon + \gamma_1 u$$

Now, if $Cov(\mu, v) = 0$, then:

$$\gamma_1 = \frac{Cov(L, x)}{Var(L)} = \frac{Cov(\mu, x)}{Var(\mu)}$$

where $Var(L) = 1$. Hence,

$$\beta_1 = \frac{Cov(L, x)}{Var(x)} = \frac{Cov(\mu, x)}{Var(x)Var(\mu)}$$

$$\beta_1 = \frac{\alpha_1}{Var(\mu)}$$

where α_1 is the corresponding parameter in the trimmed model:

$$\mu = \alpha_0 + \alpha_1 x + \varepsilon_\mu$$

Hence, consistent estimation of β_1 can be obtained if $Cov(\mu, v) = 0$ where $Cov(\mu, v) = Cov(\mu, \varepsilon) + \gamma_1 Cov(\mu, u)$. When is this assumption satisfied? Firstly, $Cov(\mu, u) = 0$ if the distributional assumption regarding L is correct. Secondly, $Cov(\mu, \varepsilon) = 0$ if x is unable to predict μ conditional on L .²³ This assumption is satisfied if the inverse conditional expectations function – $E(x|L)$ – is also linear. For a normally distributed latent variable where x and ε are independently distributed, the inverse conditional expectations function is linear and adjusted-POLS is consistent.²⁴

The proof presented above is easily extended to the multiple regression model by replacing x with \tilde{x} , where \tilde{x} is the ‘population’ residual in a model with x as outcome and the other independent variables as predictors.

A.3 A latent independent variable – extensions

Consistency

Let β_1 and β_2 be the parameters of interest as defined by the model:

$$Y = \beta_0 + \beta_1 L + \beta_2 x + z^T \beta_z + \varepsilon$$

where L is the latent variable.

Now, let $\beta_{Y,L|z}$ denote the conditional effect of L on Y in a model that excludes x . Similarly, let $\beta_{x,L|z}$ and $\beta_{L,x|z}$ denote conditional effects in models that exclude Y . Then we have that:

$$\beta_{Y,L|z} = \beta_1 + \beta_2 \beta_{x,L|z}$$

$$\beta_{Y,x|z} = \beta_2 + \beta_1 \beta_{L,x|z}$$

These are your standard textbook formulas that relate parameters of one model, to the corresponding parameters in a model with one additional control (x and L , respectively). Solving for β_1 and β_2 gives us:

²³ In other words, $\delta_2 = 0$ in the model: $\mu = \delta_0 + \delta_1 L + \delta_2 x + e$.

²⁴ Assume that L is normally distributed. If x and ε are independently distributed, then it follows that they are also normally distributed (at least if $\beta_1 \neq 0$). This is a result of Cramér’s decomposition theorem. Hence, L and x are jointly normal and their relationship – as described by $E(L|x)$ or $E(x|L)$ – is linear.

$$\beta_1 = \frac{\beta_{Y,L|Z} - \beta_{Y,x|Z}\beta_{x,L|Z}}{1 - \beta_{x,L|Z}\beta_{L,x|Z}}$$

$$\beta_2 = \frac{\beta_{Y,x|Z} - \beta_{Y,L|Z}\beta_{L,x|Z}}{1 - \beta_{x,L|Z}\beta_{L,x|Z}}$$

Hence, β_1 and β_2 can be estimated consistently if all of its components also can be estimated consistently. We have seen that $\beta_{L,x|Z}$ can be estimated consistently using adjusted-POLS. Now, $\beta_{.,L|Z(n)}$ will be a function of $\beta_{.,L|Z(n-1)}$ which in turn is a function of $\beta_{.,L|Z(n-2)}$ and so on.²⁵ Ultimately, $\beta_{.,L|Z(n)}$ is a function of $\beta_{.,L}$ which, in turn, can be estimated consistently by replacing L with μ . We know this since $L = \mu + u$ where μ and u are uncorrelated if the distributional assumption regarding L is correct.

Programming

It is rather tedious to write a code for this estimator. Below is a stepwise procedure that can aid in programming. In this example, β_2 is the parameter of interest.

Step 1: Estimate $\beta_{.,L}$

Example: With $k = 5$ independent variables (L, x, z_1, z_2 and z_3) we estimate the following five effects: $\beta_{Y,L}, \beta_{x,L}, \beta_{z_1,L}, \beta_{z_2,L}$ and $\beta_{z_3,L}$, i.e. the unconditional effects of L on each remaining variable.

Step 2: Estimate $\beta_{.,L|z_1}$

Example continued: Here we estimate the following four effects: $\beta_{Y,L|z_1}, \beta_{x,L|z_1}, \beta_{z_2,L|z_1}$ and $\beta_{z_3,L|z_1}$, i.e. the conditional effects of L on each remaining variable. These effects, in turn, are estimated using a set of five unconditional effects, i.e. the effects of z_1 on each remaining variable (Y, L, x, z_2 and z_3). We can now construct the estimates of interest, for example:

$$\hat{\beta}_{Y,L|z_1} = \frac{\hat{\beta}_{Y,L} - \hat{\beta}_{Y,z_1}\hat{\beta}_{z_1,L}}{1 - \hat{\beta}_{z_1,L}\hat{\beta}_{L,z_1}}$$

Step 3: Estimate $\beta_{.,L|z_1,z_2}$

²⁵ Notation: $z_{(n)}$ is a label for the vector z consisting of n components.

Example continued: Here we estimate the following three effects: $\beta_{Y,L|z_1,z_2}$, $\beta_{x,L|z_1,z_2}$ and $\beta_{z_1,L|z_1,z_2}$, i.e. the conditional effects of L on each remaining variable. These, in turn, are estimated using a set of four conditional effects, i.e. the effects of z_2 on each remaining variable (Y, L, x and z_3) conditional on z_1 . We construct the estimates of interest, for example:

$$\hat{\beta}_{Y,L|z_1,z_2} = \frac{\hat{\beta}_{Y,L|z_1} - \hat{\beta}_{Y,z_2|z_1}\hat{\beta}_{z_2,L|z_1}}{1 - \hat{\beta}_{z_2,L|z_1}\hat{\beta}_{L,z_2|z_1}}$$

Step 4: Estimate $\beta_{.,L|z_1,z_2,z_3}$

Example continued: Here we estimate the following two effects: $\beta_{Y,L|z_1,z_2,z_3}$ and $\beta_{x,L|z_1,z_2,z_3}$. These, in turn, are estimated using a set of three conditional effects, i.e. the effects of z_3 on each remaining variable (Y, L and x) conditional on z_1 and z_2 . We construct the estimates of interest, for example:

$$\hat{\beta}_{Y,L|z_1,z_2,z_3} = \frac{\hat{\beta}_{Y,L|z_1,z_2} - \hat{\beta}_{Y,z_3|z_1,z_2}\hat{\beta}_{z_3,L|z_1,z_2}}{1 - \hat{\beta}_{z_3,L|z_1,z_2}\hat{\beta}_{L,z_3|z_1,z_2}}$$

Step k-1: Estimate $\beta_{.,L|z_1,z_2,z_3,\dots,z_n}$

Step k: Estimate $\beta_{.,x|z_1,z_2,z_3,\dots,z_n}$

Example continued: Here we estimate the following two effects: $\beta_{Y,x|z_1,z_2,z_3}$ and $\beta_{L,x|z_1,z_2,z_3}$, i.e. the conditional effect of x on Y and L , respectively. We construct the estimate of interest:

$$\beta_2 = \frac{\beta_{Y,x|z_1,z_2,z_3} - \beta_{Y,L|z_1,z_2,z_3}\beta_{L,x|z_1,z_2,z_3}}{1 - \beta_{x,L|z_1,z_2,z_3}\beta_{L,x|z_1,z_2,z_3}}$$

Also, note that any regression where L is an outcome is estimated by replacing L with $\hat{\mu}/\widehat{Var}(\mu)$; any regression where L is a predictor is estimated by replacing L with $\hat{\mu}$.

An example using real data

In this example, we want to measure how hemoglobin (*hgb*) varies with *age* and *BMI* among women:

$$hgb = \beta_0 + \beta_1 age + \beta_2 \ln(BMI) + \varepsilon$$

where $\ln(BMI)$ is the natural logarithm of BMI. This is also our latent variable, and here we assume that we only observe if a person is obese ($BMI > 30$) or not.

Table A1 below presents the results when (1) BMI is observed and measured on a logarithmic scale (and standardized), (2) We control for being obese using a dummy, and (3) we use the estimator of interest. Here, I exploit “nhanes2” which is part of the Stata data library.

Table A1. The effect of aging on hemoglobin, using three alternative ways of controlling for BMI (obesity)

	(1)	(2)	(3)
Age (in decades)	0.0779*** (0.0091)	0.0857*** (0.0089)	0.0800*** (0.0091)
$\ln(BMI)$ (z)	0.0925*** (0.0156)	-	0.118*** (0.0244)
Obese (dummy)	-	0.207*** (0.0396)	-
Observations	5,436	5,436	5,436

(1) BMI is observed, (2) BMI is not observed; instead I use a dummy for being obese, (3) BMI is not observed, but we reconstruct the coefficients using the estimator of interest. Standard errors in parentheses; (3) Bootstrapped standard errors (20,692 reps.) *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table A1 (1) shows that each additional decade increases hemoglobin by roughly 0.08 units, when holding $\ln(BMI)$ constant. When controlling for obesity instead, this effect is slightly overestimated. The estimator of interest (see (3)) adjusts for this bias, and gives us also an estimate for the effect of BMI.

A.4 Estimating the error and latent variable variances

The error variance in the adjusted-POLS model

Consider the model:

$$L = \beta_0 + \sum \beta_k x_k + \varepsilon$$

where k is an index numbering the independent variables ($k = 1, 2, \dots, K$). The estimator for the error variance is given by:

$$\widehat{Var}(\varepsilon) = 1 - \sum_k \sum_m [b_k b_m - Cov(\widehat{b_k}, b_m)] Cov(\widehat{x_k}, x_m)$$

where b_k is the adjusted-POLS estimate for β_k . Below follows the motivation.

The variance for the latent variable is given by:

$$Var(L) = 1 = Var(\beta_0 + \sum \beta_k x_k) + Var(\varepsilon)$$

where

$$Var(\beta_0 + \sum \beta_k x_k) = \sum_k \sum_m \beta_k \beta_m Cov(x_k, x_m)$$

I estimate $\beta_k \beta_m$ with $b_k b_m - Cov(\widehat{b_k}, b_m)$; $Cov(x_k, x_m)$ is estimated by its sample counterpart. This is motivated by the following: Assume that b_k is an unbiased estimator for β_k . Then it follows that $b_k b_m$ is a biased estimator for $\beta_k \beta_m$, overestimating this parameter value by $Cov(b_k, b_m)$:

$$E(b_k b_m) = \beta_k \beta_m + Cov(b_k, b_m)$$

Hence, an unbiased estimator for $\beta_k \beta_m$ is given by $b_k b_m - Cov(b_k, b_m)$.

I use the same error variance estimator in the panel data case, where the panel-specific dummies are treated like other independent variables.

The latent variable variance in the ordered probit model

Here, I follow the convention²⁶ and estimate $Var(L_{probit})$ as:

$$\widehat{Var}(L_{probit}) = \frac{SSM}{n-1} + 1$$

where SSM is the model sum of squares in the estimated ordered probit model.

A.5 A non-normal latent variable, treated as normal

In this section, I sample from a non-normal latent variable distribution that I treat as a normal distribution. I estimate one simple and one multiple regression model. In both cases, I let the random error term be normally distributed. Hence, these models obey the distributional assumption of the

²⁶ This is how the Stata-written `spost`-program calculates the variance.

ordered probit estimator only. For both models, L has mean 0 and standard deviation 1; the first independent variable (x_1) is a dummy that takes on the values 0 and 1 with equal probability; the independent variables and the error term are independently distributed. The models include:

$$L = -0.2 + 0.4x_1 + \varepsilon \quad (\text{N1})$$

$$L = a + 0.4x_1 + 0.4x_2 + 0.4x_3 + 0.4x_4 + 0.4x_5 + \varepsilon \quad (\text{N2})$$

(N1) This implies that the error term has mean 0 and standard deviation $\sqrt{0.96}$. Measured on the probit-scale, the coefficient for x_1 has a value at roughly 0.408 ($=0.4/\sqrt{0.96}$).

(N2) Here, x_2 is a mixture between two normal distributions: $0.9N(0, \sqrt{0.4}) + 0.1N(0, \sqrt{6})$ which gives this variable excess kurtosis (outliers); x_3 follows a Weibull distribution with scale and shape parameters set to 1 and 1.5 (the location parameter is set to 0); x_4 is a dummy taking on the value 1 with 30 percent probability (and 0 otherwise) and x_5 is Poisson distributed ($\lambda = 1$). Some of these variables correlate: the correlation between x_1 and x_2 is ~ 0.29 ; the correlation between x_4 and x_5 is approximately -0.35 .²⁷ a is a constant ($-1.08...$) that gives the error term a mean at 0. This is a model where L has a slight positive skew (~ 0.05) with excess kurtosis (~ 3.30). Roughly 44 percent of the variance in L is explained by the model; the error variance is ~ 0.558 . Measured on the probit-scale, the coefficients have a value at ~ 0.535 ($= 0.4/\sqrt{0.558}...$).

Table A2a presents the characteristics of the simulated sampling distributions. In all simulations, I sample 1000 observations where the ordinal variable follows a skewed binomial distribution ($n = 4$; $p = 0.75$).²⁸ For model (N1), there is little difference in performance between adjusted-POLS and ordered-probit; for model (N2) the difference is quite clear. Here, adjusted-POLS produces biases at -0.03 to 0.02 standard deviations depending on parameter; as expected, ordered probit produces no biases worth mentioning.

Now, the performance of adjusted-POLS also depends on the ordinal variable distribution. This is exemplified in table A2b. Here, the ordinal variable is a discretized version of the latent variable: the latent variable is

²⁷ x_1 takes on the value 1 if $0.5x_2 + n > 0$, where n is a standard normal variable; x_4 takes on the value 1 if $0.641x_5 + n < 0$.

²⁸ I determine the 'theoretical cutpoints' through a simulation of its own: I sample 20 million observations on L and treat this as the theoretical distribution, i.e. I use it to determine the 'theoretical cutpoints' that are used in all of the following simulations. (When applying adjusted-POLS to the data, I estimate the cutpoints separately for each sample using sample proportions.)

Table A2a. Characteristics of the simulated sampling distributions (10,000 repetitions) with a non-normal latent variable and normal error term; $n = 1000$. The ordinal variable is binomially distributed ($n = 4$, $p = 0.75$)

	ADJUSTED-POLS						ORDERED PROBIT						
	Mean	Standard-deviation	Skew-ness	Kurtosis	$Var(\varepsilon)$ mean	MSD ($\times 100$)	Mean	Standard-deviation	Skew-ness	Kurtosis	$Var(L)$ Mean	MSD ($\times 100$)	MSD _{pols} / MSD _{probit}
<i>Covariate</i>	<i>Model (N1): A-POLS scale: $\beta = 0.4, Var(\varepsilon) = 0.96$</i>						<i>Model (N1): A-POLS scale: $\beta = 0.4, Var(L) = 1$</i>						
x_1	0.401	0.066	0.03	3.00	0.960	4.315	0.400	0.065	0.02	3.00	1	4.205	1.026
	<i>Model (N1): Probit scale: $\beta_{prob} \approx 0.408, Var(\varepsilon_{prob}) = 1$</i>						<i>Model (N1): Probit scale: $\beta_{prob} \approx 0.408, Var(L_{prob}) \approx 1.04$</i>						
x_1	0.410	0.070	0.10	3.03	1	4.900	0.409	0.069	0.08	3.02	1.04	4.774	1.027
	<i>Model (N2): A-POLS scale: $\beta = 0.4, Var(\varepsilon) \approx 0.558$</i>						<i>Model (N2): A-POLS scale: $\beta = 0.4, Var(L) = 1$</i>						
x_1	0.422	0.057	-0.03	3.00	0.585	0.377	0.399	0.055	-0.04	2.97	1	0.300	1.256
x_2	0.372	0.029	0.07	3.02	0.585	0.163	0.401	0.029	0.04	3.03	1	0.086	1.883
x_3	0.393	0.042	-0.03	2.97	0.585	0.181	0.401	0.043	-0.01	2.96	1	0.189	0.957
x_4	0.406	0.064	-0.01	3.04	0.585	0.414	0.399	0.062	0.00	3.05	1	0.383	1.080
x_5	0.390	0.027	0.02	3.00	0.585	0.080	0.400	0.028	0.03	2.99	1	0.078	1.033
	<i>Model (N2): Probit scale: $\beta_{prob} \approx 0.535, Var(\varepsilon_{prob}) = 1$</i>						<i>Model (N2): Probit scale: $\beta_{prob} \approx 0.535, Var(L_{prob}) \approx 1.79$</i>						
x_1	0.553	0.079	0.03	3.01	1	0.649	0.537	0.076	0.01	2.98	1.81	0.581	1.116
x_2	0.486	0.042	0.13	3.06	1	0.414	0.540	0.046	0.12	3.07	1.81	0.213	1.943
x_3	0.514	0.059	0.03	2.96	1	0.387	0.539	0.062	0.05	2.95	1.81	0.391	0.990
x_4	0.532	0.087	0.04	3.03	1	0.751	0.537	0.086	0.05	3.05	1.81	0.738	1.017
x_5	0.511	0.040	0.11	2.99	1	0.222	0.539	0.043	0.13	3.00	1.81	0.190	1.169

Notes: x_1 is a dummy (50/50); x_2 is a mixture between two normal distributions with excess extreme value probability (outliers); x_3 follows a Weibull distribution ($\lambda = 1$, $k = 1.5$); x_4 is a dummy (30/70) and x_5 is Poisson distributed ($\lambda = 1$). MSD is the mean square deviation: $\text{avg}(\beta - \hat{\beta})^2$.

Table A2b. Characteristics of the simulated sampling distributions (10,000 repetitions) with a non-normal latent variable and normal error term; $n = 1000$. The ordinal variable is a discretized version of the latent variable (7 pts)

	ADJUSTED-POLS						ORDERED PROBIT						
	Mean	Standard-deviation	Skew-ness	Kurtosis	$\widehat{Var}(\varepsilon)$ mean	MSD ($\times 100$)	Mean	Standard-deviation	Skew-ness	Kurtosis	$\widehat{Var}(L)$ Mean	MSD ($\times 100$)	MSD_{pol}/MSD_{prob}
<i>Covariate</i>	<i>Model (N2): A-POLS scale: $\beta = 0.4, Var(\varepsilon) \approx 0.558$</i>						<i>Model (N2): A-POLS scale: $\beta = 0.4, Var(L) = 1$</i>						
x_1	0.405	0.052	-0.02	2.96	0.563	0.274	0.401	0.052	-0.02	2.96	1	0.273	1.005
x_2	0.394	0.025	0.05	2.96	0.563	0.066	0.401	0.026	0.05	2.96	1	0.068	0.980
x_3	0.400	0.040	-0.03	2.95	0.563	0.163	0.400	0.041	-0.03	2.96	1	0.165	0.992
x_4	0.401	0.058	-0.00	3.03	0.563	0.342	0.400	0.058	-0.00	3.03	1	0.341	1.005
x_5	0.400	0.025	-0.00	2.89	0.563	0.064	0.401	0.026	0.01	2.89	1	0.065	0.980
	<i>Model (N2): Probit scale: $\beta_{prob} \approx 0.535, Var(\varepsilon_{prob}) = 1$</i>						<i>Model (N2): Probit scale: $\beta_{prob} \approx 0.535, Var(L_{prob}) \approx 1.79$</i>						
x_1	0.541	0.073	0.04	2.98	1	0.529	0.539	0.073	0.04	2.98	1.81	0.531	0.996
x_2	0.525	0.038	0.10	3.01	1	0.151	0.539	0.040	0.12	3.00	1.81	0.163	0.923
x_3	0.534	0.057	0.02	2.94	1	0.329	0.538	0.058	0.03	2.94	1.81	0.337	0.990
x_4	0.535	0.080	0.04	3.02	1	0.646	0.538	0.081	0.04	3.02	1.81	0.653	0.976
x_5	0.534	0.039	0.07	2.91	1	0.150	0.539	0.039	0.08	2.91	1.81	0.156	0.963

Notes: x_1 is a dummy (50/50); x_2 is a mixture between two normal distributions with excess kurtosis (outliers); x_3 follows a Weibull distribution ($\lambda = 1, k = 1.5$); x_4 is a dummy (30/70) and x_5 is Poisson distributed ($\lambda = 1$). MSD is the mean square deviation: $\text{avg}(\beta - \hat{\beta})^2$.

A.6 Estimating probabilities in a panel data setting using adjusted-POLS

In this section, I show how we can use adjusted-POLS to describe the probability distribution over the ordinal values, and how this distribution changes as x_{it} is manipulated. I also provide simulations that compare adjusted-POLS to ordered probit and – in the case of a binary response – OLS. For all estimators, I capture the panel-specific intercepts using a set of dummies.

The simulations suggest that adjusted-POLS can serve as a useful alternative for estimating conditional probability distributions in a panel data setting. In general, adjusted-POLS performs well in estimating ‘average effects’, ‘average marginal effects’ as well as ‘average treatment effects’ (described in more detail below). For sufficiently small samples, however, the performance of adjusted-POLS does start to suffer.

Average effects and average marginal effects using adjusted-POLS

Here, an *average effect* is the expected change in the probability of a specific outcome as x_{it} increases by one unit.

If x_{it} increases by one unit for each observation, the probability of an ordinal value at j or below becomes:

$$P(\text{ordinal score} \leq j | x_{it} = \underline{x_{it}} + 1) = N(\text{cut}_j; \mu = \beta; \sigma = 1)$$

where x_{it} denotes the random variable and $\underline{x_{it}}$ is the realization; N is the normal distribution function evaluated at the relevant cutpoint (cut_j): this cutpoint is given by $N^{-1}(\pi_j)$ where π_j is the initial probability of having an ordinal value at j or below; β is the adjusted-POLS parameter for x_{it} . By comparing this probability distribution to the original ($x_{it} = \underline{x_{it}}$), we obtain the average effects.

The average marginal effect is given by²⁹:

$$E \left[\frac{\partial P(\text{ordinal score} = j)}{\partial x_{it}} \right] =$$

$$= \frac{\beta}{\sqrt{2\pi}} \left\{ \exp \left[- \left(\frac{cut_{j-1}^2}{2} \right) \right] - \exp \left[- \left(\frac{cut_j^2}{2} \right) \right] \right\}$$

For the first category ($j = 1$) this simplifies to:

$$E \left[\frac{\partial P(\text{ordinal score} = 1)}{\partial x_{it}} \right] = - \frac{\beta}{\sqrt{2\pi}} \exp \left[- \left(\frac{cut_1^2}{2} \right) \right]$$

²⁹ The probability of belonging to category j is given by:

$$P_j = 0.5 \operatorname{erf}(z_j) - 0.5 \operatorname{erf}(z_{j-1})$$

Where $\operatorname{erf}(z)$ is the so called error function, and where z_j is given by:

$$z_j = \frac{cut_j - \hat{L}}{\sigma_\varepsilon \sqrt{2}}$$

\hat{L} is your prediction, $\hat{L} = \beta_1 x_1 + \dots + \beta_K x_K + \alpha$, and σ_ε is the error standard deviation in the corresponding model. Hence,

$$\begin{aligned} \frac{\partial P_j}{\partial x_k} &= \frac{\partial z_j}{\partial x_k} \cdot \frac{\partial 0.5 \operatorname{erf}(z_j)}{\partial z_j} + \frac{\partial z_{j-1}}{\partial x_k} \cdot \frac{\partial -0.5 \operatorname{erf}(z_{j-1})}{\partial z_{j-1}} = \\ &= \frac{\beta_k}{\sigma_\varepsilon \sqrt{2\pi}} [\exp(-z_{j-1}^2) - \exp(-z_j^2)] \\ E \left(\frac{\partial P_j}{\partial x_k} \right) &= E \left\{ \frac{\beta_k}{\sigma_\varepsilon \sqrt{2\pi}} [\exp(-z_{j-1}^2) - \exp(-z_j^2)] \right\} \\ &= \int_{-\infty}^{\infty} f(\hat{L}) \frac{\beta_k}{\sigma_\varepsilon \sqrt{2\pi}} [\exp(-z_{j-1}^2) - \exp(-z_j^2)] d\hat{L} \\ &= \int_{-\infty}^{\infty} \frac{1}{\sigma_{\hat{L}} \sqrt{2\pi}} \exp \left(- \frac{\hat{L}^2}{2\sigma_{\hat{L}}^2} \right) \frac{\beta_k}{\sigma_\varepsilon \sqrt{2\pi}} [\exp(-z_{j-1}^2) - \exp(-z_j^2)] d\hat{L} \\ &= \frac{\beta}{\sqrt{2\pi}} \left\{ \exp \left[- \left(\frac{cut_{j-1}^2}{2} \right) \right] - \exp \left[- \left(\frac{cut_j^2}{2} \right) \right] \right\} \end{aligned}$$

And for the last category ($j = J$):

$$E \left[\frac{\partial P(\text{ordinal score} = J)}{\partial x_{it}} \right] = \frac{\beta}{\sqrt{2\pi}} \exp \left[- \left(\frac{cut_{j-1}^2}{2} \right) \right]$$

This is also the relevant expression for a binary response variable.³⁰

Average treatment effects using adjusted-POLS

Here, an *average treatment effect* is the change in the probability of a specific outcome as x_{it} increases from one fixed number to another. The probability of attaining an ordinal value at j or below when x_{it} is set equal to some constant (c) is given by:

$$P(\text{ordinal score} \leq j | x_{it} = c) = N(\text{cutpoint}_j; \mu = \mu^*; \sigma = \sigma^*)$$

where μ^* is the expected value for the latent variable when x_{it} is set equal to c ; σ^* is the conditional standard deviation for the latent variable, i.e. conditional on x_{it} equaling c .

Example with simulations

Consider model (4) from section 4.2:

$$L_{it} = 0.2x_{it} + \alpha_i + \varepsilon_{it}$$

$$Var(L_{it}) = 1; Var(x_{it}) = 1; Var(\alpha_i) = 0.7; Cov(\alpha_i, 0.2x_{it}) = \sqrt{0.007}$$

Here, the unconditional variance for L_{it} is given by:

$$1 = Var(0.2x_{it}) + Var(\alpha_i) + 2Cov(\alpha_i, 0.2x_{it}) + Var(\varepsilon_{it})$$

Hence, the conditional variance for L_{it} is:

$$Var(\alpha_i) + Var(\varepsilon_{it}) = 1 - Var(0.2x_{it}) - 2Cov(\alpha_i, 0.2x_{it})$$

I estimate $2Cov(\alpha_i, 0.2x_{it})$ using its sample counterpart; $Var(0.2x_{it})$ is estimated with:

$$[b^2 - \widehat{Var(b)}] \widehat{Var(x)}$$

³⁰ For the notation above to be accurate in the binary case, we number these responses 1 and 2 (instead of 0 and 1).

where b is the adjusted-POLS estimate for the corresponding parameter (0.2).³¹ The conditional mean - μ^* - is easily estimated by replacing x_{it} with c and taking the average over the predicted values.

Table A3a presents the theoretical and estimated probabilities based on model (4). Here, I let the ordinal variable follow a skewed binomial distribution ($n = 4$; $p = 0.75$) and I number these values 1 through 5. Panel A presents the probability distribution over these values as $x_{it} = x_{it}$ and as $x_{it} = x_{it} + 1$; panel B presents the corresponding probability distribution as $x_{it} = 0$ and as $x_{it} = 1$. The sample size is set to 1000 ($i = 1, 2, \dots, 250$; $t = 1, 2, 3, 4$) and I make 10,000 repetitions for each simulated sampling distribution.

Table A3a reveals that adjusted-POLS – on average – does well in matching the theoretical values, while ordered probit is somewhat off. Example: The probability of attaining an ordinal value at five increases by 8.2 percentage points as x_{it} increases from zero to one. This is also the average estimate using adjusted-POLS; ordered-probit puts it at 7.5 percentage points. The pattern is similar when I estimate average marginal effects (not presented in table A3a).³²

Now, with a binary outcome, a probability-based measure is (perhaps) the only pertinent way of describing the effect size. For that reason, table A3b – panels A & B – repeats this exercise using a Bernoulli-distributed ordinal variable ($p = 0.2$). Here, I also compare with OLS. The results are similar to those presented in table A3a: On average, the adjusted-POLS probabilities match well with the theoretical values, while probit and OLS are somewhat off.³³ All estimators do a rather good job at estimating the average marginal effect, however.

It is worth noting, however, that adjusted-POLS relies on a rather strong assumption, i.e. that the latent variable is normally distributed. In panels C to F, this assumption is not fully met. In panel C, I estimate a model that is similar to (4) only that x_{it} is replaced by a treatment indicator: $T = 1$ if $x_{it} > 0$ and 0 otherwise; $\beta = 0.4$; $\text{Var}(\epsilon) \approx 0.126$. I refer to this as model (5). In panel D & E, I estimate a model that includes several normally distributed covariates as well as distinct time trends. I refer to this as model (6) and it is

³¹ The motivation for subtracting $\widehat{\text{Var}}(b)$ is described in section A.4.

³² The theoretical average marginal effects:

-0.2 (ordinal outcome #1), -1.9 (#2), -4.4 (#3), -0.6 (#4) and 7.1 (#5)

Adjusted-POLS matches these numbers, ordered probit puts them at:

-0.1 (#1), -1.2 (#2), -4.0 (#3), -1.5 (#4) and 6.8 (#5)

³³ Note that the average effect, the average marginal effect and the average treatment effect are the same measure when using OLS. Any differences observed in the panels are due to random chance, as each panel uses new samples.

described in more detail in the end of this section. In panel F, I estimate a model that is similar to (6) only that x_{1it} is replaced by a treatment indicator: $T = 1$ if $x_{1it} > 0$ and 0 otherwise; $\beta_1 = 0.4$; $\text{Var}(\varepsilon) \approx 0.097$. I refer to this as model (7). For model (5) and (7), I omit the ‘average effect’ as well as the ‘average marginal effect’ since the treatment indicator cannot take on values above one.

Overall, adjusted-POLS still performs well in estimating the effects of interest in these cases, although the underlying probabilities are usually overestimated. OLS is also a good alternative for estimating average marginal effects and average treatment effects, when the treatment is binary.

The performance of adjusted-POLS does start to suffer for sufficiently small samples. This is exemplified in panel F & G of table A3b which uses a sample consisting of 200 observations ($i = 1, 2, \dots, 100$; $t = 1, 2$) and estimates model (4). Due to regular non-convergence, the probit estimates are omitted here.

Below is a description of model (6) that I estimate in panel D & E of table A3b. The model includes four covariates as well as distinct time trends, combined with an IID error. The time-specific intercepts are uncorrelated with the other covariates and I control for them in the regressions. The model:

$$L_{it} = 0.2x_{1it} + 0.2x_{2it} - 0.2x_{3it} - 0.2x_{4it} + \alpha_i + \tau_t + \varepsilon_{it}$$

$$\alpha_i = 0.3\mu_{1it} + 0.3\mu_{2it} - 0.3\mu_{3it} - 0.3\mu_{4it} + e_i \quad \text{Var}(e_i) = 0.462$$

$$\tau_t = -0.4, 0, 0.1, 0.3$$

where μ_j is the expected value for the j :th covariate; the covariates are normally distributed and autocorrelated over time, with variances at unity and pairwise covariances at 0.2, but no correlation within panels. The error variance is 0.093.

Table A3a. Theoretical and empirical probabilities using adjusted-POLS and ordered probit in a panel data setting (10,000 repetitions). The ordinal variable is binomially distributed ($n = 4$, $p = 0.75$) numbered 1-5.

Panel A	Theoretical probabilities (%)			Adjusted-POLS probabilities (%)			Ordered probit probabilities (%)		
Ordinal value	$x = \underline{x}$	$x = \underline{x} + 1$	Diff.	$x = \underline{x}$	$x = \underline{x} + 1$	Diff.	$x = \underline{x}$	$x = \underline{x} + 1$	Diff.
1	0.4	0.2	-0.2	0.4 (0.3)	0.2 (0.2)	-0.2 (0.1)	0.4 (0.3)	0.2 (0.2)	-0.2 (0.1)
2	4.7	3.1	-1.6	4.7 (1.1)	3.1 (0.8)	-1.6 (0.4)	4.7 (1.1)	3.1 (0.9)	-1.6 (0.5)
3	21.1	16.8	-4.3	21.1 (2.1)	16.8 (2.0)	-4.3 (0.8)	21.1 (2.1)	16.8 (2.1)	-4.3 (1.1)
4	42.2	40.8	-1.4	42.2 (2.4)	40.8 (2.4)	-1.4 (0.6)	42.2 (2.4)	41.2 (2.7)	-1.0 (1.4)
5	31.6	39.1	7.4	31.6 (2.6)	39.1 (2.9)	7.4 (1.2)	31.6 (2.6)	38.7 (3.0)	7.0 (1.3)
Sum	100	100	0	100	100	0	100	100	0
Panel B	Theoretical probabilities (%)			Adjusted-POLS probabilities (%)			Ordered probit probabilities (%)		
Ordinal value	$x = 0$	$x = 1$	Diff.	$x = 0$	$x = 1$	Diff.	$x = 0$	$x = 1$	Diff.
1	0.1	0.1	-0.1	0.2 (0.1)	0.1 (0.1)	-0.1 (0.1)	0.2 (0.2)	0.1 (0.2)	-0.0 (0.1)
2	3.2	1.9	-1.3	3.2 (0.9)	1.9 (0.7)	-1.2 (0.3)	3.3 (1.0)	2.3 (0.9)	-1.0 (0.4)
3	20.4	15.4	-5.0	20.4 (2.1)	15.4 (2.1)	-5.0 (0.9)	20.3 (2.1)	15.4 (2.4)	-4.9 (1.3)
4	46.7	44.9	-1.8	46.7 (2.7)	44.8 (2.6)	-1.9 (0.8)	46.4 (2.7)	44.9 (3.0)	-1.5 (1.7)
5	29.6	37.8	8.2	29.6 (2.6)	37.8 (3.0)	8.2 (1.5)	29.8 (2.6)	37.3 (3.0)	7.5 (1.5)
Sum	100	100	0	100	100	0	100	100	0

Notes: The sample size is set to 1000 in all simulations ($i = 1, 2, \dots, 250$; $t = 1, 2, 3, 4$). All estimates are based on model (4), see section 4.2. Standard deviations in parenthesis.

Table A3b. The probability of a positive outcome using adjusted-POLS, probit and OLS in a panel data setting (10,000 repetitions). The ordinal variable is Bernoulli distributed ($p = 0.2$).

	Theoretical (%)	Adjusted-POLS (%)	Probit (%)	OLS (%)
Panel A				
$x = \underline{x}$	20.0	20.0 (2.2)	20.0 (2.2)	20.0 (2.2)
$x = \underline{x} + 1$	26.1	26.1 (3.0)	25.2 (2.7)	25.6 (2.9)
Difference	6.1	6.1 (1.8)	5.2 (1.3)	5.6 (1.6)
Avg. marg.	5.6	5.6 (1.6)	5.5 (1.5)	5.6 (1.6)
Panel B				
$x = 0$	17.2	17.2 (2.2)	17.4 (2.2)	19.9 (2.1)
$x = 1$	23.6	23.6 (2.7)	22.8 (2.5)	25.5 (2.8)
Difference	6.3	6.4 (2.0)	5.4 (1.4)	5.6 (1.6)
Panel C				
$T = 0$	12.3	12.6 (2.3)	13.1 (2.3)	14.4 (2.2)
$T = 1$	23.6	24.1 (2.7)	23.2 (3.1)	25.7 (2.7)
Difference	11.3	11.5 (2.9)	10.1 (3.4)	11.3 (2.8)
Panel D				
$x = \underline{x}$	20.0	20.0 (2.0)	20.0 (2.0)	20.0 (2.0)
$x = \underline{x} + 1$	26.1	26.1 (3.0)	25.1 (2.7)	25.6 (2.8)
Difference	6.1	6.1 (2.0)	5.1 (1.5)	5.6 (1.7)
Avg. marg.	5.6	5.6 (1.7)	5.2 (1.7)	5.6 (1.7)
Panel E				
$x = 0$	19.3	19.8 (2.0)	19.2 (2.1)	20.0 (2.0)
$x = 1$	25.4	25.9 (2.9)	24.1 (2.7)	25.6 (2.8)
Difference	6.2	6.2 (2.1)	4.9 (1.6)	5.6 (1.8)
Panel F				
$T = 0$	14.1	14.7 (2.1)	14.7 (2.1)	14.4 (2.2)
$T = 1$	25.5	26.0 (2.7)	24.2 (2.5)	25.6 (2.6)
Difference	11.3	11.3 (2.9)	9.5 (2.3)	11.3 (2.8)
Panel G				
$x = \underline{x}$	20.0	20.0 (3.6)	-	20.0 (3.6)
$x = \underline{x} + 1$	26.1	26.3 (6.7)	-	25.6 (6.2)
Difference	6.1	6.4 (5.5)	-	5.6 (4.7)
Avg. marg.	5.6	5.6 (4.7)	-	5.6 (4.7)
Panel H				
$x = 0$	17.2	17.7 (4.0)	-	19.9 (3.4)
$x = 1$	23.6	24.5 (5.7)	-	25.5 (6.0)
Difference	6.3	6.8 (6.1)	-	5.6 (4.7)

Notes: The Difference measures the ‘average effect’ or ‘average treatment effect’; Avg. marg. is short for the ‘average marginal effect’. Panel A & B: Based on estimates of model (4), see section 4.2; the sample size is set to 1000 ($i = 1, 2, \dots, 250$; $t = 1, 2, 3, 4$). Panel C: Based on estimates of model (5); see model (4) where x is replaced by a treatment indicator; the sample size is set to 1000. Panel D & E: Based on estimates of model (6) with several covariates and time-trends; the sample size is set to 1000. Panel F: Based on estimates of model (7); see model (6) where x_{1it} is replaced by a treatment indicator; the sample size is set to 1000. Panel G & H: Based on estimates of model (4); the sample size is set to 200 ($i = 1, 2, \dots, 100$; $t = 1, 2$). In panel F & G, the ordered probit probabilities are omitted due to regular non-convergence. Standard deviations in parenthesis.

A.7 Average percentile rank effect (APE)

Here I show that a movement up the L -distribution by β standard deviations is, on average, equal to an increase by $100 \beta / 2\sqrt{\pi}$ percentile ranks, assuming that L follows a normal distribution. In other words, the average percentile rank effect (APE)³⁴ is given by:

$$APE = E \left(\frac{\partial F(L)}{\partial x} \right) = \frac{\beta}{2\sqrt{\pi}}$$

where $F(L)$ is the distribution function for L , and β is the coefficient for x .

We start by noting that:

$$\frac{\partial F(L)}{\partial x} = f(L) \frac{\partial L}{\partial x} = \frac{\beta}{\sqrt{2\pi}} \exp(-0.5L^2)$$

Taking the expectancy on both sides gives us:

$$E \left(\frac{\partial F(L)}{\partial x} \right) = \frac{\beta}{\sqrt{2\pi}} E[\exp(-0.5L^2)] \quad (A.1)$$

where:

$$E[\exp(-0.5L^2)] = \int_{-\infty}^{\infty} [\exp(-0.5L^2)] f(L) dL = \int_{-\infty}^{\infty} \frac{\exp(-L^2)}{\sqrt{2\pi}} dL = \frac{1}{\sqrt{2}} \quad (A.2)$$

Substituting (A.2) into (A.1) gives us:

$$E \left(\frac{\partial F(L)}{\partial x} \right) = \frac{\beta}{2\sqrt{\pi}}$$

³⁴ Note that this is a general result applying also to models where the outcome is observable.

CHAPTER 3

Higher Education and the Gains in Cognitive Abilities

Abstract

This chapter uses PIAAC survey data to estimate the effect of schooling on cognitive abilities as measured by literacy and numeracy. The identification is threefold. Firstly, I exploit yearly variations in the cohort size as an instrument for schooling. Example: Some countries experienced a rising number of births in the early 1960s, but as these individuals reached adulthood, the number of college and university spots seems to have lagged behind. Exploiting this pattern, I estimate that each additional year of schooling adds 0.22 and 0.27 standard deviations to literacy and numeracy, respectively. Secondly, I use country-level data and exploit the variation in schooling-trends over age groups. As a third strategy, I measure the trends in literacy and numeracy in the years following graduation from upper-secondary school, comparing university students to others. The results from these two analyses paint a more pessimistic picture of the cognitive gains from higher education. I find that the trend for those who enroll in a university program is similar to the trend for those who move into the workforce or other activities, suggesting that university studies have small effects on literacy and numeracy (if any). I discuss potential explanations for the differences in the estimated effects.

3.1 INTRODUCTION

There is little doubt that education adds significantly to wages. Most studies find that one year of schooling increases wages by 5 to 15 percent (see, for example, the reviews by Belzil, 2007; Card, 1999, and Harmon et al., 2003). Less is known about the mechanism, i.e. whether the returns to schooling are mediated mainly through increases in cognitive skills and productivity, or whether higher education is mainly functioning as a signal for pre-determined ability.³⁵ In this chapter, I estimate the effect of schooling on cognitive skills as measured by literacy and numeracy (see chapter 1 for a description). Overall, I find evidence supporting the hypothesis that schooling has positive effects on cognitive skills. I measure this effect by contrasting three identification strategies: the instrumental variables approach, a country-level analysis and a ‘dose-response’ analysis.

Firstly, I use individual level data where schooling is instrumented by the *relative cohort size*: If you are born in a year when the cohort size is large in comparison to previous years, that would make competition for college and university spots tougher and entry into higher education less probable – assuming that the number of college and university places are fixed in the short run. This is also clearly visible in these data. Exploiting this pattern, I estimate that each additional year of schooling adds 0.22 and 0.27 standard deviations to literacy and numeracy, respectively. These are the medium- to long-term effects, with outcomes measured among individuals in the age range of 28 to 60.

Secondly, I exploit the variation in schooling that arises between individuals of varying ages (16-30). Here I use country-level data, comparing the development in literacy and numeracy with the increase in schooling, adjusted for the effects of aging. I find that one additional year of schooling adds 0.05 standard deviations to literacy and numeracy.

As a last strategy, I carry out a ‘dose-response’ analysis by comparing recently enrolled university students to those with a couple of years of university experience. I find only small differences in literacy and numeracy depending on university experience. Also, the trend for those who enroll in a university program is similar to the trend for those who move into the workforce or other activities after completing an upper-secondary degree. This suggests that university studies has small, if any, effects on literacy and numeracy.

Even though the estimates from these strategies point in the same direction, there is a qualitative difference in magnitudes that cannot easily be explained away by chance variation. While the instrumental variables

³⁵ See the discussion by Weiss, 1995.

approach suggests a sizeable effect, the country-level- and dose-response-analyses are more pessimistic. As these strategies exploit different kind of variation in schooling, it is plausible that heterogeneity in the underlying effect is part of the explanation. Also, since all strategies exploit observational data, they are open to critique regarding endogeneity.

3.2 BACKGROUND AND PREVIOUS RESEARCH

One of the main goals of education is for students to develop their cognitive skills; at least this is how educational reforms are often evaluated. So how well is this goal accomplished? Several studies have measured the relationship between schooling and cognitive performance, for example in the form of IQ-scores. There is little doubt of a positive correlation with Neisser et al. (1996) putting this number at ~ 0.55 based on the literature overall. It is unclear, however, to what degree schooling actually contributes to cognitive performance. In order to measure this effect, researchers mainly face two empirical challenges. When using data for children and adolescents, schooling is perfectly correlated with age, making it difficult to distinguish between the two.³⁶ When using data for adults, self-selection into higher education makes it difficult to separate the effect of pre-determined ability from the schooling-effect. Several recent studies have attempted to address these issues by exploiting arguably more exogenous sources of variation in schooling, such as the variation induced by educational reforms³⁷ (Banks & Mazzonna, 2012; Brinch & Galloway, 2012; Glymour et al., 2008; Meghir et al., 2013; Schneeweiss et al. 2014), by school entrance cutoff dates³⁸ (Cascio

³⁶ Alternatively, age is perfectly correlated with season of birth for students at the same grade level. Steltzl et al. (1995) measure the effect of schooling on IQ using ten year olds in the third and fourth grade, under the implicit assumption that season of birth has no influence on cognitive skills. They find that one additional year of schooling raises IQ-scores by 0.6 standard deviations. A similar approach was also used by Cahan & Cohen (1989).

³⁷ Glymour et al. (2008) exploit changes in the state-specific compulsory schooling laws during 1907-1961, and measure the effects on memory and mental status at an old age. They find a large effect (0.34 standard deviations) on memory; the effect on mental status is unclear due to large error bounds. The sample is restricted to white (non-Hispanic) Americans without a college degree that were born in 1900-1947. Banks and Mazzonna (2012) exploit the 1947 English educational reform which increased the minimum school-leaving age from 14 to 15 years. They find profound effects on late life cognition (memory and executive functioning) roughly equaling half a standard deviation. Brinch and Galloway (2012) uses a Norwegian primary school reform as an instrument for schooling. They find that one year of schooling increases IQ by 3.7 points (roughly 0.25 standard deviations). Meghir et al. (2013) find that the Swedish school reform that was introduced successively during 1949 to 1962, increased schooling by 0.23 years on average and cognitive skills by 0.07-0.15 standard deviations. The outcome was measured among 18 year old men. Using the reform as an instrument, one year of schooling translates into an increase in cognitive skills by 0.32-0.40 standard deviations for men to poorly educated fathers. However, the authors stress that this interpretation of the data could be misleading, as the reform transformed both the quantity and quality of schooling. Schneeweiss et al. (2014) exploit several European primary school reforms as an instrument for schooling. They use four different outcome measures (memory, fluency, numeracy and orientation-to-date) and measure effects among 60 year olds. They find that one additional year of schooling adds 0.1 standard deviations to memory, but find no significant effects on other outcomes.

³⁸ Leuven et al. (2004) find that one additional month of potential schooling increases test scores (language, math) by 0.06 standard deviations among disadvantaged children in the 2nd grade. They find no significant effects among non-disadvantaged children. (Note that this is 'potential schooling' as opposed to actual schooling, meaning that the actual effect is likely to be larger if

and Lewis, 2006; Leuven et al., 2004), by conditioning on earlier test-scores³⁹ (Falch & Sandgren Massih, 2011; Ritchie et al., 2013; Winship & Korenman, 1997), by exploiting quasi-random or conditionally random assignment of test dates⁴⁰ (Carlsson et al., 2012; Fitzpatrick et al., 2011) or by using control functions and structural models⁴¹ (Hansen et al., 2004).

Most of these studies use variation in schooling at the high school level or lower, and measure outcomes among adolescents. In this demographic, effect-sizes tend to fall somewhere in the neighborhood of ~0.2 standard deviations for one additional year of schooling (although there are exceptions). A couple of studies have also looked at late life cognition, with varying results. While Schneeweiss et al. (2014) find rather small or insignificant effects using several different cognitive outcome measures, Banks & Mazzonna (2012) find profound effects on memory and executive functioning amounting to roughly half of a standard deviation. Similarly, Glymour et al. (2008) find a large effect at 0.34 standard deviations for memory. All of these studies exploit educational reforms as a source of variation in schooling, and measure the outcomes among individuals several decades after graduation. Ritchie et al. (2013), on the other hand, measure the effect on IQ by conditioning on childhood IQ-scores; they find comparatively modest effect sizes at 0.16 and 0.06 standard deviations (for 70 and 79 year olds, respectively).⁴²

anything.) Cascio & Lewis (2006) find that one additional year of high school increases the Armed Forces Qualifying Test scores by 0.31-0.32 standard deviations among minority groups (Blacks and Hispanics); the effect for non-Hispanic whites is unclear. This study uses data from several states with varying school-entry cutoff dates, making it possible to control for season of birth.

³⁹ Winship & Korenman (1997) find that one additional year of schooling raises IQ-scores by 2.7 points (best guess 2-4 points) when holding early IQ-scores constant. Falch & Sandgren Massih (2011) find that one additional year of schooling increases IQ-scores at the age of 20 by roughly 0.2 standard deviations, when holding IQ at age 10 constant. Ritchie et al. (2013) find that one additional year of schooling increases IQ-scores at age 70 by 1.4 points (0.16 standard deviations) and at age 79 by 0.7 points (0.06 standard deviations), when holding IQ at age 11 constant. (The translation into standard deviations is my own, based on the standard deviations in the two samples.) They find no effects of schooling on late life processing speed.

⁴⁰ Fitzpatrick et al. (2011) find that one additional year of schooling increases test scores (reading, math) by one standard deviation for children in kindergarten and first grade (this is their conservative estimate). Carlsson et al. (2012) use data for 18 year old Swedish males, where the test date for army entrance is conditionally random. They find that 180 days (1 year) of schooling adds roughly 0.2 standard deviations to your cognitive skills.

⁴¹ Hansen et al. (2004) find that one additional year of schooling adds 2-4 percentage points to the Armed Forces Qualifying Test (0.16-0.19 standard deviations). This effect is somewhat larger for lower latent ability levels.

⁴² The conversion of the effect sizes into standard deviations is my own, based on the stated standard deviations for each sample. The raw effects are 1.4 and 0.7 points for 70 and 79 year olds, respectively.

These studies illustrate that there is little agreement on the magnitude of the schooling-effect in the medium- to long-term, and to my knowledge, no studies have looked specifically at the effects of post-secondary schooling. There are reasons to suspect that these distinctions are of importance. In the pessimistic scenario, the cognitive returns to education are short lived, with individuals regressing towards a default path in the years following graduation. For example, several studies have shown that achievement test scores tend to fall over the summer vacation⁴³, suggesting that some cognitive gains are temporary. In the optimistic scenario, on the other hand, the cognitive returns to education are amplified over time, for example due to further effects on career and lifestyle choices.⁴⁴

There are also reason to suspect that the cognitive returns to schooling might differ depending on educational level. Common sense suggests that elementary schools – teaching core abilities such as reading, writing and basic math – enhance cognitive skills at least in some sense. This is less obviously true for secondary schools, and even less so for colleges and universities which typically teach comparatively narrowly oriented subjects. Also, a couple of authors have suggested that your intelligence is more or less set by the time you reach adulthood (or even much earlier) which would suggest that the opportunities to enhance your cognitive skills are limited thereafter.⁴⁵

In comparison to the current literature, I exploit new sources of variation in schooling and measure both immediate, medium- and long-term effects. Furthermore, these data allow me to exploit variation in schooling at both the upper-secondary and tertiary levels. There are also some important limitations to this study. The IV-approach is only as good as the chosen instrument, and there is reason to suspect that your cohort size could affect both the quantity and quality of education. The country-level- and dose-response-analyses, on the other hand, are open to critique regarding confounding cohort-effects and measurement errors in schooling.

⁴³ See the review by Cooper et al., 1996.

⁴⁴ See Oreopoulos & Salvanes (2009) for lifetime effects of education on a range of outcome variables, such as work satisfaction, unemployment, health, marriage and parenting decisions, patience and risk behavior.

⁴⁵ See, for example, Jensen (1969).

3.3 DATA AND DESCRIPTIVE STATISTICS

I use PIAAC survey data that has been collected and compiled by OECD. My main sample covers roughly 77,000 adults in 25 countries⁴⁶ and has been collected during 2010-2015 using personal house interviews. The variables of central interest are those describing the respondents' cognitive abilities. Here I use two different measures – literacy and numeracy – designed to capture your ability to interpret text-based and mathematical information. See chapter 1 for a description.

My first identification strategy relies on exploiting variations in the national yearly cohort sizes as an instrument for schooling. To this end, I make the following restrictions on the sample: Each individual included was born in 1955 or later, and no one was younger than 28 years old at the time when the survey was conducted. I exclude older individuals as the instrument is not available for those born in 1954 or earlier; I exclude younger individuals as they may not have yet reached their highest educational degree. Also, I exclude anyone with missing values on key variables⁴⁷ or inconsistencies in their year of birth⁴⁸. Also, I generally exclude immigrants, unless they arrived early enough as to go through the whole educational system in the receiving country (arriving at the age of 0-5 years).⁴⁹ I include individuals with missing values on other background characteristics, and code these accordingly. This leaves me with a sample consisting of 76,709 individuals whose characteristics are summarized in table 1 below. Here, all means are weighted as to account for the country-specific survey designs; on a cross-country level, all countries are weighted equally according to the full sample (76,709 obs.).

⁴⁶ Belgium, Chile, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Greece, Ireland, Israel, Italy, Japan, Lithuania, the Netherlands, Norway, Poland, Russia, Slovakia, Slovenia, Spain, Sweden, Turkey and the United Kingdom.

⁴⁷ Including age, gender, years of schooling, literacy, numeracy or year of birth.

⁴⁸ You have an inconsistency in 'year of birth' if your age (combined with your birth-year) would suggest that the survey took place in 2009 or earlier, or 2016 or later, i.e. outside the actual time frame.

⁴⁹ Including immigrants who arrived later weakens the instrument slightly, but has essentially no effect on the main estimates.

Table 1. Weighted means for the individual characteristics

	All	Tertiary education	Below tertiary
Literacy (z-score ^a)	0.082	0.61	-0.22
Numeracy (z-score ^a)	0.097	0.62	-0.20
Female	0.50	0.54	0.48
Age	42.2	40.7	43.0
Schooling (years)	12.7	15.8	10.9
<i>Mother's education^b</i>			
Low	0.57	0.39	0.67
Medium	0.29	0.37	0.25
High	0.12	0.23	0.055
Missing	0.020	0.010	0.025
<i>Father's education^b</i>			
Low	0.49	0.32	0.59
Medium	0.32	0.36	0.29
High	0.15	0.30	0.071
Missing	0.040	0.020	0.049
<i>Immigration status</i>			
Non-immigrant	0.90	0.89	0.91
Immigrant	0.0094	0.0097	0.0092
Missing	0.091	0.10	0.081
Observations	76,709	29,479	47,230

Notes: The averages are weighted as to account for the country specific survey designs; on a cross-country level all countries are weighted equally according to the sample "All" (76,709 obs.). ^a z-scores are estimated using the full public sample of 16-65 year olds (roughly 200,000 individuals) with standard deviations estimated using within-country variation only. ^b Educational degrees: Low = Lower secondary school, primary school or less. Medium = Upper secondary and post-secondary non-tertiary degree. High = Tertiary degree.

Table 1 shows that those with tertiary degrees outperform those without such a degree; the raw difference is 0.83 and 0.82 standard deviations for literacy and numeracy, respectively.⁵⁰ There are, however, also other differences between the groups: Women and individuals with highly educated parents are overrepresented among those with tertiary degrees. Also, those with tertiary degrees are on average younger. Table 2 presents the schooling-effect adjusted for these differences, i.e. controlling for gender, age

⁵⁰ The standard deviations for literacy and numeracy have been calculated using the full PIAAC sample (roughly 200,000 individuals) and *within-country* variation only, i.e. I estimate the (weighted) variances separately for each country; the standard deviation is the square root of the average variance.

(using a second-degree polynomial), the mother's and father's education levels, immigration background and country of residence.

Table 2. The effect of schooling on literacy and numeracy (WLS-regressions)

	(1) Literacy (z-score)	(2) Literacy (z-score)	(3) Numeracy (z-score)	(4) Numeracy (z-score)
Schooling (years)	0.16*** (0.0016)	0.12*** (0.0014)	0.16*** (0.0019)	0.13*** (0.0015)
Observations	76,709	76,709	76,709	76,709
R-squared (overall)	0.234	0.357	0.248	0.383

Notes: (1) & (3) Includes no controls, (2) & (4) Controls included for gender, age (using a second degree polynomial), parental education levels, immigration status and country of residence. All observations are weighted as to account for the country specific survey designs; on a cross-country level all countries are weighted equally. Standard errors (in parenthesis) are estimated using jackknife replicate sampling weights and further adjusted for the imputation variance induced by using plausible values. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 2 shows that one additional year of schooling is associated with an increase in literacy and numeracy by 0.16 standard deviations (see (1) & (3)). This effect is partly explainable by differences in background characteristics: After adding controls, the estimates drop to 0.12 and 0.13 standard deviations for literacy and numeracy, respectively (see (2) & (4)). These effects are of similar magnitude for men and women⁵¹ and over different age groups⁵² (not presented in table 2).

Note that these estimates cannot be interpreted as causal effects. If anything, I would expect these estimates to be positively biased, as I have not (fully) controlled for differences in pre-determined ability. In the next section, I turn to the instrumental variables approach in order to exploit presumably more exogenous sources of variations in schooling.

⁵¹ For women, one additional year of schooling is associated with an increase in literacy by 0.12 standard deviations and in numeracy by 0.13 standard deviations (after adjusting for differences in background characteristics). The corresponding effects for men are 0.13 and 0.14 standard deviations.

⁵² One additional year of schooling is associated with an increase in numeracy by 0.13 standard deviations (for 28-37 year olds); by 0.14 standard deviations (for 38-47 year olds), and by 0.14 standard deviations (for 48-60 year olds).

3.4 THE INSTRUMENTAL VARIABLES APPROACH

In this section, I use the *relative cohort size* as an instrument for schooling. I find that one year of schooling adds 0.22 and 0.27 standard deviations to literacy and numeracy, respectively (95 % confidence interval: 0.14-0.30 for literacy and 0.19-0.35 for numeracy). These are rather large effects: An effect at 0.22 standard deviations is comparable to an average increase by 6 percentile points. Similarly, 0.27 standard deviations translates into an average increase by 8 percentile points.⁵³

3.4.1 The model

The instrument is your *relative cohort size*. This is measured as the deviation in the number of births in a given year from the average based on the five previous years.⁵⁴ The number of births are measured on a logarithmic scale (the natural logarithm):

$$relative\ cohort\ size_{ct} = \ln(births_{ct}) - \frac{1}{5} \sum_{i=1}^5 \ln(births_{c(t-i)})$$

where c is a country index and t is indexing year of birth, $t = 1955, 1956, \dots, 1987$. Example: If the relative cohort size is 0.01 then that cohort is, on average, one percent larger than the cohorts in the five previous years.⁵⁵ Figure 1 below shows the development in relative cohort sizes separately by country and overall. Typically, a cohort deviates from the moving average by no more than plus/minus 15 percent. South Korea is the outlier, having exceptionally large cohort sizes in the years following the Korean war.

⁵³ The 'average percentile rank effect' (APE) is given by:

$$APE = \frac{\beta}{2\sqrt{\pi}}$$

assuming that literacy and numeracy are normally distributed. Example: For literacy, the average percentile rank effect is $0.22/(2\sqrt{\pi}) \approx 0.06$. For reference, see chapter 2, section 5.

⁵⁴ Data source: UN population division, World Population Prospects 2015.

⁵⁵ This interpretation is approximate: it would be accurate if the past five cohorts were all equally large.



Figure 1. Deviation in the number of births from five year moving average – by country and overall

Notes: The deviation is measured on a logarithmic scale. Solid line represents the overall development; the dotted lines are the country-specific trends.

If you are born in a year when the cohort is large in comparison to previous years, that would make entry into higher education less probable, assuming that the supply of college and university spots is fixed in the short run. This pattern is also visible in the data, as illustrated in figure 2. Here I predict years of schooling as a function of your relative cohort size divided into six quantile groups; I rely on within-country and within-birth-year variation only. Among those belonging to the smallest cohorts, 38.2 percent attain a tertiary degree and the average education is 12.8 years; among those belonging to the largest cohorts, the corresponding numbers are 35.1 percent and 12.6 years.⁵⁶

⁵⁶ Here, the largest cohorts (the 6th quantile group) are, on average, 7 percent above the five-year average; the smallest cohorts (the 1st quantile group) are 9 percent below the five-year average.

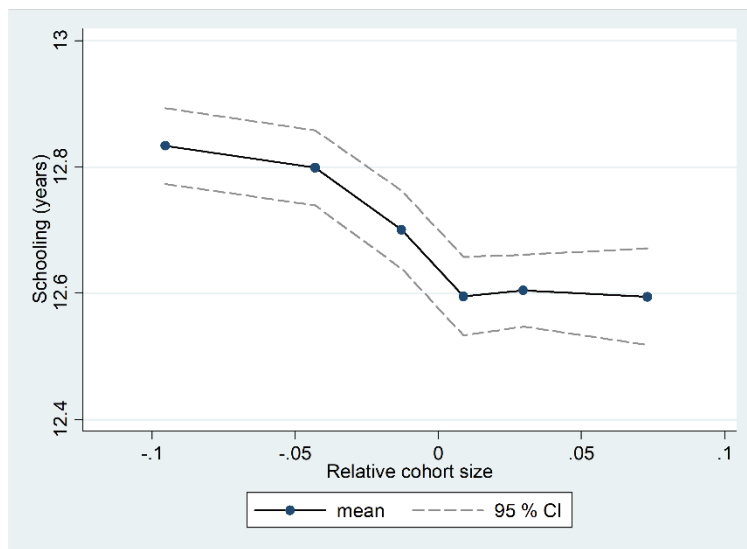


Figure 2. The relationship between the relative cohort size and schooling

Notes: Years of schooling is estimated as a function of your relative cohort size using within-country and within-birth-year variation only. The relative cohort size is divided into six quantile groups. The x-axis displays the average relative cohort size in that quantile group.

The relative cohort size is a valid instrument for schooling if any differences in skill performance between individuals from large and small cohorts can be attributed to educational differences alone. If, for example, relative cohort sizes tend to be larger among individuals born in the 1960s as compared to those born in the 80s, that could invalidate the instrument unless year of birth is appropriately controlled for. Similarly, if baby booms tend to go hand in hand with economic upturns, that could also invalidate the instrument unless the economic environment at birth is appropriately controlled for.⁵⁷ Here, I specify the *reduced form equation*⁵⁸ in the following way:

$$cog_{cti} = \beta size_{ct} + \vartheta X_{cti} + Z_{ct} + \varepsilon_{cti}$$

⁵⁷ A couple of studies have described procyclical movements in birthrates: see, for example, Sobotka et al. (2011), Hofmann & Hohmeyer (2012) and Goldstein et al., (2013). This correlation is also present in this data set. Furthermore, there is some evidence that the economic environment at birth has effects on future outcomes, with recessions correlating with worse cognitive performance late in life (Dobhammer et al., 2013).

⁵⁸ The *reduced form equation* models the *outcome* (literacy/numeracy) as a function of the instrument (relative cohort size) and other covariates. The *first stage equation* models the *independent variable of interest* (schooling) as a function of the instrument and other covariates (i.e. the same set as in the reduced form equation). The two-stage-least squares estimate is given by the ratio between the reduced form and first stage estimates ($\hat{\beta}_{reduced\ form}/\hat{\beta}_{first\ stage}$) where β is the coefficient for the instrument.

where c is a country index ($c = 1, 2, \dots, 25$); t is indexing year of birth ($t = 1955, 1956, \dots, 1987$) and i is indexing individuals. The outcome, cog_{cti} , is either literacy or numeracy; $size_{ct}$ is the relative cohort size and X_{cti} is a vector of background characteristics including gender, age (and a quadratic polynomial in age), parental education levels, and immigration background (see table 1); ε_{cti} is an individual-specific error term. The variable Z_{ct} captures the effect of being born in year t and residing in country c . I mainly consider the following model for Z_{ct} :

$$Z_{ct} = \delta_c + \tau_t + \theta_1 lgdp_{ct} + \theta_2 lgdpgap_{ct} + u_{ct}$$

where δ_c represents country-specific fixed intercepts and τ_t captures any general cohort effects. I also control for the macroeconomic environment at birth as measured by GDP per capita ($lgdp_{ct}$) and the GDP gap ($lgdpgap_{ct}$), i.e. the deviation from potential GDP per capita. The trend in potential GDP is estimated from data separately for each country using the Hodrick-Prescott filter.⁵⁹ Both GDP-variables are measured in real numbers using logarithmic scales. u_{ct} is the error term specific to that country and cohort. In the robustness section I also consider alternative specifications for Z_{ct} .

All estimates are weighted as to account for the country specific survey designs; on a cross-country level, all countries are weighted equally. Standard errors are estimated using jackknife replicate sampling weights. In the reduced-form and instrumental variables regressions I further correct the standard errors for the imputation variance added by using plausible values.

3.4.2 Main results

I find that the relative cohort size has a significant effect on schooling: when the relative cohort size increases by ten percent, schooling decreases by 0.2 years on average.⁶⁰ This is the first stage estimate, also presented in table 3. The relative cohort size is a strong instrument, with an F-value ($df = 1, 79$) at 105.6.⁶¹ Furthermore, every ten percent increase in the relative cohort size leads to a decrease in literacy by 0.045 standard deviations and in numeracy by 0.055 standard deviations.⁶² These are the reduced form estimates, also presented in table 3. Both effects are significant ($p < 0.01$).

⁵⁹ Data source (GDP): Penn World Tables.

⁶⁰ $(-2.12) \cdot \ln(1.1) \approx -0.2$.

⁶¹ Following the rule of thumb ($F > 10$) as suggested by Stock & Yogo (2002).

⁶² $(-0.47) \cdot \ln(1.1) \approx -0.045$; $(-0.58) \cdot \ln(1.1) \approx -0.055$.

Table 3. First stage and reduced form regressions

	Schooling (years)	Literacy (z-score)	Numeracy (z-score)
Relative cohort size	-2.12*** (0.21)	-0.47*** (0.090)	-0.58*** (0.089)
Observations	76,709	76,709	76,709
R-squared (overall)	0.303	0.256	0.261
F-value (1, 79)	105.6	-	-

Notes: Relative cohort size measures the number of births in the year you were born in comparison to an average based on the five previous years, where the number of births is measured on a logarithmic scale. All regressions control for gender, age (using a second-degree polynomial), parental education levels and immigration background, as well as the effects of being born in country c in year t , as modeled by Z . All regressions are weighted as to account for the country specific survey designs; on a cross-country level all countries are weighted equally. Standard errors (in parenthesis) are estimated using jackknife replicate sampling weights and further adjusted for the imputation variance in the reduced form regressions. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Taken together, we have that a ten percent increase in the relative cohort size decreases schooling by 0.2 years, literacy by 0.045 standard deviations and numeracy by 0.053 standard deviations. Thus, one full year of schooling is translated into an increase in literacy by 0.22 standard deviations and in numeracy by 0.27 standard deviations (95 % CI: 0.14-0.30 for literacy and 0.19-0.35 for numeracy). These are the instrumental variables estimates, also presented in table 4. The schooling-effect is of similar magnitude for men and women.

From a practical point of view, these effects are rather large: Adding one year of schooling to your education is comparable to a movement up the numeracy distribution by 8 percentile points. Previous research suggests that cognitive gains of these magnitudes also have effects on wages: Adding 0.27 standard deviations to your numeracy score is associated with an increase in hourly wages by roughly 3 percent. This figure is based on the pooled estimate in Hanushek et al. (2015) where a one standard deviation increase in numeracy corresponds to an increase in hourly wages by 10.7 percent when holding work experience, gender and schooling constant.

Table 4. The effect of schooling on literacy and numeracy (IV-estimates)

All	Literacy (z-score)	Numeracy (z-score)
Schooling (years)	0.22*** (0.041)	0.27*** (0.041)
Observations	76,709	76,709
By gender	Numeracy Women	Numeracy Men
Schooling (years)	0.26*** (0.047)	0.28*** (0.060)
Observations	40,863	35,846
F-value (first stage)	49.5	50.3

Notes: Both regressions control for gender, age (age, age²), parental education levels and immigration background, as well as the effects of living in country c and being born in year t, as modeled by Z. All regressions are weighted as to account for the country specific survey designs; on a cross-country level all nations are weighted equally. Standard errors (in parenthesis) are estimated using jackknife replicate sampling weights and further adjusted for the imputation variance. *** p<0.01, ** p<0.05, * p<0.1

How lasting is the cognitive effect of schooling? One possibility is that this effect is amplified with age, for example due to further effects on career paths and other life-style choices. Another possibility is that the process of aging triumphs in the long-run, making individuals regress towards a default path. Naturally, there is also the possibility of a middle-ground, where the educational gains are preserved more or less unchanged with age. In order to discriminate between these possibilities, I instrument schooling as well as schooling interacted with age and age squared.⁶³ The estimates suggest that the cognitive effect increases at a diminishing rate, peaking around age 50.⁶⁴ See the Appendix (section A.1 Heterogeneity) for the estimates.

From a policy perspective, it is of special interest to recognize potential heterogeneity in effects with regard to social background. If the cognitive returns to education are especially high in some socioeconomic groups, then that could warrant for targeted policies. However, I find no evidence of differential effects depending on parental education levels (see table A1 in the Appendix). The error margins are wide, however, leaving room for interpretations.

⁶³ The instrument is given by your relative cohort size, as well as your relative cohort size interacted with age and age squared.

⁶⁴ At the age of 30, the cognitive gain of schooling is insignificant at -0.02 standard deviations (se = 0.42). At the age 40, this effect has grown to 0.21 standard deviations (se = 0.23); at 50 the effect is 0.27 standard deviations (se = 0.049) and at 60 0.16 standard deviations (se = 0.20).

3.4.3 Robustness and validity

The estimates are robust between different variants: The results do not hinge upon including or excluding South Korea⁶⁵, or upon the details on how to measure ‘relative cohort sizes’⁶⁶. I also experiment with using two instruments (the relative cohort size & the relative cohort size squared) as suggested by Dieterle & Snell (2016) as a test of validity; this leaves the main estimate unchanged.

Of special interest for the identification is the function describing Z_{ct} which captures the country and cohort effects. I experiment with the following four alternative specifications:

$$Z_{ct} = \delta_c + \tau_t + u_{ct} \quad (i)$$

$$Z_{ct} = \delta_c + \tau_{0t} + \tau_{1t}lgdp_{ct} + \tau_{2t}lgdpgap_{ct} + u_{ct} \quad (ii)$$

$$Z_{ct} = \delta_{0c} + \tau_{0t} + \tau_{1t}lgdp_{ct} + \tau_{2t}lgdpgap_{ct} + \delta_{1c}lgdp_{ct} + \delta_{2c}lgdpgap_{ct} + u_{ct} \quad (iii)$$

$$Z_{ct} = \delta_c + \gamma_c t_t + \theta_1lgdp_{ct} + \theta_2lgdpgap_{ct} + u_{ct} \quad (iv)$$

(i) Here, I only control for country of residence (δ_c) and year of birth (τ_t) where u_{ct} represents the country- and cohort-specific error. This specification differs from the main specification only in that I exclude controls for the economic environment at birth.

(ii) Here I build on the main specification, but also allow for the cohort effects to vary by the economic environment at birth. In other words, the effect of being born in, say, 1960 is now allowed to differ between countries that were

⁶⁵ South Korea is of special interest here due to the peculiar trend in relative cohort sizes after the Korean war (see figure 1). Excluding South Korea increases the effect on numeracy from 0.27 to 0.31 standard deviations.

⁶⁶ I experiment with measuring the ‘relative cohort size’ as the deviation in the number of births in a given year from a *weighted* average based on the five previous years:

$$relative\ cohort\ size_{ct} = \ln(births_{ct}) - \sum_{i=1}^5 \frac{6-i}{15} \ln(births_{c(t-i)})$$

where c is a country index and t is indexing year of birth, $t = 1955, 1956, \dots, 1987$. This leaves the estimate practically unchanged. Furthermore, exploiting the last ten years (as opposed to five years) makes no qualitative difference (the effect increases from 0.27 to 0.35). I also experiment with calculating the relative cohort size as a log deviation from the hodrick-prescott trend. Here, the instrument was too weak for any robust inference.

rich and poor at the time, and between countries that were in economic upturns and downturns at the time. These effects are captured by the vectors τ_{1t} and τ_{2t} ; τ_{0t} is a vector of parameters setting the level for each cohort.

(iii) Here I further allow for the effects of the economic environment at birth to vary by country. For example: The effect of being born in a recession is now allowed to differ between, say, Finland and Turkey. These effects are captured by the vectors δ_{1c} and δ_{2c} ; δ_{0c} is a vector of parameters setting the level for each country.

(iv) This is similar to the main specification, but here I replace the general cohort effects with country-specific linear time trends in year of birth ($\gamma_c t_t$).

The estimates are qualitatively robust between these specifications, with the effect on numeracy varying between 0.23 and 0.29 standard deviations (see table 5). Table 5 also presents the results from two other specifications where I experiment with excluding different sets of individual characteristics: firstly parental education levels and secondly, the full set of individual characteristics including also gender, age and immigration status.⁶⁷ This has practically no effect on the estimate.

Naturally, any tests of robustness can only include a small subset of possible specifications. The estimated gains from schooling do not seem to disappear easily, however. Nevertheless, there is room for interpretation of this finding. Note that the instrument is valid if the difference in cognitive skills between individuals from small and large cohorts can be attributed to educational differences alone. It is, however, possible that the relative cohort size also affects other aspects of your upbringing such as educational quality. Most importantly, being born in a ‘popular year’ could correlate with having many classmates which potentially affects your future literacy and numeracy scores negatively.⁶⁸ In other words, this could potentially bias the estimates upwards, i.e. overstating the effect of education. At this point, it is worth making the distinction between the estimates and their interpretation: The data shows that being born as part of a relatively large cohort shows up as you attaining significantly less education and performing significantly worse on future literacy and numeracy tests. The interpretative part is that this drop in skills is brought about by less education.

Also, it is worth noting that even a valid instrument can only claim to estimate the so called ‘local average treatment effect’, meaning that the effect

⁶⁷ Here I consistently use the main specification for Z.

⁶⁸ A couple of randomized trials, as well as natural experiments, support the hypothesis of class sizes having meaningful impacts on test performance, both immediately (see, for example, Mosteller, 1995 and Angrist & Levy, 1999) and in the long-run (see, for example, Fredriksson et al., 2012).

is identified using only a subsample of individuals, i.e. those affected by the added competition of being part of a larger cohort.⁶⁹ In other words, the most academically motivated and gifted students are unlikely to contribute, as well as the academically unmotivated. If the effects of schooling are heterogeneous in this respect, then the instrumental variables estimate could be a poor estimate of the average treatment effect. From a policy perspective, this isn't necessarily a disadvantage: The students influenced by cohort sizes are also likely to be the students first gaining access to higher education if admission rates were to increase (or losing access if admission rates were to fall). These data also indicate who this 'marginal student' could be, i.e. someone with relatively poorly educated parents: The instrument does not significantly predict schooling for those with highly educated parents; dropping this group has practically no effect on the estimates.

⁶⁹ The 'local average treatment effect' is described by Imbens & Angrist (1994).

Table 5. The effect of schooling on numeracy using alternative specifications (IV-estimates)

	Main	(i)	(ii)	(iii)	(iv)	Excluding parental education	Excluding all individual characteristics
Schooling (years)	0.27*** (0.041)	0.23*** (0.035)	0.29*** (0.041)	0.29*** (0.082)	0.26** (0.11)	0.28*** (0.053)	0.29*** (0.052)
Observations	76,709	76,709	76,709	76,709	76,709	76,709	76,709
F-value (first stage)	105.6	143.3	104.22	29.8	14.4	51.1	50.5

Notes: 'Main', (i), (ii), (iii) and (iv) use different specifications for Z (which captures the country- and cohort-specific effects). The last two columns use the main specification for Z but successively exclude sets of individual characteristics. All regressions are weighted as to account for the country specific survey designs; on a cross-country level all nations are weighted equally. Standard errors (in parenthesis) are estimated using jackknife replicate sampling weights and adjusted for the imputation variance added by using plausible values. *** p<0.01, ** p<0.05, * p<0.1

3.5 A COUNTRY-LEVEL ANALYSIS

In this section, I explore an alternative identification strategy: I measure the effect of schooling by exploiting the variation that arises between cohorts of young adults (going from the age of 16 to 30). Here I use country-level data, comparing the development in literacy and numeracy with the increase in schooling, adjusted for the effect of aging. I find that one additional year of schooling adds 0.05 standard deviations to literacy and numeracy.

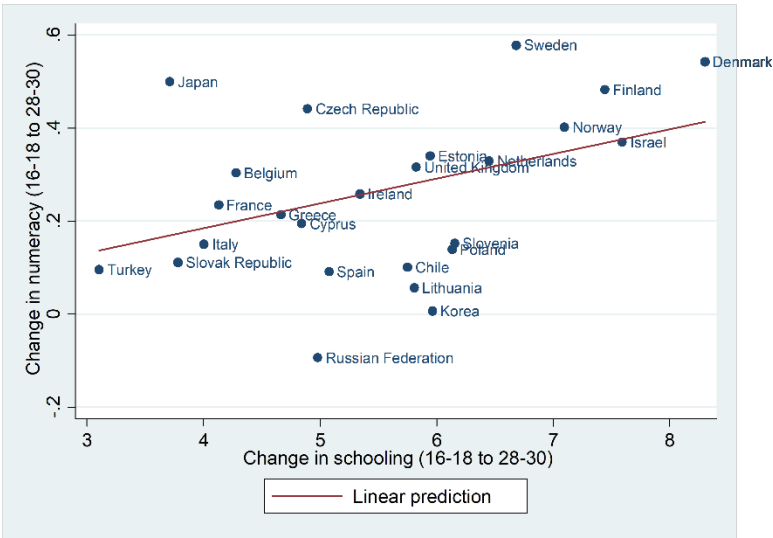


Figure 3a. The increase in schooling when comparing 16-18-year olds to 28-30-year olds, and its relationship to the growth in numeracy (country-level data)

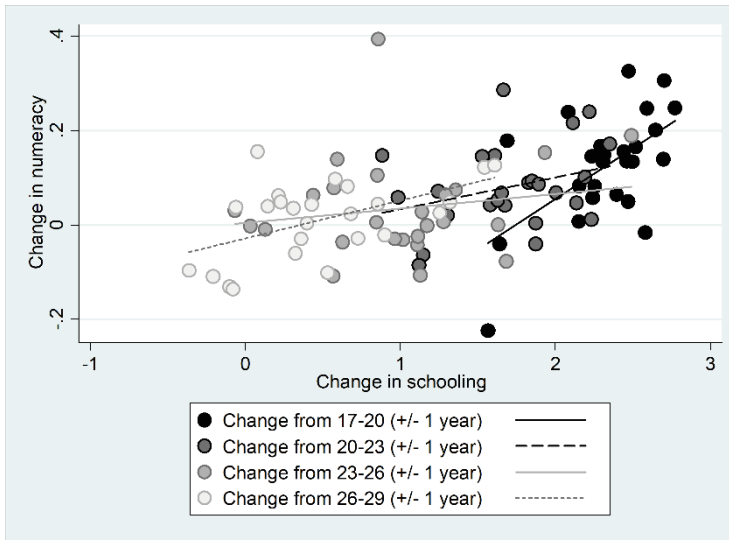


Figure 3b. The increase in schooling when comparing one co-hort to the next, and its relationship to the growth in numeracy (country-level data).

Note: One cohort consists of individuals of three ages: 16-18-year olds, 19-21-year olds, ... and 28-30-year olds.

Figure 3a illustrates the change in schooling when comparing 16-18-year olds to those in the age range of 28-30. For example, Turks add roughly three years of schooling to their education over this period, while Danes add more than eight years. This increase in schooling is also significantly correlated with the growth in numeracy, $r = 0.43$ ($p < 0.05$). Figure 3b illustrates the same relationship, but exploits also the intervening cohorts. Hence, the x-axis illustrates the change in schooling when comparing one cohort to the next; the y-axis illustrates the corresponding change in numeracy. Here, a ‘cohort’ is three years long: 16-18-year olds, 19-21-year olds, ... and 28-30-year olds. The countries where the increase in schooling is large – going from one cohort to the next – are also the ones experiencing large growths in numeracy, $r = 0.52$ ($p < 0.01$). This is especially so for the youngest cohorts, $r = 0.59$ ($p < 0.01$). In the following sections, I model and estimate this relationship more formally.

3.5.1 The model

In order to estimate the effect of schooling, I measure the development in skills within countries and over cohorts, adjusted for the effects of aging:

$$skill_{ca} = \alpha_c + \alpha_a + \beta schooling_{ca} + \vartheta pared_{ca} + \varepsilon_{ca}$$

where c is a country index and a is indexing age groups, $a = 16-18, 19-21, 22-24, 25-27$ and $28-30$. The outcome, $skill_{ca}$, is either average literacy or average numeracy; α_c and α_a represent country and age group fixed effects; $schooling_{ca}$ is the average years of schooling and $pared_{ca}$ is parental education level, measured as the share of individuals with at least one highly educated parent; ε_{ca} is the error term. All averages are weighted as to account for the country specific survey designs. I use heteroscedasticity-robust standard errors clustered on the country level.

In this model, β represents the causal effect of schooling, assuming that the differences in schooling from one cohort to the next are uncorrelated with other changes that are of importance for skill development. For example: Finns at the age of 16-18 have 10.5 years of schooling on average which increases to 13.0 years at the age of 19-21. Here, the cohort of 16-18-year olds are assumed to be comparable to the cohort of 19-21-year olds, after adjusting for the difference in age and parental education level. This assumption is more likely to hold when the means (such as average years of schooling among 16-18-years olds in Finland) are based on large sample sizes, so that the variation in pre-determined ability on the individual level cancels out on the cohort level. On average, I use 355 observations to calculate one such mean (min: 161; max: 2086).

3.5.2 Result

I find that one additional year of schooling adds 0.05 standard deviations to literacy and numeracy. The estimates are presented in table 6, where the first specification (1) controls for country of residence only; the second specification (2) adds cohort fixed effects and the third specification (3) adds parental education level. The table shows that the schooling-effect is rather robust to the inclusion of these controls, changing by no more than 0.015 standard deviations. These estimates support the hypothesis that higher education contributes to the development in literacy and numeracy. The effects are modest, however, especially compared to the IV-estimates (see section 4.2).

Figure 3b above hinted that the schooling-effect might be rather small after the age of ~ 18 . Dropping the youngest cohort (16-18 year olds) does pull the estimates down to 0.038 for numeracy and 0.029 for literacy (3), but the error margins are too wide to make any strong conclusions.

Table 6. The effect of average schooling on average literacy and numeracy (fixed effects regressions)

	(1) Literacy	(2) Literacy	(3) Literacy
Schooling	0.034*** (0.0059)	0.049*** (0.018)	0.045** (0.019)
Par. education	-	-	0.24 (0.29)
Cohort FE	No	Yes	Yes
Observations	125	125	125
Countries	25	25	25
R ² (within)	0.432	0.450	0.455
	(1) Numeracy	(2) Numeracy	(3) Numeracy
Schooling	0.047*** (0.0052)	0.055*** (0.019)	0.050** (0.020)
Par. education	-	-	0.30 (0.25)
Cohort FE	No	Yes	Yes
Observations	125	125	125
Countries	25	25	25
R ² (within)	0.615	0.621	0.628

Notes: Schooling is the average number of years of schooling; Parental education is the share of individuals with at least one highly educated parent (tertiary degree). A 'cohort' is three years long (16-18, 19-21, ..., 28-30). Standard errors (in parenthesis) are heteroscedasticity-robust and clustered on the country level; all regressions include country fixed effects.
*** p<0.01, ** p<0.05, * p<0.1

3.5.3 Validity and bias

There are mainly two threats against validity. Firstly, the model does not control for differences in cohort-specific factors that are not picked up by age or parental education. To the degree such cohort effects are present, these are likely to bias the schooling-effect upwards: If some cohorts have higher pre-determined ability levels than others, this would presumably affect both schooling and the skill outcomes positively. One central cause for such cohort-effects is that *individuals* differ with regard to pre-determined ability. For finite cohort sizes, the variation in pre-determined ability on the individual level will not completely cancel out on the cohort level, hence it is likely to be picked up as part of the schooling-effect.

What can we say regarding the size of this bias? Under general assumptions, it will be smaller than the corresponding bias using individual-level data, and it tends towards zero as the sample size on the group-level

increases. (Here, a *group* denotes every one of the same age living in the same country.) To see this, we begin by describing the bias on the individual level. Assume we want to estimate β as defined by the model:

$$numeracy = \alpha_c + \beta schooling + \gamma ability + \theta age + u$$

where α_c represents country-specific intercepts and u is the individual error term. *Schooling*, in turn, is assumed to be a function of your *group* (α_g) and your *ability*. I set the effect of *ability* to equal unity without loss of generality:

$$schooling = \alpha_g + ability + \varepsilon$$

Lacking data on *ability*, the omitted-variable-bias equals $\delta\gamma$, where δ is the conditional effect of *schooling* on *ability*, i.e. conditional on *age* and *country*. This effect (δ) can be described as the ratio between the conditional variance for *ability* and the conditional variance for *schooling*:

$$bias = \delta\gamma = \frac{Var(ability|age, country)}{Var(schooling|age, country)} \cdot \gamma \quad (1)$$

assuming that α_g does not correlate with *ability* conditional on *age* and *country*. Example: If Finland has a steeper aging-trend in schooling than the United Kingdom, then this does not reflect a steeper aging-trend in ability among Finns. I derive expression (1) in the Appendix (A.4).

Now, turning to group-level data, the corresponding bias becomes:

$$bias' = \delta'\gamma = \frac{Var(\overline{ability}|age, country)}{Var(\overline{schooling}|age, country)} \cdot \gamma$$

where δ' is the effect of *average schooling* on *average ability* with averages measured on the group-level. This bias, $\delta'\gamma$, can be expressed as:

$$\delta'\gamma = \frac{\sigma_v^2 \cdot \gamma}{m \cdot \sigma_{g|age, country}^2 + \sigma_v^2 + \sigma_\varepsilon^2} \quad (2)$$

$$< \delta\gamma = \frac{\sigma_v^2 \cdot \gamma}{\sigma_{g|age, country}^2 + \sigma_v^2 + \sigma_\varepsilon^2}$$

where m is the sample size on the group level, assumed to be the same for all groups; $\sigma_{g|age, country}^2$ is the conditional variance for the group-effects, i.e. this variance reflects the differences in schooling-trends between countries; σ_v^2 is the within-country variance in ability. See the Appendix (A.4) for the intermediate steps.

These equalities reveal two points. Firstly, when using group-level data, the bias tends towards zero as the group size, m , increases (assuming

$\sigma_{g|age,country}^2 > 0$). Also, for any finite group size ($m > 1$) the bias will be smaller when using group-level data as compared to individual-level data (assuming $\sigma_{g|age,country}^2 > 0$). Furthermore, these expressions allow me to estimate how much the omitted variable bias falls by turning to group-level data. Here, the bias is predicted to fall by 95 percent when using groups consisting of 355 individuals (the average for this data set).⁷⁰ In other words, it is reasonable to assume that this kind of omitted variable bias is significantly suppressed here.

Omitted variable bias is one threat against validity, potentially pulling the estimates upwards. Perhaps more importantly, however, schooling is likely measured with error, potentially pulling the estimates downwards. This measurement error arises since schooling is not directly observable for individuals who are still in school. Instead, I infer ‘years of schooling’ using the assumption that you stay in school until you achieve your highest degree or drop out, after which you never reenter the educational system again.⁷¹ For students, schooling equals your age; for non-students, schooling equals your age at graduation or dropping out.⁷² This also implies that countries where students regularly have gap years or work while studying will have steeper schooling-trends only for those reasons.

⁷⁰ The bias-ratio can also be expressed as:

$$\frac{\delta'\gamma}{\delta\gamma} = \frac{Var(schooling|age, country)}{Var(schooling|age, country) \cdot m}$$

I replace m with the average group size (355) and the variances with their sample counterparts. This gives me a ratio at ~ 0.05 , i.e. a bias-reduction by 95 percent.

⁷¹ Furthermore, I assume that everyone from the same country enter school at the same age, which is realistic on a group level.

⁷² Naturally, schooling is now overestimated for everyone. However, this error is equally large for everyone from the same country, assuming they started school at the same age. Hence, this error does not affect the estimates. However, there are also some individuals who are *enrolled* in an educational program but do not consider themselves students (but rather workers or home-keepers, for example). For these individuals, I assume that you have been a student up until last year.

3.6 THE TREND IN COGNITIVE SKILLS AMONG UNIVERSITY STUDENTS

In this section, I measure the trends in literacy and numeracy in the years following graduation from upper-secondary school. I find that the trend for those who enroll in a university program is similar to the trend for those who move into the workforce or other activities. This suggests that university studies have small immediate effects on literacy and numeracy if any; the large gaps between the groups is better explained by self-selection.

In order to measure these trends, I select individuals who graduated from upper-secondary school⁷³ at the age of 17 to 21. Some of these individuals continue onto university programs while others continue into working life or other activities.⁷⁴ I observe individuals 1 to 4 years later when some have accumulated a couple of years of university studies while others have accumulated a couple of years of other experiences. I compare the trends in literacy and numeracy over these years and groups. I also control for country of residence, gender, immigration status, age at graduation (17-21), upper-secondary degree (converted into years of schooling) and parental education levels. The result is presented in table 7, where panel A shows estimates for the full sample and panel B separately by gender (for literacy only). I find that university students gain ~0.03 standard deviations in literacy and numeracy for each passing year. These trends are significant⁷⁵, but not significantly different from the trend of non-students. Using a dummy-specification gives similar results (see the Appendix, section A.2).

This analysis suggests that the observed cognitive gap between university students and others is mainly due to selection; the immediate gains from university studies seem to be rather small (to the degree they exist at all). At 'year 1' there is a large gap between the groups: Those who are enrolled in a university program score roughly half of a standard deviation above those who are not (see table 7). For each year thereafter, this gap increases modestly by 0.02-0.03 standard deviations depending on outcome, but this increase is insignificant or marginally significant (for women).

⁷³ Includes any upper secondary program (academic or vocational) except for ISCED 3C short (shorter than 2 years).

⁷⁴ Some individuals continue onto other educational programs; these are not included here. (This is because the length of these programs are not clear.) Also, some individuals continue studying but later drop out. These are also not included. Naturally, I cannot exclude individuals who will drop out in the future. Hence, these are part of the student sample.

⁷⁵ For each additional year since graduation, the university students add an estimated 0.033 standard deviations to their literacy score ($p = 0.046$) and 0.031 standard deviations to their numeracy score ($p = 0.021$). These are the trends from the main specification presented in table 7. When I estimate the corresponding model using university students only, these estimates increase to 0.051 ($p = 0.001$) and 0.052 ($p = 0.001$) for literacy and numeracy, respectively.

It's worth noting that the error margins are rather wide. Hence, these regressions do not exclude the possibility of modest gains from university studies (up to ~8 percent of a standard deviation per year on literacy). In order to increase the statistical power, I experiment with less stringent sample selection criteria (see the Appendix, section A.3). These regressions strengthen the case for the effect being small or non-existent. Example: In one sample I include all students enrolled in post-secondary and tertiary programs, and observe them 0-4 years after graduation. Here, the gains from higher education are still 0.030 standard deviations per year for literacy and 0.020 for numeracy, but now more precisely measured (with standard errors at 0.018 and 0.017, respectively). In the Appendix (section A.3) I also compare the trend of students to that of workers, but the conclusions remain the same. Similarly, I find no significant differences in trends depending on educational area (see section A.3, table A3).

Table 7. Trends in literacy and numeracy after graduating from upper-secondary school

Panel A: Full sample	Literacy (z-score)	Numeracy (z-score)
Enrolled in a University program	0.46*** (0.068)	0.48*** (0.067)
Years later (1-4)	0.0025 (0.019)	0.012 (0.024)
Years later*Enrolled	0.031 (0.025)	0.020 (0.024)
Observations	6,316	6,316
R-squared (overall)	0.319	0.318
Panel B: By gender	Literacy Women	Literacy Men
Enrolled in a University program	0.35*** (0.091)	0.56*** (0.10)
Years later (1-4)	-0.0081 (0.028)	0.0076 (0.026)
Years later*Enrolled	0.060* (0.033)	0.0033 (0.038)
Observations	3,183	3,133
R-squared (overall)	0.313	0.342

Notes: All regressions include controls for country of residence, gender, immigration status, age at graduation, upper-secondary degree (converted into years of schooling) and parental education levels. All regressions are weighted as to account for the country specific survey designs, on a cross-country level all countries are weighted equally in all regressions. Standard errors (in parenthesis) are estimated using jackknife replicate sampling weights and adjusted for the imputation variance. *** p<0.01, ** p<0.05, * p<0.1

There are a couple of caveats worth mentioning at this point. First, note that these estimates are based on cross-sectional data. Example: The individuals observed in 'year 1' is an independent sample from those observed in year 2, 3 or 4. For university students, this means that 'year 1' also includes those who will graduate or drop out before 'year 4'; 'year 4' only includes those who did not. For this reason, I would expect the university-trend to be overly optimistic (although the opposite is also possible if 'year 4'-students are overrepresented by individuals who did not graduate on time). In a similar fashion, one could expect the control-trend to be overly pessimistic; some of the individuals in 'year 1' may enter a university program later on, leaving only those who never did so in 'year 4'.

I do find some evidence of such sorting: The probability of having highly educated parents is marginally higher later on (in the student sample) and significantly lower later on (in the control group).⁷⁶ This would suggest that the gap in trends between the groups is overestimated, if anything. This does not change the main conclusion, however.

There is also reason to suspect that the estimated university-trend is an understatement of the true trend, since I only observe the number of years since graduation, and not 'years as enrolled'. Now, students who graduated three years ago are likely to have more university education compared to those who graduated two years ago, but probably not a full year more. However, this measurement error needs to be rather large in order for the university-trend to change in important ways.⁷⁷

⁷⁶Among the students: 53.4 percent have at least one highly educated parent in 'year 1' which grows to 55.7 percent by 'year 4' ($p = 0.331$). Similarly, there is no significant difference in schooling, age at graduation, the probability of being female or immigrant depending on 'years since graduation'. For the control group: 25.8 percent have at least one highly educated parent in 'year 1'. By 'year 4' this number has dropped significantly to 19.7 percent ($p = 0.023$). There are no other significant differences in background characteristics depending on 'years since graduation' in this group. In these regressions, 'years since graduation' is included linearly. I have not included any controls except for country of residence.

⁷⁷ As a thought experiment: Assume that 25 % of students go directly from upper secondary school to higher education; 35 % wait one year; 20 % wait two years, 10 % wait three years and 5 percent wait 4 years (the rest wait five years or longer). In this scenario, one additional year since graduation would correspond to 0.76 additional years of education.

3.7 DISCUSSION

In this chapter, I measure the effect of schooling on cognitive abilities by contrasting three identification strategies: the instrumental variables approach, a country-level analysis and a dose-response analysis. The instrumental variables approach suggests that the effect of schooling is rather large: With each additional year, numeracy is expected to increase by 27 percent of a standard deviation using the main specification (95 % confidence interval: 19-35 percent). This prediction fits rather well with the results from previous studies, but is hard to reconcile with my other two findings. If one year of schooling increases numeracy by 0.27 standard deviations, then why isn't performance growing more rapidly as individuals age? For example, Finns add 3.8 years of schooling to their education between the ages of 20 and 25.⁷⁸ This predicts an increase in numeracy by roughly one standard deviation, assuming no change in the absence of schooling. The real increase, however, is much smaller at 0.36 standard deviations. A similar pattern can be observed for any other country in this data set. Also, the fact that experienced university students perform only somewhat above those recently enrolled, suggests that the cognitive return to education is rather small at this level.

Assuming that all three identification strategies are valid, what does this say about the effect of schooling on cognitive abilities? There are a couple of possible interpretations. Most importantly, the instrumental variables approach measures the outcomes years, and even decades, after graduation, whereas the country-level and dose-response analyses measure the immediate effects. Indeed, the instrumental variables approach suggests that the cognitive return to education peaks at age ~50. Secondly, the identification strategies use variation in schooling at partly different levels: while the dose-response analysis looks at university students specifically, the instrumental variables approach uses any variation in schooling influenced by the instrument (which could include both secondary and tertiary levels). Thirdly, the instrumental variables technique estimates a 'local average treatment effect'. It is possible that this effect differs markedly from the average treatment effect. Indeed, the IV-estimate is roughly twice as big as the corresponding ordinary regression estimate, which, in turn, is expected to be positively biased if anything. This is a common finding in the 'returns to education'-literature, where IV-estimates tend to overshoot the corresponding OLS-estimates, suggesting that the 'marginal student' has more to

⁷⁸ I calculate this number by assuming that you stay in school until you achieve your highest degree or drop out, after which you never reenter the educational system again. For students, schooling equals your age; for non-students, schooling equals your age at graduation or dropping out. Hence, the estimate is probably overly liberal, but even so, my point remains.

gain from additional schooling than the average student.⁷⁹ Furthermore, the dose-response analysis is not either measuring an average treatment effect, but rather, the average treatment effect on the treated. It is possible that this effect is comparatively small. Indeed, recently enrolled university students perform roughly half of a standard deviation above others, where additional gains might be harder to come by. Lastly, the treatment – schooling – may differ depending on identification strategy. Note that the country-level analysis exploits differences in institutions or norms between countries, where some countries observe a larger portion of their youth continuing onto higher education. Now, countries with a large uptake into colleges and universities could potentially also contain a higher degree of tracking due to the diversity in abilities among the student population. This ‘tracking-effect’ would be part of the estimate in the country-level-analysis, but presumably not in the instrumental variables approach.

Naturally, there is also the possibility of bias. One could suspect that the instrumental variables approach overestimates the effect of schooling, if it also captures the effect of educational quality. For example, individuals born into relatively small cohorts would probably have experienced higher teacher-to-student ratios growing up. On the other hand, the country-level and dose-response analyses are sensitive to critique regarding unobserved cohort-effects and measurement errors, the compounded effect of which is uncertain.

⁷⁹ See the discussion by Carneiro et al. (2011).

3.8 CONCLUSIONS

In order to measure the effect of schooling on cognitive abilities, I use three alternative identification strategies: the instrumental variables approach, a country-level-analysis and a dose-response-analysis. These strategies have different strengths and weaknesses. The IV-approach is open to critique regarding instrument validity; the country-level- and dose-response-analyses, on the other hand, are open to critique regarding confounding cohort-effects and measurement errors in schooling. With these caveats in mind, I draw the following two main conclusions:

1) Post-primary schooling has positive and significant effects on cognitive abilities later in life. This is supported by the fact that individuals born into relatively large cohorts have significantly less education later on and perform significantly worse on future literacy and numeracy tests. The instrumental variables approach suggests that one additional year of schooling increases numeracy by 0.27 standard deviations among 28-60 year olds. This effect is robust to a wide range of specifications, with effect sizes varying between 0.23 and 0.30 standard deviations. Naturally, this effect may not be representative for all groups of individuals. For example, those who choose to study at this level could experience smaller gains from education than the marginal student targeted by the instrument.

This finding is in line with the literature at large which usually find clear effects of schooling on cognitive abilities, although the effects on late life cognition vary substantially between studies. This finding is more difficult to reconcile with the ordinary regression estimates based on these data, which are substantially smaller. Likewise, the country-level-analysis also suggests a clearly smaller *immediate* schooling-effect, amounting to 0.05 standard deviations among 16-30 year olds. One plausible explanation is heterogeneity in the underlying effect, although I cannot exclude the possibility of bias.

2) The *immediate* effects of university studies are likely to be rather modest, at least for individuals choosing to study at this level. University students experience a yearly growth in numeracy by 0.03-0.04 standard deviations. This effect is significant, but not significantly different from the trend of non-students.

References

- Angrist, J. D., & Lavy, V. (1999). Using Maimonides' rule to estimate the effect of class size on scholastic achievement. *The Quarterly Journal of Economics*, 114(2), 533-575.
- Banks, J., & Mazzonna, F. (2012). The effect of education on old age cognitive abilities: evidence from a regression discontinuity design. *The Economic Journal*, 122(560), 418-448.
- Belzil, C. (2007). The return to schooling in structural dynamic models: a survey. *European Economic Review*, 51(5), 1059-1105.
- Brinch, C. N., & Galloway, T. A. (2012). Schooling in adolescence raises IQ scores. *Proceedings of the National Academy of Sciences*, 109(2), 425-430.
- Cahan, S., & Cohen, N. (1989). Age versus schooling effects on intelligence development. *Child Development*, 1239-1249.
- Card, D. (1999). The causal effect of education on earnings. *Handbook of Labor Economics*, 3, 1801-1863.
- Carlsson, M., Dahl, G. B., Öckert, B., & Rooth, D. O. (2015). The effect of schooling on cognitive skills. *Review of Economics and Statistics*, 97(3), 533-547.
- Carneiro, P., Heckman, J. J., & Vytlačil, E. J. (2011). Estimating marginal returns to education. *The American Economic Review*, 101(6), 2754-2781.
- Cascio, E. U., & Lewis, E. G. (2006). Schooling and the armed forces qualifying test evidence from school-entry laws. *Journal of Human Resources*, 41(2), 294-318.
- Cooper, H., Nye, B., Charlton, K., Lindsay, J., & Greathouse, S. (1996). The effects of summer vacation on achievement test scores: A narrative and meta-analytic review. *Review of Educational Research*, 66(3), 227-268.
- Dieterle, S. G., & Snell, A. (2016). A simple diagnostic to investigate instrument validity and heterogeneous effects when using a single instrument. *Labour Economics*, 42, 76-86.
- Doblhammer, G., van den Berg, G. J., & Fritze, T. (2013). Economic conditions at the time of birth and cognitive abilities late in life: evidence from ten European countries. *PLoS One*, 8(9), e74915.
- Falch, T., & Sandgren Massih, S. (2011). The effect of education on cognitive ability. *Economic Inquiry*, 49(3), 838-856.
- Fitzpatrick, M. D., Grissmer, D., & Hastedt, S. (2011). What a difference a day makes: Estimating daily learning gains during kindergarten and first grade using a natural experiment. *Economics of Education Review*, 30(2), 269-279.
- Fredriksson, P., Öckert, B., & Oosterbeek, H. (2012). Long-term effects of class size. *The Quarterly Journal of Economics*, 128(1), 249-285.
- Glymour, M. M., Kawachi, I., Jencks, C. S., & Berkman, L. F. (2008). Does childhood schooling affect old age memory or mental status? Using state schooling laws as natural experiments. *Journal of Epidemiology & Community Health*, 62(6), 532-537.
- Goldstein, J., Kreyenfeld, M., Jasilioniene, A., & Örsal, D. D. K. (2013). Fertility reactions to the "Great Recession" in Europe: Recent evidence from order-specific data. *Demographic Research*, 29, 85-104.

- Hansen, K. T., Heckman, J. J., & Mullen, K. J. (2004). The effect of schooling and ability on achievement test scores. *Journal of Econometrics*, 121(1), 39-98.
- Hanushek, E. A., Schwerdt, G., Wiederhold, S., & Woessmann, L. (2015). Returns to skills around the world: Evidence from PIAAC. *European Economic Review*, 73, 103-130.
- Harmon, C., Hogan, V., & Walker, I. (2003). Dispersion in the economic return to schooling. *Labour Economics*, 10(2), 205-214.
- Hofmann, B., & Hohmeyer, K. (2013). Perceived economic uncertainty and fertility: Evidence from a labor market reform. *Journal of Marriage and Family*, 75(2), 503-521.
- Imbens, G., & Angrist, J. (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica*, 62(2), 467-475.
- Jensen, A. (1969). How much can we boost IQ and scholastic achievement. *Harvard Educational Review*, 39(1), 1-123.
- Leuven, E., Lindahl, M., Oosterbeek, H., & Webbink, D. (2004). New evidence on the effect of time in school on early achievement.
- Meghir, C., Palme, M., & Simeonova, E. (2013). *Education, cognition and health: Evidence from a social experiment* (No. w19002). National Bureau of Economic Research.
- Mosteller, F. (1995). The Tennessee study of class size in the early school grades. *The Future of Children*, 113-127.
- Neisser, U., Boodoo, G., Bouchard Jr, T. J., Boykin, A. W., Brody, N., Ceci, S. J., Halpern, D. F., Loehlin, J. C., Perloff, R., Sternberg, R. J., & Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, 51(2), 77.
- Oreopoulos, P., & Salvanes, K. G. (2009). *How large are returns to schooling? Hint: Money isn't everything* (No. w15339). National Bureau of Economic Research.
- Ritchie, S. J., Bates, T. C., Der, G., Starr, J. M., & Deary, I. J. (2013). Education is associated with higher later life IQ scores, but not with faster cognitive processing speed. *Psychology and Aging*, 28(2), 515.
- Schneeweis, N., Skirbekk, V., & Winter-Ebmer, R. (2014). Does education improve cognitive performance four decades after school completion?. *Demography*, 51(2), 619-643.
- Sobotka, T., Skirbekk, V., & Philipov, D. (2011). Economic recession and fertility in the developed world. *Population and Development Review*, 37(2), 267-306.
- Stelzl, I., Merz, F., Ehlers, T., & Remer, H. (1995). The effect of schooling on the development of fluid and cristallized intelligence: A quasi-experimental study. *Intelligence*, 21(3), 279-296.
- Stock, J. H., & Yogo, M. (2002). Testing for weak instruments in linear IV regression.
- Weiss, A. (1995). Human capital vs. signalling explanations of wages. *The Journal of Economic Perspectives*, 9(4), 133-154.
- Winship, C., & Korenman, S. (1997). Does staying in school make you smarter? The effect of education on IQ in The Bell Curve. In *Intelligence, Genes, and Success* (pp. 215-234). Springer New York.

Appendix

A.1 Heterogeneity

Table A1. Heterogeneous effects of schooling on numeracy (IV-estimates)

<i>By age</i>	Numeracy
Schooling	-0.083 (0.46)
Schooling x (Age-28)	0.035 (0.021)
Schooling x (Age-28) ²	-0.00086** (0.00034)
Chi2 (age-interactions)	7.02**
Chi2 (schooling, age-interactions)	20.74***
Observations	76,709
<i>By parental education</i>	Numeracy
Schooling	0.28*** (0.042)
(ref. Parents have primary education)	
Schooling x Parent has a secondary education	0.00057 (0.069)
Schooling x Parent has a tertiary education	-0.053 (0.51)
Chi2 (parental-interactions)	0.01
Chi2 (schooling, parental-interactions)	101.29***
Observations	76,709

Notes: Your parents have 'primary education' if both have a lower secondary degree or less; your parents have 'secondary education' if at least one has an upper-secondary degree and none has a tertiary degree; your parents have 'tertiary education' if at least one has a tertiary degree. I also include individuals with missing values on parental education levels: If both are missing, your parent have primary education; if one is missing, the non-missing parent determines your group. All regressions control for gender, age (age, age²), parental education levels and immigration background, as well as the effects of living in country c and being born in year t, as modeled by Z. All regressions are weighted as to account for the country specific survey designs; on a cross-country level, all nations are weighted equally in all regressions. Standard errors (in parenthesis) are estimated using jackknife replicate sampling weights and further adjusted for the imputation variance. *** p<0.01, ** p<0.05, * p<0.1

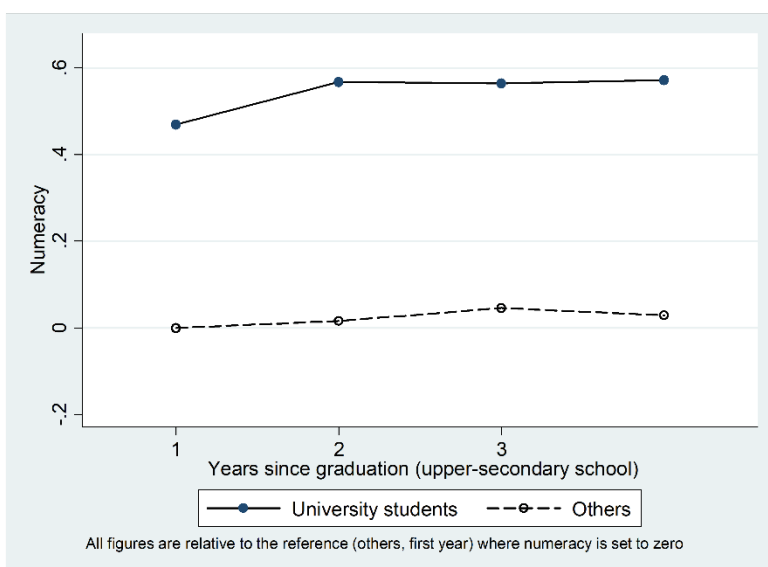


Figure A2. The trend in numeracy among university students and others

A.3 Alternative sample selection criteria

Here I compare the trend in literacy and numeracy of students to the trend of non-students, using alternative samples. In the original sample, I include individuals who graduated from upper-secondary school one to four years ago; the students are all enrolled at universities and the control group include all others (69 percent are working, 15 percent are unemployed and 16 percent are not in the labor force). Here I extend (limit) the sample in several ways as to gain statistical power (see table A2). I make the following adjustments to the sample selection criteria:

- 1) 1-3 years: Here I exclude anyone in their fourth year after graduation, as these students might be especially vulnerable to selection (these students are, presumably, overrepresented by master students and students who did not earn a bachelor on time).
- 2) 0-4 years: Here I include individuals who graduated from upper-secondary school 0-4 years ago. Including 'year 0' may be problematic: These individuals have recently graduated and it seems likely that some of those in the 'control group' actually belong in the 'treatment group', i.e. they will soon enroll. However, the estimates remain mainly unchanged but the standard errors are further reduced.

3) 0-4 years, All students: Here I further broaden the 'treatment group' to all students who are enrolled in a post-secondary or tertiary program (not only university students). This leaves the estimates mainly unchanged.

4) 0-4 years, All students vs Workers: Here I further restrict the 'control group' to workers with a 'reasonable' amount of work experience. Example: If you graduated two years ago then you need at least two years of work experience in order to be included. This closes the gap in trends between the student and the control group.

Table A2. Trends in literacy and numeracy after graduating from upper-secondary school, alternative samples

<i>LITERACY</i>	1-3 years	0-4 years	0-4 years All stud.	0-4 years All stud. vs Work
Enrolled	0.45*** (0.081)	0.44*** (0.045)	0.16*** (0.050)	0.16*** (0.055)
Enrolled*Uni	-	-	0.27*** (0.036)	0.29*** (0.038)
Years later	0.0075 (0.029)	-0.0039 (0.015)	-0.0056 (0.015)	0.021 (0.020)
Years later*Enrolled	0.033 (0.039)	0.032* (0.019)	0.030 (0.018)	0.0029 (0.023)
Observations	5,091	7,644	9,374	7,703
R-squared (overall)	0.311	0.309	0.294	0.265
<i>NUMERACY</i>	1-3 years	0-4 years	0-4 years All stud.	0-4 years All stud. vs Work
Enrolled	0.46*** (0.080)	0.48*** (0.043)	0.19*** (0.050)	0.19*** (0.059)
Enrolled*Uni	-	-	0.26*** (0.037)	0.27*** (0.037)
Years later	0.021 (0.032)	0.0095 (0.014)	0.0076 (0.014)	0.036* (0.019)
Years later*Enrolled	0.024 (0.036)	0.018 (0.018)	0.020 (0.017)	-0.0086 (0.022)
Observations	5,091	7,644	9,374	7,703
R-squared (overall)	0.307	0.310	0.292	0.261

Notes: 1-3 years: Includes individuals who graduated from upper-secondary school 1-3 years ago; 1-4 years: Includes individuals who graduated from upper-secondary school 1-4 years ago; 0-4 years: Includes individuals who graduated from upper-secondary school 0-4 years ago; 0-4 years, All stud.: Here I further broaden the treatment group to all students at a post-secondary or tertiary level; 0-4 years, All stud. vs Work: Here I further restrict the control group to workers with a 'reasonable' amount of work experience. All regressions include controls for country of residence, gender, immigration status, age at graduation, upper-secondary degree (converted into years of schooling) and parental education levels. All regressions are weighted as to account for the country specific survey designs, on a cross-country level all countries are weighted equally in all regressions. Standard errors (in parenthesis) are estimated using jackknife replicate sampling weights and adjusted for the imputation variance. *** p<0.01, ** p<0.05, * p<0.1

Table A3. Trends in literacy and numeracy after graduating from upper-secondary school, by educational area

	Literacy (z-score)	Numeracy (z-score)
Enrolled (<i>ref. Social sciences, business, law</i>)	0.45*** (0.090)	0.43*** (0.10)
Enrolled*HumEduc	-0.13 (0.11)	-0.18 (0.12)
Enrolled*ScienceMath	0.36*** (0.10)	0.41*** (0.13)
Enrolled*Engineering	-0.017 (0.13)	0.13 (0.13)
Enrolled*Else	-0.12 (0.11)	-0.039 (0.13)
Years later	0.0020 (0.019)	0.012 (0.019)
Years later*Enrolled (<i>ref. Social sciences, business, law</i>)	0.039 (0.033)	0.029 (0.037)
Years later*Enrolled*HumEduc	0.014 (0.039)	0.023 (0.044)
Years later*Enrolled*ScienceMath	-0.082* (0.042)	-0.075 (0.049)
Years later*Enrolled*Engineering	-0.011 (0.049)	0.0016 (0.047)
Years later*Enrolled*Else	0.037 (0.042)	0.015 (0.049)
Observations	6,316	6,316
R-squared (overall)	0.325	0.329

Notes: HumEduc is an indicator for those studying humanities, languages and art; ScienceMath includes science, mathematics and computing; Engineering includes engineering, manufacturing and construction; Else includes, for example, teacher training, agriculture and health. All regressions include controls for country of residence, gender, immigration status, age at graduation, upper-secondary degree (converted into years of schooling) and parental education levels. All regressions are weighted as to account for the country specific survey designs, on a cross-country level all countries are weighted equally in all regressions. Standard errors (in parenthesis) are estimated using jackknife replicate sampling weights and adjusted for the imputation variance. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

A.4 Bias reduction

Equation (1) states that:

$$bias = \delta\gamma = \frac{Var(ability|age, country)}{Var(schooling|age, country)} \cdot \gamma \quad (1)$$

Here I show that:

$$\begin{aligned} \delta &= \frac{Cov(ability, schooling|age, country)}{Var(schooling|age, country)} \\ &= \frac{Var(ability|age, country)}{Var(schooling|age, country)} \end{aligned}$$

Using the schooling-equation, we can re-write the numerator as:

$$\begin{aligned} &Cov(ability, schooling|age, country) \\ &= Cov(ability, \alpha_g + ability + \varepsilon|age, country) \\ &= Var(ability|age, country) \end{aligned}$$

using the assumption that α_g does not correlate with ability, conditional on age and country: $Cov(ability, \alpha_g|age, country) = 0$.

Equation (2) states that:

$$\delta'\gamma = \frac{\sigma_v^2 \cdot \gamma}{m \cdot \sigma_{g|age, country}^2 + \sigma_v^2 + \sigma_\varepsilon^2} \quad (2)$$

Here I show that:

$$\frac{Var(\overline{ability}|age, country)}{Var(\overline{schooling}|age, country)} = \frac{\sigma_v^2}{m \cdot \sigma_{g|age, country}^2 + \sigma_v^2 + \sigma_\varepsilon^2}$$

Writing ability as:

$$ability = \mu_c + v$$

where μ_c are country-specific constants and v are individual IID deviations: $E(v) = 0$ and $Var(v) = \sigma_v^2$. The average ability on the group-level is described by:

$$\overline{ability}_g = \mu_c + \bar{v}_g$$

and the conditional variance is given by:

$$\begin{aligned} \text{Var}(\overline{ability}_g | age, count) &= \underbrace{\text{Var}(\mu_c | age, count)}_{=0} + \text{Var}(\bar{v}_g | age, count) \\ &= \frac{\sigma_v^2}{m} \end{aligned}$$

where m is the sample size on the group level, assumed the same for all groups. For the denominator in (2) we have:

$$\overline{schooling}_g = \alpha_g + \overline{ability}_g + \bar{\varepsilon}_g$$

$$\text{Var}(\overline{schooling}_g | age, country) = (m \cdot \sigma_{g|age, country}^2 + \sigma_v^2 + \sigma_\varepsilon^2) / m$$

CHAPTER 4

Some Evidence for a Cognitive Decline from Leaving School in a Recession

Abstract

Several recent studies have shown that graduating in a bad economy has large and persistent adverse labor market effects. In this study, I investigate whether graduating in a bad economy also affects your cognitive skills. I use PIAAC survey data covering roughly 48,000 individuals, mostly from European OECD-countries. I exploit three complementary identification strategies. Firstly, I measure the effect of the unemployment rate at school leaving, controlling for observable confounders. Secondly, I instrument the unemployment rate using the unemployment rate in the expected time of graduation. Thirdly, I measure the effect of the unemployment rate after leaving school, conditional on the unemployment rate in the year of leaving. Overall, there is some evidence for a small cognitive decline. The main estimate (IV) suggests that a one-percentage point increase in the unemployment rate at school leaving causes a 0.006 standard deviation drop in literacy ($p < 0.1$) with an even smaller effect on numeracy. Here, the unemployment rate is measured as a three-year average starting from the year of leaving school. Outcomes are measured 14 years after leaving school on average, for individuals at an average age of 35.

4.1. INTRODUCTION

Many young adults struggle to find their first job.⁸⁰ As reported in several studies, these struggles are exacerbated for those who graduate in a slow economy. Emerging evidence suggests that other areas of your life may also be affected, including your physical and mental wellbeing, self-esteem, sleep quality and family life (see Cutler et al., 2015; Maclean, 2013; Maclean, 2015; Maclean & Hill, 2015; Maclean, 2016, Maclean et al., 2016; Maclean & Hill, 2017). Here, I investigate whether your cognitive skills are also affected. To this end, I use survey data from The Programme for the International Assessment of Adult Competencies (PIAAC). This is a pooled cross-section covering tens of thousands of individuals from mostly European OECD-countries. The variables of central interest are literacy and numeracy, measuring your text-based and mathematical problem-solving skills (see chapter 1). The identification is three-fold. First, I compare individuals who left school when the unemployment rate was high to those who left school when it was low, controlling for differences in observable confounders. Below, this strategy is called the ‘standard regression’. Secondly, I instrument the school leaving unemployment rate using the unemployment rate in the expected time of graduation. Thirdly, I measure the effect of the unemployment rate in the first year(s) following school leaving, conditional on the unemployment rate at school leaving.

I find some evidence for the existence of a small negative effect. The ‘standard regression’ and the instrumental variables approach give similar results: A one-percentage point increase in the unemployment rate at school leaving predicts a 0.005-0.006 standard deviation drop in literacy; the effect on numeracy is even smaller. Here, the unemployment rate is measured as a three-year average starting from the year of leaving school. Outcomes are measured 14 years after leaving school on average, for individuals at an average age of 35. Hence, a deep recession – increasing this unemployment rate by four percentage points – predicts a future loss in literacy by 0.02 standard deviations or roughly 0.6 percentile points. The third identification strategy provides additional support for a cognitive decline: High unemployment rates in the year(s) following school leaving predict significantly lower performances on literacy and numeracy later on, conditional on the unemployment rate at school leaving.

It is worth noting that this is a modest effect indeed. The state of the labor market at graduation is unlikely to be an economically significant contributor to the variation in cognitive functioning among cohorts later in life. On the other hand, graduating in a bad economy is a gentle stressor in comparison

⁸⁰ Based on NLSY79: The median school leaver takes 4.6 years before finding stable employment (a relationship lasting at least three years). (Yates, 2005)

to being unable to find a job. Hence, this study does not exclude the possibility that some groups of the labor force are cognitively severely affected.

4.2 BACKGROUND AND PREVIOUS RESEARCH

Several studies have shown that graduating in a recession has large and persistent adverse labor market effects (Brunner & Kuhn, 2014; Cutler et al., 2015; Genda et al., 2010; Hershbein, 2012; Kahn, 2010; Kondo, 2015; Kwon et al., 2010; Liu et al., 2016; Oreopoulos et al., 2012; Oyer, 2006 and Raaum & Røed, 2006). For example, Liu et al. use Norwegian data and find that a one-percentage point increase in the unemployment rate at graduation increases the short-term risk of unemployment by 1.4 percentage points. Such labor market effects are typically larger early on, and last a couple of years or decades, depending on study and demographic.⁸¹

As suggested by several of the authors, these effects may be driven by cyclical skill mismatch, which seem to be difficult to swiftly recover from. One potential reason for this is that your job – or lack thereof – also influences your human capital formation. Indeed, this assumption underlies several models of persistent unemployment (see, for example, Pissarides, 1992, and Acemoglu, 1995). For example, Acemoglu argues that employers use unemployment-duration as a signal for productivity, further alienating segments of the labor force. In his model, unemployment itself becomes a cause of unemployment, fueled by human capital deterioration. Whether your job position actually affects your skill formation, and to what degree, is therefore central for the understanding of career developments.

A priori, there are both pecuniary and non-pecuniary motivations for a relationship between your work situation and cognitive performance. Grossman (1972) introduced the idea of health as an endogenous variable, formed by both market and non-market inputs. This model presents a couple of potential mechanisms, if we allow for the concept of health to also include cognitive health. First, there is the ‘investment-argument’: Some jobs have higher returns on cognitive investments. Hence, we might expect individuals with cognitively challenging jobs to make more of these investments. Then there is the ‘consumption-argument’, where cognitive health is viewed as a normal good comparable to a gym-membership or a healthy diet. On the other hand, well-paid jobs may also increase the opportunity cost of certain health-promoting activities, such as sleeping.

One can also make a case for a stress-related pathway, i.e. that the stresses of a poor labor market position affects your cognitive functioning. There is robust empirical evidence of a relationship between unemployment and several other adverse outcomes, including general dissatisfaction, poor mental health and depression (Clark et al., 2001; Dooley et al., 1994; Kassenboehmer & Haisken-DeNew, 2009; Krueger & Mueller, 2011; Kuhn et al.,

⁸¹ Brunner & Kuhn (2014) is an exception in this regard, finding larger effects later on.

2009; Winkelmann & Winkelmann, 1998)⁸², social isolation and marital disruption (Brand & Burgard, 2008; Eliason, 2012; Hansen, 2005) as well as declining physical health and mortality (Beale & Nethercott, 1985; Browning & Heinesen, 2012; Eliason & Storrie, 2009; Sullivan & von Wachter, 2009).⁸³ Among the employed, a couple of job characteristics – including type of contract and earnings – have been linked to mental and physical well-being.⁸⁴

Of special interest for this study is a couple of recent findings showing that graduating in an economic downturn predicts several adverse health-related outcomes, including lowered life satisfaction, greater obesity and more smoking and drinking (Cutler et al., 2015). Among men, worse physical functioning (Maclean, 2013), more heavy drinking (Maclean, 2015), lowered chance of being married and having children (Maclean et al., 2016), lowered body weight (Maclean, 2016) and lowered self-esteem (Maclean & Hill, 2015) have also been reported. These patterns are often different for women, who generally seem to fair better or even benefit from graduating in a recession (see the work by Maclean and others).

Based on this body of research, it is not very controversial to suggest that a poor labor market position can be stressful. The emerging view in the psychological field is that high or chronic stress impairs cognitive functioning as measured by explicit memory (Sandi, 2013).⁸⁵ There is also several epidemiological studies showing that poor health – as captured by metabolic indicators – is successful in predicting cognitive decline in both younger and older adults (see the review by Yates et al., 2012). Furthermore, some health-related behaviors, including exercise and quality sleep, have been shown to improve cognitive functioning (Alhola & Polo-Kantola, 2007; Loprinzi &

⁸² See also the reviews by Björklund & Eriksson (1998) and Paul & Moser (2009).

⁸³ There are also studies casting doubt on the importance of unemployment for mental and physical health. Böckerman & Ilmakunnas (2009) find no effect on self-reported health connected to the event of becoming unemployed using Finnish panel data. Similar results are reported for Germany (Schmitz, 2011) and United States (Salm, 2009) when looking at various health measures, and exploiting longitudinal data and plant closures. Similarly, Browning et al. (2006) find no effect of job displacement on stress-related hospitalizations.

⁸⁴ Employees on temporary contracts and part-time workers experience poor physical and mental health as compared to standard workers (for physical health, see Kim et al., 2008; Pirani & Salvini, 2015 and Virtanen et al., 2005; for mental health, see Han et al., 2017; Ferrie et al., 2002; Kim et al., 2006; Quesnel-Vallée et al., 2010). There is also some evidence that higher earnings have mental health benefits. Changes in incomes correlate positively with changes in self-reported health (see the review by Gunasekara, 2011; Meraya et al., 2018). Studies that exploit lottery winnings, inheritances or the Social Security Notch as exogenous sources of variation in income have also found beneficial effects on mental health (Apouey & Clark, 2015; Au & Johnston, 2014; Golberstein, 2015; Lindqvist et al., 2018); the physical health effects are debatable, however (see Apouey & Clark, 2015; Au & Johnston, 2014; Au & Johnston, 2015; Kim & Ruhm, 2012).

⁸⁵ Mild and transitory stressors, on the other hand, can have beneficial cognitive effects.

Kane, 2015; Hwang et al., 2017; Winter et al., 2007). To sum up: the poor labor market outcomes resulting from graduating in a bad economy may lead to long-term stress, declining mental and physical health, all of which may contribute to cognitive impairment.

To date, the empirical evidence on labor market-related skill formation is sparse, however. To my knowledge, there is only one study looking at this relationship directly: Edin & Gustavsson (2008) use Swedish panel data and find that a full year of non-employment decreases literacy by five percentile points. Also related is a study by Mani et al. (2013) showing that periodical poverty causes losses in IQ by exploiting the cyclical variation in incomes among Indian sugarcane farmers. I contribute to this emerging literature on labor market-related skill formation by studying whether leaving school in a bad economy affects your future cognitive functioning as measured by literacy and numeracy. To my knowledge, this is the first study of its kind. This study does not, however, discriminate between pecuniary and non-pecuniary pathways.

4.3. EMPIRICAL STRATEGIES

The aim is to measure the cognitive effect of leaving school in a bad economy. The empirical challenge is to disentangle the causal effect from other unobservable confounders. Naturally, graduating dates are not randomly assigned. Hence, there might be relevant differences between individuals who graduate while the unemployment rate is high, and those who graduate while it is low. Most importantly, individuals might choose this date strategically, postponing or preponing graduation depending on their labor market prospects. Ex ante, I expect high-achievers to be less affected by business cycles, as their employment opportunities might also be less affected (Cutler et al., 2015 and Hoynes et al., 2012, among others, show that recessions hit poorly educated individuals harder). On the other hand, high-achievers might have better adjustment opportunities, enabling them to be more flexible. Hence, any potential bias could go in either direction.

In order to capture the cognitive effect, I contrast three complementary identification strategies. First, I measure the effect of the unemployment rate at school leaving, controlling for observable confounders. Secondly, I instrument this unemployment rate using the unemployment rate in the *expected* time of graduation. Lastly, I measure the effect of the unemployment rate in the first year(s) following school leaving, conditional on the unemployment rate in the year of leaving school. The first two strategies have been employed in several studies (Brunner & Kuhn, 2014; Cutler et al., 2015; Kahn, 2010; Kondo, 2015; Oreopoulos et al., 2012 and the studies by Maclean and others, all exploit this or similar instruments).

4.3.1 Measuring the effect, adjusted for observable confounders

As an opening strategy, I measure the effect of the unemployment rate at school leaving – graduation or dropping out – while controlling for other variables that are expected to predict literacy and numeracy (*cog*):

$$cog_{ci} = \alpha_c + \beta unemp_{ci} + \vartheta X_{ci} + \varepsilon_{ci} \quad (1)$$

where c is a country index and i is indexing individuals; α_c represents country-specific fixed intercepts; $unemp$ is the unemployment rate at school leaving with β representing the parameter of central interest; X_{ci} is a vector of background characteristics including age (using age-dummies), gender, immigration status, years of schooling, as well as the mother's and father's educational levels; ε_{ci} is the random error term.

It is worth noting that I do not control for the unemployment rate in later years. Hence, a factor such as the average unemployment rate since leaving school is likely to be larger, on average, among those who left school in a

recession. I interpret these later labor market conditions as part of the stimuli of interest. The contemporaneous unemployment rate will have little to no influence on the estimates, however, as there is little within-country variation in interviewing dates.

This identification strategy is valid if there are no relevant differences in background characteristics between those who leave school while the economy is strong and those who leave it while it is weak (conditional on the other covariates). If students choose their school leaving date strategically, however, then there is reason to suspect that β may not fairly reflect the causal effect. I explore this issue further in section 6.

4.3.2 The instrumental variables technique

My main identification strategy relies on instrumenting the unemployment rate at school leaving, using the unemployment rate in the *expected* time of graduation.⁸⁶ I specify the *reduced form equation*⁸⁷ in the following way:

$$cog_{ci} = \alpha_c + \beta expunemp_{ci} + \vartheta X_{ci} + \varepsilon_{ci} \quad (2)$$

where *expunemp* is the instrument measuring the unemployment rate in the expected time of graduation; X_{ci} represents the same vector of background characteristics described in section 3.1. Similarly, α_c and ε_{ci} represent the country-specific fixed intercepts and error term for this equation.

The unemployment rate in the *expected* time of graduation is a valid instrument, assuming that β owns its value only due to it mediating the effect of the actual unemployment rate.⁸⁸ Put differently, individuals are ‘allowed’ to choose their school leaving date, but not their expected graduation date. Hence, students might choose to postpone graduation when they fail to find a job in a slow economy. If they, however, choose to enter a new educational program that they would not have entered otherwise, then there is reason to suspect that β may not fairly reflect the causal effect. I discuss this possibility further in section 6.

⁸⁶ The expected year of graduation is inferred from data based on your desired degree and country of residence, as described in section 4.2.

⁸⁷ Using two-stage-least-squares I estimate the ratio: β/β_{fs} , where β is the coefficient for the instrument (*expunemp*) from the reduced form equation (see (2)); β_{fs} is the coefficient for the instrument from the first stage equation. The first stage equation models the independent variable of interest – the unemployment rate at school leaving – as a function of the instrument and the same set of covariates (X) described in equation (2).

⁸⁸ Also, the unemployment rate in the expected time of graduation needs to be a sufficiently strong predictor of the actual unemployment rate.

4.3.3 The conditional effect of the unemployment rate in the year(s) following graduation

As a last strategy, I estimate the effect of the unemployment rate *in the first year(s) following school leaving*, conditional on the unemployment rate in the year of leaving school:

$$cog_{ci} = \alpha_c + \beta_1 postunemp_{ci} + \beta_2 unemp_{ci} + \vartheta X_{ci} + \varepsilon_{ci} \quad (3)$$

Example: Assume two individuals, A and B, who both graduated while the unemployment rate was 7.5 percent ($unemp = 7.5$). In the following year, A experiences an unemployment rate at eight percent ($postunemp = 8$) while B experiences an unemployment rate at seven percent ($postunemp = 7$). In this model, the parameter of interest is β_1 , capturing the effect of this labor market difference on literacy and numeracy (cog). I also control for the same set of covariates (X) described in section 3.1, as well as country fixed effects (α_c).

Now, assuming that individuals can react to the current unemployment rate, but do not adjust their school leaving date depending on future unemployment rates, then conditioning on the unemployment rate at school leaving is arguably making the future unemployment rate exogenous. The validity of this strategy is called into question if students have well-formed expectations regarding future labor market conditions and react to these strategically, or if they choose to re-enter the educational system as a response to their post-graduate labor market situation. I return to these possibilities in section 6.

Naturally, there would make little sense to estimate β_1 in this model, unless the conditional unemployment rate in the year(s) following school leaving also predicts your future career path. Previous research gives me reason to believe that this is, indeed, the case. Oreopoulos et al. (2012) find that a large part of the effect, i.e. the wage deficit from leaving school in a recession, can be attributed to the unemployment rate in the first year after leaving school. The current data also show that the conditional unemployment rate following graduation predicts your future employment prospects (as discussed further in section 5.3).

4.3.4 Weights and standard errors

There are some survey design issues to consider in the estimations. Firstly, the sample has been selected using a stratified cluster-design. Secondly, the testing procedure for literacy and numeracy are based on item response theory and multiple imputations using so called ‘plausible values’. I account for the survey design by using appropriate weights as assigned by OECD; on

a cross-country level, I weigh each country equally in each regression. I estimate the standard errors using jackknife replicate sampling weights and account for the imputation variance added by using plausible values.

4.4 DATA & SAMPLE SELECTION

I use PIAAC survey data that has been collected and compiled by OECD. The sample I exploit is a cross-sectional data covering roughly 48,000 individuals in 24 countries⁸⁹ and has been collected during 2010-2015 using personal house interviews. The outcome variables of central interest are *literacy* and *numeracy*, measuring your ability to solve text-based and mathematical problems. See chapter 1 for a description of these concepts.

4.4.1 Sample selection and descriptive statistics

The main identification relies on instrumenting the unemployment rate at school leaving using the unemployment rate in the expected time of graduation. In order to measure this effect, I select individuals who were expected to graduate in 1982 to 2008, varying somewhat depending on country and interviewing date.⁹⁰ Furthermore, I exclude individuals who are still enrolled in an educational program, who graduated or dropped out in the last year (as these could be in-between educational programs), those who have inconsistencies in their graduating date⁹¹, who are known to have immigrated to the current country of residence after leaving school, who earned their highest degree abroad, who were expected to graduate at the age of 13 or younger, as well as those who have missing values on key

⁸⁹ Belgium (Flanders), Chile, Cyprus, Czech Republic, Denmark, Finland, France, Greece, Ireland, Israel, Italy, Japan, Korea, Lithuania, the Netherlands, Norway, Poland, Russian Federation (excluding Moscow municipal area), Slovak Republic, Slovenia, Spain, Sweden, Turkey and United Kingdom (England & Northern Ireland).

⁹⁰ Making the selection based on the *actual* year of leaving school (as opposed to the expected year) could potentially bias the estimates: The probability of being included in the sample could then become a function of your cognitive abilities. Example: For most countries, I observe unemployment rates from 1980 going forward. Hence, I restrict the sample to those who were expected to graduate in 1982 or later, so that 'early leavers' are included from the start, i.e. those leaving school in 1980 or 1981. For some countries*, mostly post-Soviet states, unemployment rates are not observed until years later, and I adjust the sample selection criteria accordingly. Similarly, the typical individual took the survey in 2011. Hence, I include such an individual if she were expected to graduate in 2008 or earlier, so that 'late leavers' are included till the end. *This includes Cyprus (1982), Czech Republic (1995), Ireland (1985), Lithuania (1999), Poland (1990), Russia (1992), Slovak Republic (1993) and Slovenia (1992). Example: Slovenians are included if they were expected to graduate in 1994 or later. Data on unemployment rates from World Economic Forum.

⁹¹ This is identified by calculating the current year based on your graduating date and the number of years elapsed since graduation. If the proposed 'current year' falls outside the survey period (2010-2015) you are excluded.

variables.⁹² This also includes individuals who are observed leaving school unusually late, at the age of 35 or older: These individuals are likely to have spent years outside the educational system at some earlier date; hence, their ‘relevant’ school leaving date is likely not the one observed. Naturally, this cutoff is somewhat arbitrary, and I return to this issue in the robustness section (5.4). This leaves me with a sample consisting of 47,842 observations whose mean characteristics are described in table 1 below. Here I also divide the sample into two groups: those who graduated when the unemployment rate was high and those who graduated when it was low, where a ‘high’ unemployment rate is one at or above the country median for the relevant time period. The table includes the full set of covariates regularly included in the preceding regressions, excluding only country of residence.

Table 1 shows that the individual background characteristics are strikingly similar among those who graduated when the unemployment rate was low, and those who graduated when it was high. Neither is there any noticeable difference in literacy or numeracy that would favor those who graduated in a strong economy.

On average, the sample individuals score almost 0.2 standard deviations above average on literacy and numeracy, meaning that they score above the expected value for a randomly chosen individual from the PIAAC target population.⁹³ Furthermore, 48 percent of individuals are female; 4 percent are known to be immigrants; on average, the individuals are 35 years old with 13 years of schooling.⁹⁴ The age range is wide, however, stretching from 17 to 60 years, with outcomes measured 1 to 34 years after leaving school (average: 14 years).

⁹² Including age, gender, literacy/numeracy, schooling, the year of leaving school or the corresponding unemployment rate, and the expected year of leaving school. I allow for missing values on other covariates, and code these accordingly.

⁹³ The target population consists of everyone aged 16-65 from the relevant set of countries, i.e. the 31 countries included in the first two rounds of PIAAC (totally 197,754 individuals). The standard deviation for literacy and numeracy are estimated using within-country variation only: First, I estimate the (weighted) variance separately for each country. These variances are then averaged and the standard deviation of interest is the square root of that variance.

⁹⁴ The distribution for the highest (completed) degrees is as follows: Primary education (1 %), Lower secondary school (11 %), Upper secondary school (42 %), Post-secondary non-tertiary degree (4 %), Tertiary degree (42 %), Research degree (1 %).

Table 1. Weighted means for the individual characteristics

	All	Low unemp. ^a	High unemp. ^a
Literacy (z-score ^b)	0.18	0.18	0.18
Numeracy (z-score ^b)	0.19	0.18	0.19
Female	0.48	0.48	0.48
<i>Immigration status</i>			
Immigrant	0.035	0.036	0.034
Native	0.96	0.96	0.97
Unknown	0.0052	0.0053	0.0050
Age	35.2	35.3	35.0
Schooling (years)	13.2	13.2	13.2
<i>Educational field</i>			
Lower educ./General program	0.26	0.26	0.25
Teaching/Educational sciences	0.053	0.051	0.054
Hum./Languages/Art	0.064	0.064	0.063
Social sciences/Business/Law	0.16	0.17	0.16
Science/Math/Computing	0.075	0.075	0.076
Engineering/Manufact./Constr.	0.19	0.19	0.19
Agriculture/Veterinary	0.026	0.027	0.026
Health/Welfare	0.067	0.065	0.068
Services	0.075	0.072	0.077
Doesn't know	0.032	0.032	0.032
<i>Mother's education^c</i>			
Low	0.47	0.47	0.47
Medium	0.35	0.35	0.36
High	0.15	0.16	0.15
Unknown	0.023	0.022	0.024
<i>Father's education^c</i>			
Low	0.41	0.41	0.41
Medium	0.37	0.36	0.37
High	0.19	0.19	0.18
Unknown	0.034	0.034	0.034
Unemployment, % (year of leaving school)	8.4	6.3	10.2
Observations	47,842	22,432	25,410

Notes: The means are weighted as to account for the country specific survey designs; on a cross-country level, all countries are weighted equally according to the sample "All" (47,842 obs.). ^a'Low unemployment' means that the rate is below the country median for the studied period; 'high unemployment' means that the rate is at or above the country median. ^bz-scores are estimated using the full sample of 16-65 year olds (roughly 200,000 individuals) with standard deviations estimated using within-country variation only. ^cEducational degrees: Low = Lower secondary school, primary school or less. Medium = Upper secondary and post-secondary non-tertiary degree. High = Tertiary degree.

4.4.2 Determining the expected year of graduation

The main identification relies on instrumenting the unemployment rate at school leaving using the unemployment rate in the expected time of graduation. Your expected year of graduation is calculated based on the *median* age at graduation in an appropriate reference group. Your reference group is made up of individuals from your country who graduated with the same educational degree as yourself (for dropouts – the degree you were aiming for). The reference group excludes anyone who graduated at the age of 35 or older, as the target population consists of individuals who finish school ‘for the first time’. Your educational degree is based on the ISCED-scale, measured in roughly 10 categories⁹⁵ depending on country. In the robustness section (5.4) I also experiment with using other criteria for constructing an instrument.

The most common expected ages for graduating are 18 years (21 % of sample), 19 years (14 %), 23 years (14 %) and 25 years (11 %). Of the sample individuals, 30 percent left school on time; another 25 percent left school one or two years before expected and 22 percent left school one or two years later than expected. See the Appendix, section A.1, for the expected age at graduation depending on country and degree.

⁹⁵ These categories are: ISCED 1 (Primary education), ISCED 2 (Lower secondary school), ISCED 3C short (Upper-secondary school, vocational track, less than two years), ISCED 3C long, ISCED 3A-B (Upper-secondary school, academic track), ISCED 3 long without distinction A-B-C, ISCED 4C (Post-secondary non-tertiary degree, vocational track), ISCED 4A-B (Post-secondary non-tertiary degree, academic track), ISCED 4 without distinction A-B-C, ISCED 5B (Tertiary education, professional), ISCED 5A bachelor, ISCED 5A master, ISCED 5A without distinction bachelor-master, ISCED 6 (Research degree), ISCED master-research without distinction. (No country is represented by all of these categories.)

4.5 RESULTS

Overall, these data provide some evidence for the existence of a small cognitive decline from graduating in a recession; higher unemployment rates around the time of leaving school predict lower cognitive performances later on. The ‘standard regressions’ and the instrumental variables approach give similar estimates: As the three-year average unemployment rate at school leaving increases by one percentage point, your future literacy score is predicted to drop by 0.005-0.006 standard deviations depending on estimator. This effect is significant or marginally significant (IV). The effect on numeracy is even smaller and insignificant for both estimators. Furthermore, higher unemployment rates in the years following school leaving predict significantly lower performances on literacy and numeracy later on, conditional on the unemployment rate at school leaving.

4.5.1 The effect of the unemployment rate at school leaving, model (1)

In this section, I measure the effect of the unemployment rate at school leaving, adjusted for observable confounders (see model (1) in section 3). I find that high unemployment rates at the time of leaving school are associated with lower cognitive performances later on. These effects, however, are small and only significant for literacy when the unemployment rate is measured as a three-year average (starting from the year of leaving school). Here, a one-percentage point increase in the mean unemployment rate predicts a 0.005 standard deviation drop in literacy (95 % confidence interval: -0.0095 to -0.0007). This effect is modest indeed, comparable to a decline by 0.14 percentile points.⁹⁶ For numeracy, this effect is even smaller and insignificant. The results are presented in table 2 for literacy; see the Appendix, section A.2, for numeracy.

⁹⁶ The ‘average percentile rank effect’ (APE) is given by:

$$APE = \frac{\beta}{2\sqrt{\pi}}$$

assuming that literacy is normally distributed. For reference, see section 5 in chapter 2.

Table 2. The effect of the unemployment rate at school leaving on literacy (WLS-regressions, model (1))

LITERACY (Z)	ALL	MEN	WOMEN	AGE: 17-30 (mean: 27)	AGE: 31-40 (mean: 35)	AGE: 41-60 (mean: 45)	BELOW TERTIARY	TERTIARY DEGREE
Unemployment, % (year of leaving school)	-0.0017 (0.0021)	-0.0030 (0.0032)	-0.00087 (0.0025)	-0.0035 (0.0045)	-0.00086 (0.0035)	-0.0092 (0.0060)	0.00024 (0.0028)	-0.00087 (0.0031)
Mean unemployment, % (first two years)	-0.0036* (0.0021)	-0.0049 (0.0033)	-0.0024 (0.0026)	-0.0053 (0.0033)	-0.0025 (0.0034)	-0.0083 (0.0053)	-0.0020 (0.0030)	-0.0027 (0.0031)
Mean unemployment, % (first three years)	0.0051** (0.0022)	-0.0067** (0.0033)	-0.0035 (0.0027)	-0.0063 (0.0041)	-0.0036 (0.0034)	-0.0087 (0.0055)	-0.0039 (0.0031)	-0.0045 (0.0032)
Observations	47,842	23,391	24,451	13,767	19,924	14,151	27,735	20,107

Notes: Each estimate corresponds to a separate regression. All regressions include the full set of controls described in section 3.1 and listed in table 1 (as well as fixed effects for country). All effects are expressed in standard deviations and weighted as to account for the country-specific survey designs; on a cross-country level, all nations are given equal weights in all regressions. Standard errors (in parenthesis) are estimated using jackknife replicate sampling weights and corrected for the imputation variance. *** p<0.01, ** p<0.05, * p<0.1

A couple of studies have reported that men are more severely affected than women are by leaving school in a bad economy.⁹⁷ This pattern also shows up in this data set, where the cognitive decline for men is larger than that for women and only significant for men (see table 2). The difference in effects is insignificant, however.⁹⁸

Table 2 also presents estimates separately for other groups. Here, it is of special interest to divide the sample by age. Several studies have shown that the labor market effects of graduating in a bad economy tend to decrease over time. Hence, one might expect any cognitive effects to be larger among the younger cohorts. Yet, I find no such pattern in these data. Here I divide the sample into three groups: those aged 17-30, 31-40 and 41-60. There is no discernable age pattern, however, with effects being rather small and insignificant in all three groups.

Business cycles have been shown to affect poorly and highly educated individuals differently. Hoynes et al. (2012) show that historical fluctuations in unemployment rates are distinctly stronger among those with high school degrees or less, as compared to college graduates. Therefore, one might expect poorly educated individuals to experience larger cognitive deficits from graduating in a bad economy. I observe no such pattern in these data however (see table 2).

Although the cognitive effects of graduating in a bad economy are small as judged from these estimates, the same is not true regarding the labor market outcomes. This is of some importance here, since any cognitive effects are assumed to work through the career path. See the Appendix, section A.3, for the labor market outcomes.

⁹⁷ For example, Maclean (2012) finds that leaving school in a bad economy predicts significantly worse physical functioning and more depressive symptoms for men; for women, she finds no significant effect on physical functioning and a lowering of depressive symptoms.

⁹⁸ The gender difference in effects is 0.0032 standard deviations with a standard error at 0.0043 = $(0.0033^2 + 0.0027^2)^{(1/2)}$ standard deviations.

4.5.2 Instrumenting the unemployment rate at school leaving, model (2)

In this section, I present the results from the main identification strategy where I instrument the unemployment rate at school leaving, using the unemployment rate at the expected time of graduation. The unemployment rate at the expected time of graduation is measured in three ways: firstly, as the unemployment rate in the year when you are expected to graduate; secondly, as the mean unemployment rate in the first two years and thirdly, as the mean unemployment rate in the first three years. In the second and third case, I also instrument the corresponding two-year and three-year actual unemployment rates.

The unemployment rate at the expected time of graduation is a strong instrument: As this unemployment rate increases by one percentage point, the unemployment rate at school leaving increases by 0.65-0.69 percentage points (depending on how the unemployment rate is measured). The first stage F-values are large by any standards in all regressions.⁹⁹ See the Appendix, section A.4, for the first stage estimates.

The unemployment rate in the expected time of graduation is also a significant predictor of your future employment prospects (see section A.5 in the Appendix). Yet, the effects on literacy and numeracy are small and only marginally significant for literacy when I instrument the three-year average unemployment rate. Here, a 1.45 percentage point increase in the unemployment rate at the expected time of graduation increases the *actual* (three-year) unemployment rate by one point, while literacy decreases slightly by 0.006 standard deviations (95 % confidence interval: -0.012 to 0.0007). This is the instrumental variables estimate also presented in table 3 for literacy; see section A.6 in the Appendix for numeracy.

Table 3 also presents estimates separately for different subgroups. The estimated cognitive decline from an increase in the three-year average unemployment rate is somewhat larger for men, for older cohorts and for those with tertiary degrees. None of the group differences is significant, however.¹⁰⁰ In general, estimates are small and insignificant in all groups, one exception being the 31-40 year olds: here, a one-percentage point increase in the three-year average unemployment rate predicts a 0.016 standard deviation drop in literacy ($p < 0.01$).

⁹⁹ As a rule of thumb, F-values should be no lower than 10 (Stock & Yogo, 2002). Here, F-values are no lower than 755.

¹⁰⁰ Tested by calculating the standard error for a difference as $\sqrt{se_1^2 + se_0^2}$, where se_j is the standard error for the point estimate in subgroup j ; $j = 0, 1$. Example: The difference in effects between men and women is 0.0023 standard deviations with a standard error at $\sqrt{0.0042^2 + 0.0045^2} \approx 0.0062$.

Table 3. The effect of the unemployment rate at school leaving on literacy (2SLS-regressions, see model (2))

LITERACY (Z)	ALL	MEN	WOMEN	AGE: 17-30 (mean: 27)	AGE: 31-40 (mean: 35)	AGE: 41-60 (mean: 45)	BELOW TERTIARY	TERTIARY DEGREE
Unemployment, % (year of leaving school)	-0.0014 (0.0033)	-0.00091 (0.0045)	-0.0032 (0.0044)	-0.0028 (0.0067)	-0.0083 (0.0057)	-0.0093 (0.013)	0.0042 (0.0039)	-0.0056 (0.0058)
F.S. F-value (1, 79)	7542.00	3534.31	3833.22	1456.86	1940.54	846.62	3912.99	3157.52
Mean unemployment, % (first two years)	-0.0034 (0.0032)	-0.0040 (0.0043)	-0.0038 (0.0044)	-0.0045 (0.0065)	-0.011** (0.0057)	-0.012 (0.011)	0.0026 (0.0038)	-0.0094 (0.0059)
F.S. F-value (1, 79)	8948.12	4460.52	3849.32	1492.02	2540.40	805.39	5911.76	3531.62
Mean unemployment, % (first three years)	0.0055* (0.0031)	-0.0070 (0.0042)	-0.0047 (0.0045)	-0.0044 (0.0063)	-0.016*** (0.0057)	-0.0095 (0.0099)	-0.00035 (0.0039)	-0.012* (0.0061)
F.S. F-value (1, 79)	9988.77	5016.69	4118.33	1424.29	3255.47	755.36	6261.76	3083.00
Observations	47,842	23,391	24,451	13,767	19,924	14,151	27,735	20,107

Notes: Each estimate corresponds to a separate regression. All regressions include the full set of controls described in section 3.1 and listed in table 1 (as well as fixed effects for country). All effects are expressed in standard deviations and weighted as to account for the country-specific survey designs; on a cross-country level all nations are given equal weights in all regressions. Standard errors (in parenthesis) are estimated using jackknife replicate sampling weights and corrected for the imputation variance. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

4.5.3 The conditional effect of the unemployment rate in the first year following school leaving, model (3)

Lastly, I also experiment with using an alternative identification strategy: I estimate the effect of the unemployment rate in the year(s) following school leaving, conditional on the unemployment rate at school leaving. In other words, I ‘compare’ individuals who experience the same unemployment rate when they first enter the labor market, but where some thereafter experience higher unemployment rates than others. This has a significant effect on your future employment prospects (see the Appendix, section A.7). For example, a one-percentage point increase in the unemployment rate in the first year following school leaving predicts an increase in the risk of not having had a paid job in the last twelve months by 0.32 percentage points ($p < 0.05$).

Furthermore, the post-graduate unemployment rate also has significant effects on literacy and numeracy. As the unemployment rate in the first year following school-leaving increases by one percentage point, literacy and numeracy are predicted to drop by roughly 0.01 standard deviations ($p < 0.01$). This effect is somewhat smaller, but still highly significant, when the post-graduate unemployment rate is measured as a two-year average. These estimates are presented in table 4 for literacy; see section A.8 in the Appendix for numeracy.

Table 4 also shows that the estimated cognitive decline is somewhat larger for men than for women, and smaller for the oldest cohort (41-60 year olds). None of the group differences is significant, however.

Table 4 also reveals that individuals who leave school in a year when the unemployment rate is high are predicted to perform significantly better on literacy later on, conditional on the unemployment rate in the first year following school leaving. Now, a causal interpretation of this effect seems far-fetched. One alternative explanation is sorting; individuals who leave school in a slow economy are overrepresented by high-performers. In section 6, I return to the question of sorting and validity.

Table 4. The effect of the unemployment rate measured *after* leaving school, conditional on the unemployment rate in the year of leaving. Outcome: Literacy (WLS-estimates, model (3))

LITERACY (Z)	ALL	MEN	WOMEN	AGE: 17-30 (mean: 27)	AGE: 31-40 (mean: 35)	AGE: 41-60 (mean: 45)	BELOW TERTIARY	TERTIARY
Unemployment, % (year <i>after</i> leaving school)	-0.013*** (0.0041)	-0.015** (0.0059)	-0.010** (0.0052)	-0.011 (0.0067)	-0.011* (0.0063)	0.0065 (0.015)	-0.016** (0.0062)	-0.013** (0.0059)
Unemployment, % (year of leaving school)	0.0098** (0.0043)	0.0098 (0.0061)	0.0079 (0.0053)	0.0060 (0.0081)	0.0083 (0.0069)	-0.014 (0.016)	0.014** (0.0061)	0.010 (0.0062)
Mean unemployment, % (1 st & 2 nd year after)	-0.011*** (0.0031)	-0.013*** (0.0046)	-0.0079* (0.0041)	-0.0080 (0.0054)	-0.0084* (0.0050)	0.00024 (0.012)	-0.013*** (0.0049)	-0.011** (0.0045)
Unemployment, % (year of leaving school)	0.0062* (0.0033)	0.0065 (0.0048)	0.0048 (0.0040)	0.0026 (0.0065)	0.0051 (0.0056)	-0.0090 (0.013)	0.0096* (0.0046)	0.0074 (0.0047)
Observations	47,842	23,391	24,451	13,767	19,924	14,151	27,735	20,107

Notes: Each estimate corresponds to a separate regression. All regressions include the full set of controls described in section 3.1 and listed in table 1 (as well as fixed effects for country). All effects are expressed in standard deviations and weighted as to account for the country-specific survey designs; on a cross-country level all nations are given equal weights in all regressions. Standard errors (in parenthesis) are estimated using jackknife replicate sampling weights and corrected for the imputation variance. *** p<0.01, ** p<0.05, * p<0.1

4.5.4 Robustness

The estimates are fairly robust to different specifications and sample selection criteria. I experiment with the following six modifications:

(1) Extended set of controls: Here, I allow for all effects to vary by country, keeping only the effect of the unemployment rate (and its instrument) common. In this specification, I also make the following adjustments regarding functional forms: I replace the age dummies with polynomial age trends of the third degree, as to save degrees of freedom. Also, I include 'years of schooling' non-linearly, using a second degree polynomial.

(2) Restricted set of controls: Here, I only include controls for country of residence, age, gender and schooling (linearly).

In the following modifications, I include the same set of covariates as in the main specification:

(3) Extended ('uncleaned') sample: The original sample is restricted in two important ways: First, I only use individuals who were expected to graduate in 1982 to 2008 (for some countries this time frame is shorter). Secondly, I exclude individuals who are observed leaving school at the age of 35 or older. Here I lift these restrictions, which increases the sample by ~11,000 observations.

(4) Extended sample with adult students: The original sample excludes anyone who is still in school. This, however, also excludes adult students, i.e. those who left school years ago but later re-entered. Here I attempt at identifying these individuals: I reinstate those who are currently enrolled in an educational program if they are at least 30 years old and earned their highest degree or dropped out at least ten years ago. This increases the sample by ~1,700 observations.

(5) New instrument: Here I exchange the instrument for the average unemployment rate measured at the age of 18 to 20, and restrict the sample to individuals with an upper-secondary degree at the very least, including also individuals who dropped out from such a program.

(6) Log-specification: Here I measure the unemployment rate on a logarithmic scale (the natural logarithm). I calculate the average unemployment rate by taking the mean of the logarithmic unemployment rates (as opposed to the logarithm of the mean).

The results from these modified regressions are presented in table 5, together with the main specification (0). Here I only use literacy as outcome,

and apply the modifications to all three identification strategies, presented separately by panel.

The estimates are fairly insensitive to the chosen set of controls. For the IV-estimator, I observe no statistically significant differences in point estimates between the main specification (0) and the specification with the extended set of controls (1), nor between the extended set (1) and the restricted set (2).¹⁰¹ This is of some importance here, as instrumental validity is not, a priori, assumed to be conditional on controlling for other background factors, such as parental education levels. For the other two identification strategies, estimates are also fairly stable between these specifications.

Modifications (3) and (4) exploit extended samples, with IV-estimates being more pronounced in sample (3) and somewhat suppressed in sample (4). Taken together, I regard these modifications to strengthen the case for a cognitive decline, if anything.

Perhaps the strongest argument against a cognitive decline is given by specification (6) which measures the unemployment rate on a logarithmic scale. Here, effect sizes are smaller and insignificant when combined with the IV-estimator.¹⁰² Also, using the alternative instrument (5) fails to provide any additional support for a cognitive decline.

¹⁰¹ Compare the main specification (0) to the one with an extended set of controls (1). The difference in point estimates at 0.0017 standard deviations is insignificant, as the standard deviation for this difference is given by $\sqrt{Var(\hat{\beta}_{(0)}) + Var(\hat{\beta}_{(1)}) - 2Cov(\hat{\beta}_{(0)}, \hat{\beta}_{(1)})}$ which is estimated to be 0.0014 standard deviations or higher. Here, $Cov(\hat{\beta}_{(0)}, \hat{\beta}_{(1)})$ is unknown, but it can be no larger than $\sqrt{Var(\hat{\beta}_{(0)}) \cdot Var(\hat{\beta}_{(1)})}$. In any case, the difference at 0.0017 standard deviations is insignificant. Similarly, the difference in point estimates at 0.0027 standard deviations comparing specification (1) and (2) is, at most, marginally significant.

¹⁰² An increase in the average unemployment rate from eight to nine percent predicts a loss in literacy by 0.0055 standard deviations (se: 0.0031) using specification (0) and a loss by 0.0038 standard deviations (se: 0.0034) using specification (6).

Table 5. Robustness: The effect of the unemployment rate at school leaving on literacy using alternative specifications, samples and/or instruments

1. 'Standard regressions'	(0)	(1)	(2)	(3)	(4)	(5)	(6)
Mean unemployment ^a (three years)	-0.0051** (0.0022)	-0.0043* (0.0024)	-0.0047** (0.0022)	-0.0062*** (0.0018)	-0.0044** (0.0021)	-	-0.036* (0.019)
Observations	47,842	47,842	47,842	59,251	49,584		47,842
2. IV	(0)	(1)	(2)	(3)	(4)	(5)	(6)
Mean unemployment ^a (three years)	-0.0055* (0.0031)	-0.0072 (0.0045)	-0.0045 (0.0031)	-0.0088*** (0.0025)	-0.0045 (0.0030)	-0.0036 (0.0062)	-0.032 (0.029)
First stage F-value (1,79)	9988.77	3344.43	10,022.79	9584.16	10,360.79	1665.25	15,420.37
Observations	47,842	47,842	47,842	59,251	49,584	46,702	47,842
3. Conditional unemployment	(0)	(1)	(2)	(3)	(4)	(5)	(6)
Mean unemployment ^a (1 st & 2 nd year after)	-0.011*** (0.0031)	-0.0086*** (0.0032)	-0.011*** (0.0031)	-0.0098*** (0.0029)	-0.011*** (0.0030)	-	-0.081*** (0.027)
Unemployment ^a (year of leaving school)	0.0062* (0.0033)	0.0050 (0.0035)	0.0064* (0.0034)	0.0040 (0.0028)	0.0071** (0.0032)		0.047* (0.026)
Observations	47,842	47,842	47,842	59,251	49,584		47,842

Notes: ^aMean unemployment is the average unemployment rate (%) measured over three years, starting from the year of leaving school. The exception is specification (5) where the unemployment rate is measured on a logarithmic scale (the same goes for the instrument). (0) is the main specification, (1) uses an extended set of controls, (2) uses a restricted set of controls, (3) and (4) use an extended samples, (5) uses an alternative instrument (the mean unemployment rate at age 18-20) and (6) uses a logarithmic scale. All effects are expressed in standard deviations and weighted as to account for the country-specific survey designs; on a cross-country level all nations are given equal weights in all regressions. Standard errors (in parenthesis) are estimated using jackknife replicate sampling weights and corrected for the imputation variance. *** p<0.01, ** p<0.05, * p<0.1

4.6 VALIDITY

In this study, I contrast three empirical strategies relying on different assumptions. Here, I test these assumptions indirectly. The first identification strategy relies on measuring the effect of the unemployment rate at school leaving, adjusted for observable confounders. The validity of this strategy is called into question if individuals postpone or prepone school leaving depending on the state of the labor market. The second identification strategy instruments the school leaving unemployment rate using the unemployment rate in the expected time of graduation. The validity of this approach becomes questionable if individuals also choose their expected school leaving date strategically. This would be the case if students forgo, or enroll in, educational programs depending on labor market conditions.

A couple of studies have investigated these hypotheses, usually finding evidence of such effects (see, for example, Betts & McFarland, 1995; Hershbein, 2012 and Kahn, 2009). For example, Betts & McFarland find that a one-percent increase in the adult unemployment rate increases full-time attendance at American community colleges by four percent, which also carries over to degrees earned. The research evidence is not entirely streamlined, however. Johnson (2013) finds that American graduate school enrollments are only countercyclical for women, but procyclical for men. Raaum & Røed (2006) use Norwegian data and find that students at age 16 and 19 postpone graduation when local unemployment rates are high, but find no evidence of effects on educational attainment.

The current data also provides evidence consistent with the hypothesis that individuals choose their school leaving date strategically, at least to some degree. I test for this by running several logit-regressions, estimating the probability of continuing your studies as a function of the unemployment rate. First, I select everyone who was still in school at the age of 16 and measure whether they continued studying beyond that. To this end, I restrict the sample to individuals who are ‘currently’ at least 30 years old.¹⁰³ I regard you as having continued your studies if you earned your highest degree or dropped out at the age of 17 or later. I also control for country of residence, age (using a second-degree polynomial), gender, immigration status and parental education levels. I repeat this exercise, using the same sample selection criteria, for individuals who were still in school at the age of 17, then 18, and so on, up to age 25. The result is presented in table 6.

¹⁰³ I choose to create this ‘data gap’ for the intervening ages, as to avoid having to code the outcome based on your ‘current’ status as a student or non-student. Instead, I can use the same coding criteria for everyone independently of age, by applying the assumption that anyone who continued studying would have either earned a higher degree or dropped out by the age of 30.

Table 6. The probability of continuing your studies for those who were still in school at age X, as a function of the unemployment rate in that year and later (logit coefficients)

	X = 16	X = 17	X = 18	X = 19	X = 20	X = 21	X = 22	X = 23	X = 24	X = 25
Unemployment, % (at age X)	-0.016 (0.016)	0.039*** (0.012)	0.0090 (0.0090)	0.032*** (0.011)	0.0000076 (0.010)	0.015 (0.011)	0.0075 (0.013)	0.038*** (0.012)	0.0021 (0.012)	0.013 (0.012)
Continued studying, %	95.6	93.3	81.8	85.8	88.6	89.6	85.2	83.5	82.8	81.6
Observations	34,889	35,546	35,354	30,894	27,465	25,541	23,812	21,291	19,023	16,985
	X = 16	X = 17	X = 18	X = 19	X = 20	X = 21	X = 22	X = 23	X = 24	X = 25
Mean unemployment, % (at age X, X+1 & X+2)	-0.018 (0.020)	0.048*** (0.013)	0.025** (0.0097)	-0.026** (0.013)	0.00083 (0.011)	0.010 (0.013)	0.016 (0.014)	0.031*** (0.012)	0.011 (0.012)	0.035** (0.013)
Observations	34,889	35,546	35,354	30,894	27,465	25,541	23,812	21,291	19,023	16,985
	X = 16	X = 17	X = 18	X = 19	X = 20	X = 21	X = 22	X = 23	X = 24	X = 25
Mean unemployment, % (at age X+1 & X+2)	0.0084 (0.027)	0.034 (0.022)	0.045*** (0.016)	0.014 (0.018)	0.0029 (0.019)	-0.019 (0.024)	0.042* (0.022)	-0.024 (0.020)	0.025 (0.018)	0.060*** (0.018)
Unemployment, % (at age X)	0.0098 (0.022)	0.014 (0.021)	-0.023 (0.015)	0.043*** (0.016)	-0.0022 (0.018)	0.031 (0.019)	-0.026 (0.021)	0.055*** (0.021)	-0.015 (0.019)	-0.029 (0.018)
Joint sig., F(2, 78)	0.48	7.18***	4.81**	5.27***	0.01	1.77	1.86	5.21***	1.18	5.99***
Observations	34,889	35,546	35,354	30,894	27,465	25,541	23,812	21,291	19,023	16,985

Notes: All regressions control for country of residence, age, gender, immigration status and parental education levels. All effects are expressed as log odds ratios and weighted as to account for the country-specific survey designs; on a cross-country level all nations are given equal weights in all regressions. Standard errors (in parenthesis) are estimated using jackknife replicate sampling weights. *** p<0.01, ** p<0.05, * p<0.1

Typically, the unemployment rate is positively associated with the probability of continuing your studies (see the top and middle panel of table 6). For those who were still in school at age 17, 18, 23 or 25, this positive relationship is significant, although sometimes only when the unemployment rate is measured as a three-year average (see the middle panel). For example, at the age of 17, 93.3 percent of students continue their studies. Hence, a deep recession, increasing the unemployment rate by five percentage points in that year, is predicted to postpone school leaving for 11 out of 1000 students.¹⁰⁴ For those who were still in school at age 19, however, the effect is significantly negative. It is not fully clear why this would be the case, but here is one possibility: Unemployment rates are known to be strongly auto-correlated. Perhaps a high unemployment rate at age 18 postponed graduation for a group of students, who then graduated at the age of 19 while the unemployment rate was still high. This could be the case if further postponement is especially costly at this age, perhaps demanding entry into tertiary programs. This idea is consistent with the fact that the unemployment rate at age 19 loses any predictive power, once the unemployment rate in the previous year is controlled for (not reported in the table).

Table 6 also shows that there is some evidence that students respond to *future* unemployment rates (see the bottom panel of table 6). For example, students who will experience high future unemployment rates if they leave school at age 18 are significantly more likely to stay in school, conditional on the current unemployment rate. This could reflect a tendency to re-enter the educational system for individuals with little labor market success. In any case, this has implications for this study, as one identification strategy relies on the assumption that the future unemployment rate is exogenous (conditional on the unemployment rate at school leaving). This exercise does weaken the support for this assumption.

Even if these data provide evidence of students choosing their school leaving date strategically, I find no evidence of them also choosing their *expected* school leaving date strategically. I test this by firstly selecting everyone with an upper-secondary degree that are 'currently' at least 30 years old. I predict the probability of having a post-secondary degree as a function of the unemployment rate in the year when you were expected to graduate from upper-secondary school. I control for the same set of covariates as earlier. I repeat this exercise for everyone with a bachelor's

¹⁰⁴ First, 933 out of 1000 students continue their studies, i.e. the odds of continuing is $933/67 = 13.925\dots$. If the unemployment rate increases by one percentage point, the odds of continuing increases by a factor of $\exp(0.039)$; if the unemployment rate increases by five percentage points, the odds increases by a factor of $\exp(0.039)^5 = 1.215\dots$. Hence, the odds of continuing is now $16.923\dots$, which is roughly equal to 944 students out of 1000.

degree. In both cases, the unemployment rate is unable to significantly predict the event of earning a higher degree (or dropping out from a higher educational program). See section A.9 in the Appendix.

4.7 CONCLUSIONS

Graduating in a bad economy is potentially harmful for your future cognitive functioning. As judged from these data, such an effect is likely to be small, however. The main estimate from the instrumental variables approach suggests that a one-percentage point increase in the unemployment rate at school leaving decreases literacy by 0.006 standard deviations, with an even smaller effect on numeracy. Here, the unemployment rate is measured as a three-year average, starting from the year of leaving school; outcomes are measured 1-34 years after leaving school (mean: 14 years). To make it concrete, Finnish students who graduated in 1988 experienced a labor market where the unemployment rate averaged at 3.5 percent over the following three years (1988-1990). For those who graduated five years later, in 1992, the corresponding unemployment rate was 16.2 percent. This massive change in employment prospects is predicted to have only a modest effect on your future literacy performance, equal to a loss at 0.07 standard deviations or roughly two percentile points.

It can be noted, however, that a conservative interpretation of the data, heavily punishing a type one error, would also make room for a null effect: Firstly, the IV-estimate is only marginally significant. Furthermore, the effect becomes smaller and insignificant when combining the IV-estimator with a log-specification. Perhaps more importantly, however, the mere fact that the estimate is close to zero, makes its existence sensitive to the influence of confounders. The validity tests do support the hypothesis of students choosing their school leaving date strategically. It is also possible that they choose their *expected* school leaving date strategically, at least to some degree, although I find no evidence for this in the data. In either case, the unemployment rate at graduation is likely to be an economically unimportant factor for explaining the cohort-to-cohort variation in literacy and numeracy among adults later in life.

References

- Acemoglu, D. (1995). Public policy in a model of long-term unemployment. *Economica*, 161-178.
- Alhola, P., & Polo-Kantola, P. (2007). Sleep deprivation: Impact on cognitive performance. *Neuropsychiatric Disease and Treatment*.
- Apouey, B., & Clark, A. E. (2015). Winning big but feeling no better? The effect of lottery prizes on physical and mental health. *Health Economics*, 24(5), 516-538.
- Au, N., & Johnston, D. W. (2014). Self-assessed health: what does it mean and what does it hide?. *Social Science & Medicine*, 121, 21-28.
- Au, N., & Johnston, D. W. (2015). Too much of a good thing? Exploring the impact of wealth on weight. *Health Economics*, 24(11), 1403-1421.
- Beale, N., & Nethercott, S. (1985). Job-loss and family morbidity: a study of a factory closure. *JR Coll Gen Pract*, 35(280), 510-514.
- Betts, J. R., & McFarland, L. L. (1995). Safe port in a storm: The impact of labor market conditions on community college enrollments. *Journal of Human Resources*, 741-765.
- Björklund, A., & Eriksson, T. (1998). Unemployment and mental health: evidence from research in the Nordic countries. *International Journal of Social Welfare*, 7(3), 219-235.
- Brand, J. E., & Burgard, S. A. (2008). Job displacement and social participation over the lifecourse: Findings for a cohort of joiners. *Social Forces*, 87(1), 211-242.
- Browning, M., & Heinesen, E. (2012). Effect of job loss due to plant closure on mortality and hospitalization. *Journal of Health Economics*, 31(4), 599-616.
- Browning, M., Moller Dano, A., & Heinesen, E. (2006). Job displacement and stress-related health outcomes. *Health Economics*, 15(10), 1061-1075.
- Brunner, B., & Kuhn, A. (2014). The impact of labor market entry conditions on initial job assignment and wages. *Journal of Population Economics*, 27(3), 705-738.
- Böckerman, P., & Ilmakunnas, P. (2009). Unemployment and self-assessed health: evidence from panel data. *Health Economics*, 18(2), 161-179.
- Clark, A., Georgellis, Y., & Sanfey, P. (2001). Scarring: The psychological impact of past unemployment. *Economica*, 68(270), 221-241.
- Cutler, D. M., Huang, W., & Lleras-Muney, A. (2015). When does education matter? The protective effect of education for cohorts graduating in bad times. *Social Science & Medicine*, 127, 63-73.
- Dooley, D., Catalano, R., & Wilson, G. (1994). Depression and unemployment: panel findings from the Epidemiologic Catchment Area study. *American Journal of Community Psychology*, 22(6), 745-765.
- Edin, P. A., & Gustavsson, M. (2008). Time out of work and skill depreciation. *ILR Review*, 61(2), 163-180.

- Eliason, M. (2012). Lost jobs, broken marriages. *Journal of Population Economics*, 25(4), 1365-1397.
- Eliason, M., & Storrie, D. (2009). Does job loss shorten life?. *Journal of Human Resources*, 44(2), 277-302.
- Ferrie, J. E., Shipley, M. J., Stansfeld, S. A., & Marmot, M. G. (2002). Effects of chronic job insecurity and change in job security on self reported health, minor psychiatric morbidity, physiological measures, and health related behaviours in British civil servants: the Whitehall II study. *Journal of Epidemiology & Community Health*, 56(6), 450-454.
- Genda, Y., Kondo, A., & Ohta, S. (2010). Long-term effects of a recession at labor market entry in Japan and the United States. *Journal of Human Resources*, 45(1), 157-196.
- Golberstein, E. (2015). The effects of income on mental health: evidence from the social security notch. *The Journal of Mental Health Policy and Economics*, 18(1), 27.
- Grossman, M. (1972). On the concept of health capital and the demand for health. *Journal of Political Economy*, 80(2), 223-255.
- Gunasekara, F. I., Carter, K., & Blakely, T. (2011). Change in income and change in self-rated health: Systematic review of studies using repeated measures to control for confounding bias. *Social Science & Medicine*, 72(2), 193-201.
- Han, K. M., Chang, J., Won, E., Lee, M. S., & Ham, B. J. (2017). Precarious employment associated with depressive symptoms and suicidal ideation in adult wage workers. *Journal of Affective Disorders*, 218, 201-209.
- Hansen, H. T. (2005). Unemployment and marital dissolution: A panel data study of Norway. *European Sociological Review*, 21(2), 135-148.
- Hanushek, E. A., Schwerdt, G., Wiederhold, S., & Woessmann, L. (2015). Returns to skills around the world: Evidence from PIAAC. *European Economic Review*, 73, 103-130.
- Hershbein, B. J. (2012). Graduating high school in a recession: Work, education, and home production. *The BE Journal of Economic Analysis & Policy*, 12(1).
- Hoynes, H., Miller, D. L., & Schaller, J. (2012). Who suffers during recessions?. *Journal of Economic Perspectives*, 26(3), 27-48.
- Hwang, J., Castelli, D. M., & Gonzalez-Lima, F. (2017). The positive cognitive impact of aerobic fitness is associated with peripheral inflammatory and brain-derived neurotrophic biomarkers in young adults. *Physiology & Behavior*, 179, 75-89.
- Johnson, M. T. (2013). The impact of business cycle fluctuations on graduate school enrollment. *Economics of Education Review*, 34, 122-134.
- Kahn, L. B. (2010). The long-term labor market consequences of graduating from college in a bad economy. *Labour Economics*, 17(2), 303-316.
- Kassenboehmer, S. C., & Haisken-DeNew, J. P. (2009). You're fired! The causal negative effect of entry unemployment on life satisfaction. *The Economic Journal*, 119(536), 448-462.

- Kim, M. H., Kim, C. Y., Park, J. K., & Kawachi, I. (2008). Is precarious employment damaging to self-rated health? Results of propensity score matching methods, using longitudinal data in South Korea. *Social Science & Medicine*, 67(12), 1982-1994.
- Kim, I. H., Muntaner, C., Khang, Y. H., Paek, D., & Cho, S. I. (2006). The relationship between nonstandard working and mental health in a representative sample of the South Korean population. *Social Science & Medicine*, 63(3), 566-574.
- Kim, B., & Ruhm, C. J. (2012). Inheritances, health and death. *Health Economics*, 21(2), 127-144.
- Kondo, A. (2015). Differential effects of graduating during a recession across gender and race. *IZA Journal of Labor Economics*, 4(1), 23.
- Krueger, A. B., & Mueller, A. (2011). Job search, emotional well-being, and job finding in a period of mass unemployment: Evidence from high-frequency longitudinal data. *Brookings Papers on Economic Activity*, 2011(1), 1-57.
- Kuhn, A., Lalive, R., & Zweimüller, J. (2009). The public health costs of job loss. *Journal of Health Economics*, 28(6), 1099-1115.
- Kwon, I., Milgrom, E. M., & Hwang, S. (2010). Cohort effects in promotions and wages evidence from Sweden and the United States. *Journal of Human Resources*, 45(3), 772-808.
- Lindqvist, E., Östling, R., & Cesarini, D. (2018). *Long-run Effects of Lottery Wealth on Psychological Well-being* (No. w24667). National Bureau of Economic Research.
- Liu, K., Salvanes, K. G., & Sørensen, E. Ø. (2016). Good skills in bad times: Cyclical skill mismatch and the long-term effects of graduating in a recession. *European Economic Review*, 84, 3-17.
- Loprinzi, P. D., & Kane, C. J. (2015, April). Exercise and cognitive function: a randomized controlled trial examining acute exercise and free-living physical activity and sedentary effects. In *Mayo Clinic Proceedings* (Vol. 90, No. 4, pp. 450-460). Elsevier.
- Maclean, J. C. (2013). The health effects of leaving school in a bad economy. *Journal of Health Economics*, 32(5), 951-964.
- Maclean, J. C. (2015). The lasting effects of leaving school in an economic downturn on alcohol use. *ILR Review*, 68(1), 120-152.
- Maclean, J. C. (2016). Does leaving school in an economic downturn persistently affect body weight? Evidence from panel data. *Industrial Relations: A Journal of Economy and Society*, 55(1), 122-148.
- Maclean, J. C., Covington, R., & Sikora Kessler, A. (2016). Labor Market Conditions at School-Leaving: Long-Run Effects on Marriage and Fertility. *Contemporary Economic Policy*, 34(1), 63-88.
- Maclean, J. C., & Hill, T. D. (2015). Leaving school in an economic downturn and self-esteem across early and middle adulthood. *Labour Economics*, 37, 1-12.

- Maclean, J. C., & Hill, T. D. (2017). Economic conditions at school leaving and sleep patterns across the life course. *The BE Journal of Economic Analysis & Policy*, 17(2).
- Mani, A., Mullainathan, S., Shafir, E., & Zhao, J. (2013). Poverty impedes cognitive function. *Science*, 341(6149), 976-980.
- Meraya, A. M., Dwibedi, N., Tan, X., Innes, K., Mitra, S., & Sambamoorthi, U. (2018). The dynamic relationships between economic status and health measures among working-age adults in the United States. *Health Economics*, 27(8), 1160-1174.
- OECD (2016). *Skills Matter: Further Results from the Survey of Adult Skills*, OECD Skills Studies, OECD Publishing, Paris
- Oreopoulos, P., Von Wachter, T., & Heisz, A. (2012). The short-and long-term career effects of graduating in a recession. *American Economic Journal: Applied Economics*, 4(1), 1-29.
- Oyer, P. (2006). Initial labor market conditions and long-term outcomes for economists. *The Journal of Economic Perspectives*, 20(3), 143-160.
- Paul, K. I., & Moser, K. (2009). Unemployment impairs mental health: Meta-analyses. *Journal of Vocational Behavior*, 74(3), 264-282.
- Pirani, E., & Salvini, S. (2015). Is temporary employment damaging to health? A longitudinal study on Italian workers. *Social Science & Medicine*, 124, 121-131.
- Pissarides, C. A. (1992). Loss of skill during unemployment and the persistence of employment shocks. *The Quarterly Journal of Economics*, 107(4), 1371-1391.
- Quesnel-Vallée, A., DeHaney, S., & Ciampi, A. (2010). Temporary work and depressive symptoms: a propensity score analysis. *Social Science & Medicine*, 70(12), 1982-1987.
- Raaum, O., & Røed, K. (2006). Do business cycle conditions at the time of labor market entry affect future employment prospects?. *The Review of Economics and Statistics*, 88(2), 193-210.
- Salm, M. (2009). Does job loss cause ill health?. *Health Economics*, 18(9), 1075-1089.
- Salthouse, T. A. (2006). Mental exercise and mental aging: Evaluating the validity of the "use it or lose it" hypothesis. *Perspectives on Psychological Science*, 1(1), 68-87.
- Sandi, C. (2013). Stress and cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, 4(3), 245-261.
- Schmitz, H. (2011). Why are the unemployed in worse health? The causal effect of unemployment on health. *Labour Economics*, 18(1), 71-78.
- Stock, J. H., Wright, J. H., & Yogo, M. (2002). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business & Economic Statistics*, 20(4), 518-529.
- Sullivan, D., & Von Wachter, T. (2009). Job displacement and mortality: An analysis using administrative data. *The Quarterly Journal of Economics*, 124(3), 1265-1306.

- Virtanen, M., Kivimäki, M., Joensuu, M., Virtanen, P., Elovainio, M., & Vahtera, J. (2005). Temporary employment and health: a review. *International Journal of Epidemiology*, 34(3), 610-622.
- Winkelmann, L., & Winkelmann, R. (1998). Why are the unemployed so unhappy? Evidence from panel data. *Economica*, 65(257), 1-15.
- Winter, B., Breitenstein, C., Mooren, F. C., Voelker, K., Fobker, M., Lechtermann, A., Krueger, K., Fromme, A., Korsukewitz, C., Floel, A., Knecht, S. (2007). High impact running improves learning. *Neurobiology of Learning and Memory*, 87(4), 597-609.
- Yates, J. A. (2005). The transition from school to work: education and work experiences. *Monthly Lab. Rev.*, 128, 21.
- Yates, K. F., Sweat, V., Yau, P. L., Turchiano, M. M., & Convit, A. (2012). Impact of metabolic syndrome on cognition and brain: a selected review of the literature. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 32(9), 2060-2067.

Appendix

A.1 Median age at graduation

Table A1. Median age at graduation, by educational degree

Country	Prim.	Sec. low.	Sec. upp. short	Sec. upp. long	Post sec.	Tert. prof.	Bach.	Mast.	Res.
Belgium	14	15	-	18	19	21	22	23	29
Chile	14	15	-	18	-	23	25	29	29 ^a
Cyprus	-	15	-	18	-	21	24	25	30
Czech R.	-	15	18	19	21	23	24	25	26
Denmark	-	16	21	21	22	25	26	27	29
Finland	15	16	-	19	24	23	25	27	29
France	14	16	-	18	-	22	23	24	27
Greece	-	15	-	18	21	23	24	25	25 ^a
Ireland	14	16	-	18	20	22	23	25	27
Israel	-	15	-	18	-	22	26	28	28 ^a
Italy	-	14	17	19	23	-	26	28	28 ^a
Japan	-	15	18	18	19	20	22	24	24 ^a
Korea	-	16	-	19	-	23	25	27	31
Lithuania	-	16	-	18	20	22	23	24	25 ^a
Netherl.	-	16	-	19	-	23	23	26	31
Norway	-	16	19	20	22	24	25	26	31
Poland	15	15	-	20	21	-	23	25	25 ^a
Russia	15	16	-	17	19	20	22	23	25
Slovakia	15	15	18	19	22	-	23	24	27
Slovenia	14	15	18	19	-	25	25	28	30
Spain	14	16	18	18	21	20	23	24	30
Sweden	15	16	16	19	22	24	25	26	29
Turkey	-	15	-	18	-	22	23	24	25 ^a
United K.	-	17	16	18	-	22	23 ^b	23 ^b	-

Notes: Prim. = Primary education (ISCED 1). Sec. low. = Lower secondary school (ISCED 2). Sec. upp. short = Upper secondary school short (ISCED 3C less than two years). Sec. upp long = Upper secondary school long (ISCED 3C two years or more, ISCED 3A-B or ISCED 3 two years or more without distinction A-B-C, depending on country; for a country represented by at least two of these categories, the most abundant one is reported). Post sec. = Post-secondary, non-tertiary degree (ISCED 4C, ISCED 4A-B or ISCED 4 without distinction A-B-C, depending on country. For a country represented by at least two of these categories, the most abundant one is reported). Tert. prof. = Tertiary degree, professional (ISCED 5B). Bach. = Bachelor (ISCED 5A bachelor) or ^bBachelor/Master (ISCED 5A bachelor-master without distinction). Mast. = Master (ISCED 5A, master) or ^bBachelor/Master (ISCED 5A without distinction). Res. = Research degree (ISCED 6) alternatively ^aMaster/Research (ISCED 5A master & ISCED 6 without distinction).

Table A2. The effect of the unemployment rate at school leaving on numeracy (WLS-regressions, model (1))

NUMERACY (Z)	ALL	MEN	WOMEN	AGE: 17-30 (mean: 27)	AGE: 31-40 (mean: 35)	AGE: 41-60 (mean: 45)	BELOW TERTIARY	TERTIARY
Unemployment, % (year of leaving school)	- 0.00058 (0.0021)	-0.0015 (0.0032)	0.00010 (0.0027)	-0.0054 (0.0046)	0.0011 (0.0032)	-0.010* (0.0057)	0.00092 (0.0028)	-0.0019 (0.0036)
Mean unemployment, % (first two years)	-0.0021 (0.0022)	-0.0035 (0.0033)	-0.00066 (0.0027)	-0.0069 (0.0046)	-0.00039 (0.0033)	-0.0087* (0.0050)	-0.0010 (0.0028)	-0.0033 (0.0036)
Mean unemployment, % (first three years)	-0.0032 (0.0023)	-0.0052 (0.0034)	-0.00093 (0.0028)	-0.0077 (0.0046)	-0.0011 (0.0034)	-0.0092* (0.0049)	-0.0023 (0.0028)	-0.0045 (0.0036)
Observations	47,842	23,391	24,451	13,767	19,924	14,151	27,735	20,107

Notes: Each estimate corresponds to a separate regression. All regressions include the full set of controls listed in table 1 as well as fixed effects for country. All effects are expressed in standard deviations and weighted as to account for the country-specific survey designs; on a cross-country level, all nations are given equal weights in all regressions. Standard errors (in parenthesis) are estimated using jackknife replicate sampling weights and further corrected for the imputation variance. *** p<0.01, ** p<0.05, * p<0.1

A.2 The effect of the unemployment rate at school leaving on numeracy, model (1)

Summary of table A2: The unemployment rate at school leaving has no significant effect on numeracy in any sample, independently of how the unemployment rate is measured (at the year of leaving school, or as a two- or three-year average).

A.3 Predicting employment prospects as a function of the unemployment rate at school leaving (WLS-regressions)

Summary of table A3: As the three-year average unemployment rate at school leaving increases by one percentage point, the risk of not having had a paid job in the last year increases by 0.35 percentage points ($p < 0.01$) and the risk of never having had a paid job increases by 0.19 percentage points ($p < 0.01$). Estimates are similar, but somewhat smaller in magnitude, when the unemployment rate is measured as a one- or two-year average.

Table A3. The effect of the unemployment rate at school leaving on labor market outcomes (WLS-regressions)

	No paid job (last 12 months)	No paid job ever
Unemployment, % (year of leaving)	0.0027*** (0.00080)	0.0013*** (0.00047)
Mean unemployment (first two years)	0.0031*** (0.00084)	0.0016*** (0.00047)
Mean unemployment (first three years)	0.0035*** (0.00086)	0.0019*** (0.00049)
Observations	47,834	47,837

Notes: Each estimate corresponds to a separate regression. All regressions include the full set of controls described in section 3.1 and listed in table 1 (including also country fixed effects). All effects are weighted as to account for the country-specific survey designs; on a cross-country level all countries are given equal weights in all regressions. Standard errors (in parenthesis) are estimated using jackknife replicate sampling weights. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

A.4 First stage estimates

Summary of table A4: As the unemployment rate in the expected year of graduation increases by one percentage point, the actual unemployment rate increases by 0.65 percentage points. This relationship is strong with an F-value at 7542 (df = 1, 79). The relationships are similar when the unemployment rate is measured as a two- or three-year average.

Table A4. First stage estimates

FIRST STAGE	Actual unemployment (first year)	Actual unemployment (first 2 years)	Actual unemployment (first 3 years)
Expected unemployment ^a	0.65*** (0.0075)	0.67*** (0.0071)	0.69*** (0.0069)
F-value (1, 79)	7542.00	8948.12	9988.77
Observations	47,842	47,842	47,842

^a*Expected unemployment* is the unemployment rate (%) in the expected time of graduation. This is measured in three ways: as the unemployment rate in the year when you are expected to graduate (with the corresponding estimate presented in the first column); as the mean unemployment rate in the first two years (second column) and as the mean unemployment rate in the first three years (third column). All regressions include the full set of controls described in section 3.1 and listed in table 1 (including also country fixed effects). All effects are weighted as to account for the country-specific survey designs; on a cross-country level all countries are given equal weights in all regressions. Standard errors (in parenthesis) are estimated using jackknife replicate sampling weights. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

A.5 Predicting employment prospects as a function of the unemployment rate at school leaving (IV-estimates)

Summary of table A5: As the three-year average unemployment rate in the *expected* time of graduation increases by 1.45 percentage points, the corresponding actual unemployment rate increases by one point ($p < 0.01$) and the risk of not having had a paid job in the last year increases by 0.30 percentage points ($p < 0.05$). Furthermore, the risk of never having had a paid job increases by 0.21 percentage points ($p < 0.01$). The effects are smaller in magnitude – and generally insignificant – when the unemployment rate is measured as a one- or two-year average. All effects are adjusted for the covariates described in section 3.1.

Table A5. The effect of the school leaving unemployment rate on employment prospects (IV-regressions)

	No job last 12 months	No job ever
Unemployment, % (year of leaving)	0.0019 (0.0012)	0.00089 (0.00061)
Mean unemployment, % (first two years)	0.0024* (0.0012)	0.0014** (0.00061)
Mean unemployment; % (first three years)	0.0030** (0.0012)	0.0021*** (0.00063)
Observations	47,834	47,837

Notes: All regressions include the full set of controls described in section 3.1 and listed in table 1 (including also country fixed effects). All effects are weighted as to account for the country-specific survey designs; on a cross-country level all countries are given equal weights in all regressions. Standard errors (in parenthesis) are estimated using jackknife replicate sampling weights. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

A.6 The effect of the unemployment rate at school leaving on numeracy, 2SLS-estimates

Summary of table A6: I find no significant effects of the unemployment rate at school leaving on numeracy when using the instrumental variables approach. Estimates are generally small in magnitude as compared to the corresponding WLS-estimates, and small in magnitude as compared to the corresponding literacy-estimates.

Table A6. The effect of the unemployment rate at school leaving on numeracy (2SLS-regressions, see model (2))

NUMERACY (Z)	ALL	MEN	WOMEN	AGE: 17-30 (mean: 27)	AGE: 31-40 (mean: 35)	AGE: 41-60 (mean: 45)	BELOW TERTIARY	TERTIARY
Unemployment, % (year of leaving school)	0.00057 (0.0035)	0.0020 (0.0045)	-0.0019 (0.0050)	-0.0020 (0.0068)	-0.0032 (0.0058)	-0.013 (0.012)	0.0049 (0.0041)	-0.0041 (0.0051)
F.S. F-value (1, 79)	7542.00	3534.31	3833.22	1456.86	1940.54	846.62	3912.99	3157.52
Mean unemployment, % (first two years)	-0.0010 (0.0034)	-0.0010 (0.0043)	-0.0018 (0.0049)	-0.0031 (0.0066)	-0.0052 (0.0059)	-0.012 (0.011)	0.0033 (0.0040)	-0.0069 (0.0051)
F.S. F-value (1, 79)	8948.12	4460.52	3849.32	1492.02	2540.40	805.39	5911.76	3531.62
Mean unemployment, % (first three years)	-0.0029 (0.0033)	-0.0042 (0.0043)	-0.0021 (0.0049)	-0.0042 (0.0063)	-0.0079 (0.0060)	-0.0067 (0.0098)	0.00069 (0.0041)	-0.0085 (0.0052)
F.S. F-value (1, 79)	9988.77	5016.69	4118.33	1424.29	3255.47	755.36	6261.76	3083.00
Observations	47,842	23,391	24,451	13,767	19,924	14,151	27,735	20,107

Notes: Each estimate corresponds to a separate regression. All regressions include the full set of controls listed in table 1 as well as fixed effects for country. All effects are expressed in standard deviations and weighted as to account for the country-specific survey designs; on a cross-country level all nations are given equal weights in all regressions. Standard errors (in parenthesis) are estimated using jackknife replicate sampling weights and corrected for the imputation variance. *** p<0.01, ** p<0.05, * p<0.1

A.7 Predicting labor market outcomes as a function of the unemployment rate in the year(s) following school leaving, conditional on the unemployment rate at school leaving

Summary of table A7: As the conditional unemployment rate in the first year following school leaving increases by one percentage point, the risk of not having had a paid job in the last year increases by 0.32 percentage points ($p < 0.05$) and the risk of never having had a paid job increases by 0.20 percentage points ($p < 0.05$). These estimates are similar in magnitude, but somewhat more precisely measured, when the post-graduate unemployment rate is measured as a two-year average. All effects are adjusted for the covariates described in section 3.1.

Table A7. The effect of the unemployment rate in the year(s) following school leaving, conditional on the unemployment rate at school leaving (WLS-estimates)

	No paid job (last 12 months)	No paid job ever
Unemployment, % (year after leaving)	0.0032** (0.0016)	0.0020** (0.00096)
Unemployment, % (year of leaving)	-0.000063 (0.0015)	-0.00045 (0.00098)
Mean unemployment (1 st and 2 nd year after)	0.0028** (0.0013)	0.0020*** (0.00074)
Unemployment, % (year of leaving)	0.00059 (0.0013)	-0.00017 (0.00074)
Observations	47,834	47,837

Notes: All regressions include the full set of controls described in section 3.1. All effects are weighted as to account for the country-specific survey designs; on a cross-country level, all countries are given equal weights in all regressions. Standard errors (in parenthesis) are estimated using jackknife replicate sampling weights. Here, your work experience is top-coded so that the maximum value is given by age-15. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table A8. The effect of the unemployment rate measured *after* leaving school, conditional on the unemployment rate in the year of leaving. Outcome: Numeracy (WLS-estimates, model (3))

NUMERACY (Z)	ALL	MEN	WOMEN	AGE: 17-30 (mean: 27)	AGE: 31-40 (mean: 35)	AGE: 41-60 (mean: 45)	BELOW TERTIARY	TERTIARY
Unemployment, % (year <i>after</i> leaving school)	-0.011*** (0.0041)	-0.015** (0.0059)	-0.0050 (0.0048)	-0.0094 (0.0068)	-0.0097 (0.0061)	0.0094 (0.015)	-0.013** (0.0056)	-0.0092 (0.0061)
Unemployment, % (year of leaving school)	0.0089** (0.0042)	0.012* (0.0061)	0.0044 (0.0050)	0.0031 (0.0077)	0.0093 (0.0061)	-0.018 (0.016)	0.012** (0.0058)	0.0061 (0.0065)
Mean unemployment, % (1 st & 2 nd year after)	0.0083** (0.0033)	-0.013*** (0.0047)	-0.0030 (0.0041)	-0.0072 (0.0055)	-0.0063 (0.0051)	0.0020 (0.014)	-0.010** (0.0043)	-0.0082 (0.0050)
Unemployment, % (year of leaving school)	0.0055* (0.0033)	0.0078 (0.0048)	0.0023 (0.0041)	0.000074 (0.0061)	0.0055 (0.0049)	-0.011 (0.014)	0.0082* (0.0044)	0.0041 (0.0053)
Observations	47,842	23,391	24,451	13,767	19,924	14,151	27,735	20,107

Notes: All regressions include the full set of controls listed in table 1 as well as fixed effects for country. All effects are expressed in standard deviations and weighted as to account for the country-specific survey designs; on a cross-country level all nations are given equal weights in all regressions. Standard errors (in parenthesis) are estimated using jackknife replicate sampling weights and corrected for the imputation variance. *** p<0.01, ** p<0.05, * p<0.1

A.8 The conditional effect of the unemployment rate in the first year(s) following school leaving on *numeracy*, model (3)

Summary of table A8: As the conditional unemployment rate in the first year following graduation increases by one percentage point, the risk of not having had a paid job in the last year increases by 0.18 percentage points ($p > 0.1$) and the risk of never having had a paid job increases by 0.20 percentage points ($p < 0.05$). The effect on average work experience is insignificant at -0.023 years. For the first two outcomes, estimates are based on linear probability models. All effects are adjusted for the covariates described in section 3.1. These estimates are similar, but somewhat larger in magnitude and more precisely measured, when the post-graduation unemployment rate is calculated as a two-year average.

A.9 The probability of entering a higher educational program

Summary of table A9: Looking specifically at individuals with an upper secondary degree, I find no significant effect on the probability of earning a higher degree depending on the unemployment rate in the year when you were *expected* to earn your upper secondary degree. Similarly, the unemployment rate in the year when you were expected to earn your bachelors degree, does not significantly predict the probability of earning a higher degree.

Table A9. The probability of entering a higher educational program for those with degree X, as a function of the unemployment rate in the year when you were expected to claim that degree (logit coefficients)

	X = Upper secondary	X = Bachelor
Unemployment, % (expected year)	-0.0035 (0.0076)	0.017 (0.012)
Observations	37,097	13,907

Notes: All regressions control for country of residence, age, gender, immigration status and parental education levels. All effects are expressed as log odds ratios and weighted as to account for the country-specific survey designs; on a cross-country level all nations are given equal weights in all regressions. Standard errors (in parenthesis) are estimated using jackknife replicate sampling weights. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

CHAPTER 5

The Effect of Retiring on Cognitive Functioning and Subjective Health

Abstract

This study uses PIAAC survey data to estimate the effect of retiring on cognitive abilities and subjective health, where cognitive abilities are measured via literacy and numeracy. The data is a pooled cross-section covering roughly 40,000 individuals in 25 countries in the age range of 50 to 65. The identification relies on instrumenting *retirement* as well as *years as retired* by exploiting the variation in incentives created by country-specific retirement ages. I find no significant discontinuity in cognitive functioning as individuals reach the retirement age. However, I do find a significant trend-break, suggesting that retiring slows down the age-related decline in literacy and numeracy among men. For women, the evidence points towards a positive health-effect.

5.1 INTRODUCTION

Effective retirement ages have been slowly increasing in most OECD-countries over the 21st century. As compared to the year 2000, men and women in 2014 were retiring 1.4 and 2.0 years later on average.¹⁰⁵ A majority of OECD-countries are now in the process of postponing official retirement ages or planning such reforms. These reforms are driven by the pressure on social security systems; the effect on the cognitive and physical health of the elderly is unclear.

In this study, I investigate if retiring affects your cognitive functioning and subjective health, where your cognitive functioning is measured by literacy and numeracy (see chapter 1). The data is a cross-section covering roughly 40,000 individuals from 25 countries in the age range of 50 to 65. The identification relies on instrumenting *retirement* as well as your *retirement duration* by exploiting the incentives created by country-specific retirement ages. I find no significant *discontinuity* in cognitive functioning as you reach the retirement age. However, I do find a significant trend-break among men, suggesting that retiring slows down the age-related decline in literacy and numeracy. The main estimate suggests that retiring causes a performance advantage in literacy that grows by 0.08 (SE 0.03) standard deviations for each additional year as retired; the effect is smaller and only marginally significant for numeracy. These estimates should not, however, be extrapolated more than a couple of years into retirement.¹⁰⁶

For women, I find no robust evidence for a cognitive effect of retiring, although the trend is similar to that of men. On the other hand, there is evidence of a positive effect on health: entering retirement is estimated to increase your subjective health score by almost 60 percent of a standard deviation (SE 18). The evidence on health is less conclusive for men.

¹⁰⁵ OECD Pensions at a glance, 2015

¹⁰⁶ For the men in the data who have reached the retirement age: roughly 90 percent reached it within the last 5.5 years (the average is 2.6 years).

5.2 BACKGROUND AND PREVIOUS RESEARCH

The results from previous studies are mixed, finding both significant benefits and harms from retiring; there is no general agreement on the effects on cognitive functioning nor health. A priori, there are reasonable arguments on both sides. Firstly, many jobs are presumably mentally stimulating as compared to retirement, possibly preventing age-related declines in cognitive functioning. This is the so called ‘use it or lose it’-hypothesis, which has also gained some empirical support. For example, Fisher et al. (2014) show that individuals with intellectually demanding jobs experience slower rates of cognitive decline (both in the years before and after retirement). Intervention studies have also found that mental stimulation can improve cognitive performance among the elderly (see the meta-analysis by Kelly et al., 2014).¹⁰⁷ A related argument is presented in the human capital framework by Mazzonna & Peracchi (2012) building on Grossman (1972) where the incentives for cognitive investments decline after retirement as the productivity motive is lost. Also, some authors characterize retirement as a stressful life-event where individuals lose their work-identity and access to social networks, and are forced to adjust to new routines.

On the other hand, working life can also be stressful, especially among the elderly in a transforming work environment. As compared to workers, American retirees of the same age spend more time socializing, relaxing, reading, watching TV, sleeping, gardening and engaging in other leisure- and sport activities (Krantz-Kent & Stewart, 2007). This pattern can also be fitted into a health capital framework, where your health can improve after retirement if the opportunity cost of making health investments (e.g. exercising or sleeping) decrease.

As these arguments leave much room for speculation, several recent studies have measured the effect of retiring empirically. The descriptive data tend to show that retirees have lower cognitive skills and more health issues than workers. There are, however, several possible interpretations of these relationships. Partly, they could reflect a causal connection, but it’s also likely that confounders and reversed causation influence the correlation. To address the endogeneity of retirement, mainly two different strategies have been employed.

Firstly, there are a number of studies that exploit longitudinal datasets and follow individuals over time as they retire (see, for example, Dave et al., 2006; de Grip et al. 2015 and Mein et al., 2003 as well as the reviews by Meng et al., 2017 and van der Heide et al., 2013). It should be noted, however, that

¹⁰⁷ See also Salhouse (2010) who argue that the empirical support for this hypothesis is unconvincing as of yet.

this strategy is still open to critique regarding reversed causation; if individuals tend to opt for retirement as their cognitive or physical health start to decline, this would be captured as part of the estimate.¹⁰⁸ As for health, the findings from these studies are mixed. While Dave et al. find significant and negative effects on several health outcomes using American data, Mein et al. find that retirement at age 60 has no effect on physical health using British data. Based on a review of 22 longitudinal studies, van der Heide et al. conclude that the evidence are contradictory. The picture isn't much clearer with regard to cognitive functioning. Based on seven longitudinal studies, Meng et al. conclude that there is 'weak evidence' supporting the notion that retirement accelerates the rate of decline in crystallized intelligence among those retiring from with socially complex jobs.

Secondly, there are a number of studies that exploit retirement ages, pension windows and reforms as instruments (e.g. Atalay & Barrett, 2014; Bloemen et al., 2013; Bonsang et al., 2012; Coe et al., 2012; Coe & Zamarro, 2011; Eibich, 2015; Hagen, 2016; Hallberg et al., 2015; Hernaes et al., 2013; Hessel, 2016; Johnston & Lee, 2009; Mazzonna & Peracchi, 2012; Mazzonna & Peracchi, 2017; Neuman, 2008 and Rohwedder & Willis, 2010). These studies typically find beneficial effects on health outcomes¹⁰⁹ but harmful effects on cognitive outcomes¹¹⁰. However, the magnitude of these effects differ rather sharply between different studies. For example, while Rohwedder & Willis find that retiring decreases your word recall score by a substantial amount (roughly 1.35 standard deviations), Coe et al. find that your retirement duration has little or positive effects on several cognitive outcome measures (including word recall).

There are several possible explanations for the variety in estimated effects, including differences in sampling frame, outcome measures and modeling assumptions. For example, it's possible that effects vary between countries due to differences in norms, labor market conditions or social security systems. The results may also be sensitive to the choice of outcome

¹⁰⁸ Some of these studies have attempted to correct for the endogeneity of the retirement decision. For example, Bonsang et al. (2012) use the eligibility age for social security as instrument (included in the review by Meng et al.).

¹⁰⁹ Of these 15 studies, 11 had some measure of physical health (self-reported or objectively measured) as outcome. Of these, seven found a positive effect of retirement; four found insignificant effects and one found a negative effect (one study has been double-counted as having found both a positive and an insignificant effect). Four studies with *mortality* as outcome have also been included (Bloemen et al., 2013; Hagen, 2016; Hallberg et al., 2015 and Hernaes et al., 2013) where two studies found that retirement (or early retirement) decreases the risk of dying and two studies found no significant effects.

¹¹⁰ Of these 15 studies, six had some measure of cognitive functioning as outcome. Of these, four found significant negative effects and one found a significant positive effect (although only among blue-collar workers); one finding was insignificant.

measure; while research on self-reported health tend to show beneficial effects of retirement, the results are more mixed with regard to objective health measures such as mortality. Modeling assumptions may also be of importance. Bingley & Martinello (2013) point out that several cross-country studies that use retirement ages as instrument fail to control for schooling, and argue that this explains a large part of the negative cognitive effects that have been found in the literature. I find that the parametric assumptions underlying the specification of the age effect is of more importance for these data, however. Specifically, it is of importance to allow for different age-trends from one country to the next. Also, different pictures may emerge depending on how the retirement-effect is modeled. While some studies model retirement as having a constant (added) effect to the outcome, others allow for this effect to depend on the retirement duration. Here I contrast both strategies.

5.3 DATA AND DESCRIPTIVE STATISTICS

I use PIAAC survey data that has been collected and compiled by OECD. The subsample I exploit covers roughly 40,000 individuals aged 50-65 years in 25 countries¹¹¹ and has mainly been collected during 2010-2014 using house interviews. The variables of special interest are those describing the respondents' cognitive abilities and subjective health. For the cognitive outcomes, I use two different measures – literacy and numeracy – designed to capture your ability to interpret text-based and mathematical information (see chapter 1 for a description). In the following section, I describe the other key variables including health and retirement.

5.3.1 Variable description: Health and retirement

Self-reported health

Self-reported health is measured on a 5-point scale (1 = Poor, 2 = Fair, 3 = Good, 4 = Very good, 5 = Excellent). It's not clear how one should appropriately interpret such a measure, as health isn't fully observable to individuals themselves. Part of the variation in perceived health is therefore likely due to indicators for health, such as lifestyle and overall wellbeing. But even if health were fully observable to individuals themselves, assigning a number to your health status is not straightforward, making this outcome susceptible to measurement errors.¹¹² As suggested in the literature, it's also possible that self-reported health suffers from a 'justification bias' (see, for example, the discussion by Dwyer & Mitchell, 1999). Similarly, self-reported health could also be sensitive to a placebo effect; personal beliefs regarding the effect of retirement may influence your rating.

¹¹¹ Belgium, Chile, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Greece, Ireland, Israel, Italy, Japan, Korea, Lithuania, the Netherlands, Norway, Poland, Russia, Slovakia, Slovenia, Spain, Sweden, Turkey and the United Kingdom. In practice, neither Ireland nor Norway contribute, however, since their retirement ages are falling outside the sample age frame. For men, Israel is not either contributing for the same reason.

¹¹² For example, women have significantly poorer self-reported health than men (the gap is 0.1 points when keeping age constant). From this, I would not dare to conclude that women have worse health than men, nor that they experience their health being poorer, only that they rate their health being poorer, i.e. it is possible that the gender gap is fully explainable by a gender difference in rating norms. The existence of rating norms can be viewed as a measurement error (if this norm varies from one person to the next). In practice, this will only be a problem here if retiring affects your rating norm.

Retirement

I define an individual as being retired if she is outside the labor force, i.e. neither working, looking for work, nor waiting to start a job. Here, you are a *worker* if you work for at least 10 hours per week (this includes unpaid work for a family business) or if you have a job but the working hours are unknown. You are *looking for work* if you were looking for paid work at any time within the last four weeks, or if you are waiting for a reply on a job application or for help from a training agency. You are *waiting to start a job* if you are waiting to start a job for which you've already been hired, or if you are temporarily absent from a job that you plan on returning to.¹¹³

I also consider an alternative way of measuring retirement using self-reported status.

5.3.2 Sample selection and descriptive statistics

The focus of this study is on the effect of retiring on cognitive outcomes and self-reported health. To this end, I make the following restrictions on the sample: Each individual included was aged 50 to 65 at the time of the survey. I exclude individuals with missing values on the outcome variable of interest, age, gender or labor market status. I include individuals with missing values on other covariates, and code these accordingly. This leaves me with a sample of 43,978 individuals (for literacy and numeracy) and 42,817 (for self-reported health). Self-reported health is unknown for individuals from Turkey, which is the main difference between the samples.

Table 1 presents weighted means for the outcome variables and background characteristics separately for retirees and those in the labor force ('active'). Retirees score significantly below those in the labor force: The raw gaps in literacy and numeracy are 0.38 and 0.45 standard deviations¹¹⁴; the gap in self-reported health is 0.55 points (on a 5-point scale). The groups also differ in several other ways, however. Most importantly, retirees are on average older, less educated and their parents are also less educated. Family life also differs between the groups; not surprisingly, retirees are overrepresented among those having a partner who is retired.

¹¹³ I allow for missing data to some extent: If it is known that an individual does not work, but it is unknown whether she is looking for work or waiting to start a job, then I treat that individual as retired (4.5 percent of retirees fall into this category; of those in the labor force, 93 percent are working and the remaining 7 percent are looking for work or waiting to start a job).

¹¹⁴ Here, one standard deviation is measured using within-country variation only and the full PIAAC sample of 16-65 year olds.

Table 1. Weighted means for the outcomes and background variables, separately for retirees and those in the labor force ('active')

	All	Retirees	Active
Literacy (z-score) ^a	-0.32	-0.55	-0.17
Numeracy (z-score) ^a	-0.28	-0.55	-0.098
Self-reported health (1-5) ^b	2.89	2.55	3.10
Female	0.51	0.62	0.45
Age	57.2	59.3	55.8
Immigrant	0.087	0.078	0.092
Schooling (years) ^c	11.6	10.5	12.3
<i>Mother's education^d</i>			
Low	0.77	0.82	0.73
Medium	0.15	0.12	0.18
High	0.048	0.025	0.063
<i>Father's education^d</i>			
Low	0.66	0.71	0.62
Medium	0.22	0.19	0.23
High	0.087	0.057	0.11
<i>Children</i>			
Have children	0.89	0.89	0.90
Don't have children	0.11	0.11	0.10
Number of children (0-25) ^c	2.10	2.09	2.11
<i>Partner</i>			
Have partner	0.75	0.73	0.76
Don't have partner	0.11	0.11	0.11
<i>If have partner:</i>			
Partner working fulltime	0.44	0.28	0.55
Partner working halftime	0.10	0.061	0.13
Partner retired	0.24	0.43	0.12
None of the above or missing	0.22	0.23	0.21
Observations	43,978	17,477	26,501

Notes: ^aThe sample individuals score roughly 0.3 standard deviations below average, meaning that they score 0.3 standard deviations below a randomly chosen individual from the PIAAC target population which consists of everyone aged 16-65 from the 31 countries included in the first two rounds of PIAAC. The standard deviations for literacy and numeracy are estimated using within-country variation only. ^bSelf-reported health is known for 42,817 individuals in total. ^cSchooling is measured for those with non-missing values only which constitute 98.5 percent of the sample. Similarly, number of children is non-missing for 88.5 percent of the sample. There are missing values on other variables also. For example: The percentages on 'Mother's education' sum up to 97 meaning that the remaining 3 percent are missing. ^dEducational degree: Low = Lower secondary school, primary school or less. Medium = Upper secondary and post-secondary non-tertiary school. High = Tertiary degrees. The averages are weighted as to account for the country specific survey designs; on a cross-country level all countries are weighted equally according the sample "All" (43,978 obs.)

In the Appendix (A1. Adjusting for observable heterogeneity) I present the adjusted gaps in literacy, numeracy and self-reported health between retirees and those in the labor force, controlling for the background characteristics listed in table 1, as well as country of residence. The gaps in the cognitive outcomes are now rather modest at 0.13 and 0.14 standard deviations on literacy and numeracy, respectively. The gap in self-reported health, on the other hand, is mainly unaffected by adjusting for these covariates. Furthermore, the negative effects on literacy and numeracy disappear altogether when I identify retirees by their self-reported status.¹¹⁵ These estimates suggest that there are no large harmful cognitive effects of retiring, assuming that any bias would inflate the gaps between the groups if anything.

5.3.3 Country-level analysis

In this section, I use country-level data and measure the increase in retirements over the ages of 50-55 and 60-65, and how these increases relate to changes in the outcomes. The main advantage of this exercise is that the variation in retirements over age groups (and between countries) is presumably more exogenous than the variation between individuals. For example, I observe that the fraction of retirees increases by 15 percentage points in Chile while the corresponding number in Slovenia is 70 percentage points. This difference is presumably largely driven by institutional differences between the countries, and less influenced by personal health-related choices (although institutional rules can also be endogenous). With that caveat in mind, I find no support for to the notion that retiring would have negative consequences for cognitive functioning, at least as measured by literacy and numeracy. Figures 1a-1c illustrate these relationships graphically. Countries where the increase in retirements is large experience smaller declines in literacy and numeracy on average, but this correlation is only marginally significant for literacy. In some samples, however, these correlations are more pronounced: The positive correlations with literacy and numeracy are significant when retirees are identified using self-reported status, or when limiting the sample to women only.

¹¹⁵ Individuals who self-identify as retirees have 0.0047 standard deviations worse performance on *literacy* and 0.052 standard deviations better performance on *numeracy* as compared to all others (controlling for the background characteristics listed in table 1 and country of residence). When restricting the reference group to self-identified workers, the corresponding gaps are 0.050 and 0.013 standard deviations favoring workers. Here, I also restrict the sample to individuals who have had a paid job at some point in their life. For more detail, see the Appendix (A.1).

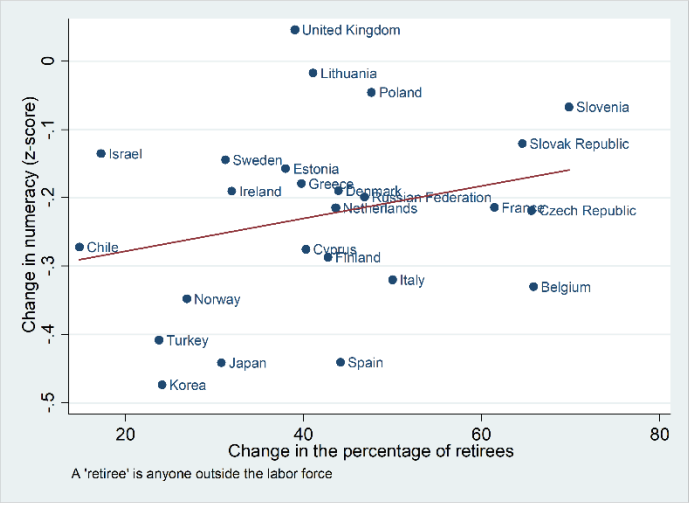


Figure 1a. Retirements and numeracy, ages 50-55 to 60-65

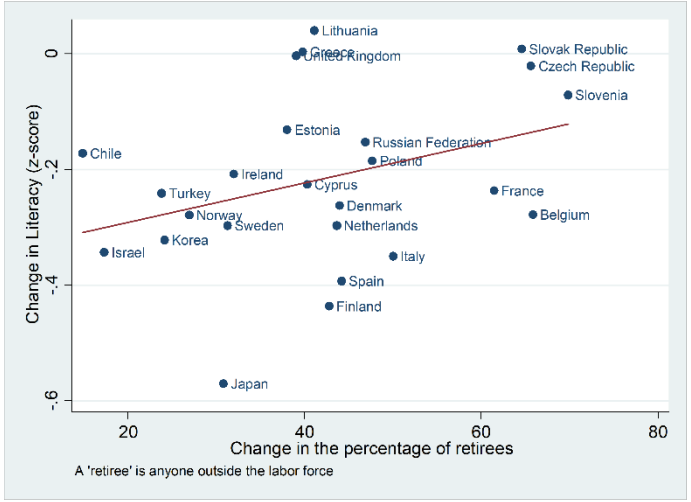


Figure 1b. Retirements and literacy, ages 50-55 to 60-65

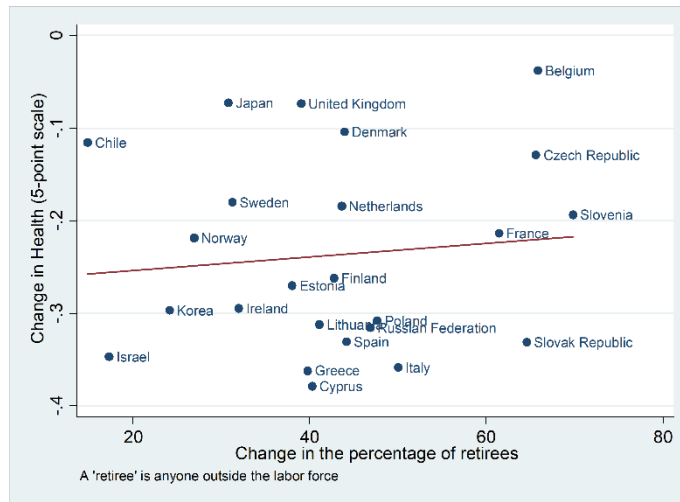


Figure 1c. Retirements and health, ages 50-55 to 60-65

5.4 INSTRUMENTING RETIREMENT AND RETIREMENT DURATION

In this section, I instrument *retirement* as well as *years as retired* by exploiting the variation in incentives created by country- and gender-specific retirement ages. I find no significant discontinuity in literacy or numeracy as individuals reach the retirement age. However, I do find a significant trend-break, suggesting that retiring slows down the age-related decline in literacy and numeracy among men. For women, there is evidence of a positive effect on health.

5.4.1 The models

The instrument is a dummy taking on the value 1 for individuals who have reached the retirement age and the value 0 for others. Several countries have both a ‘normal’ and ‘early’ retirement age; in these cases, I use both.¹¹⁶ In general, the fraction of retirees are higher in the years following the retirement date, but there is no clear discontinuity; retirees are common also in the years leading up to that date. Figure 2 below illustrates the proportion of retirees and how it increases with age, together with the trends in the fraction of individuals who have reached the early and normal retirement ages.

¹¹⁶ Retirement ages by country (normal/early): Belgium (65/60), Chilean males (65/-), Chilean females (60/-), Cyprus (65/63), Czechian males (62.33/60), Czechian females (59/-), Denmark (65/-), Estonian males (63/60), Estonian females (61.5/58.5), Finland (65/62), France (61.17/60), Greece (62/-), Ireland (66/-), Israeli males (67/-), Israeli females (62/-), Italy (60/-), Japan (65/60), Korea (60/55), Lithuanian males (63/-), Lithuanian females (61/-), Netherlands (65/-), Norway (67/62), Polish males (65/-), Polish females (60/-), Russian males (60/-), Russian females (55/-), Slovakian males (62/-), Slovakian females (58.75/-), Slovenian males (58.67/-), Slovenian females (58.33/-), Spain (65/61), Sweden (65/61), Turkish males (60/-), Turkish females (58/-), British males (65/-), British females (60.75/-). Main source: OECD, Pensions at a glance 2015 & 2013.

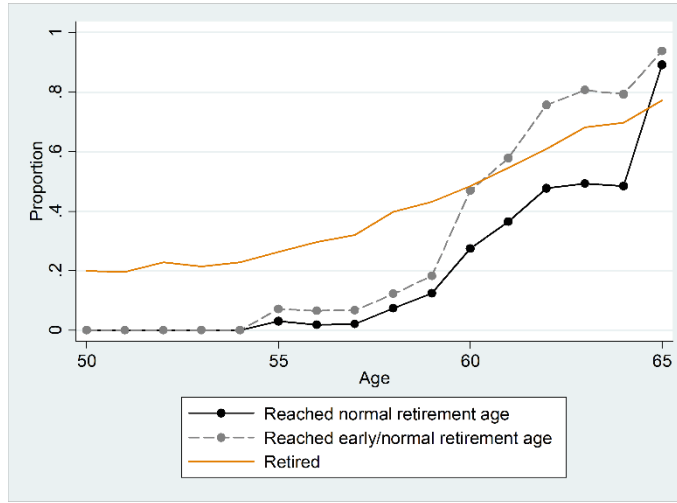


Figure 2. Retirement ages and actual retirements

‘Having reached the retirement age’ is a valid instrument for being retired if any differences in the outcomes between individuals who have reached the retirement age and those who have yet to reach it, are fully explainable by a difference in the probability of actually being retired. Naturally, age is one such distinguishing factor, but I adjust for age as well as other background characteristics. I specify the *reduced form equation* in the following way:

$$y_{ci} = \alpha_c + \beta_1 normal_{ci} + \beta_2 early_{ci} + \vartheta X_{ci} + \varepsilon_{ci} \quad (1)$$

where c is a country index and i is indexing individuals. The outcome, y_{ci} , is either literacy, numeracy or subjective health; α_c represents country-specific fixed intercepts; *normal* and *early* are dummies for those who have reached the normal and early retirement ages; X_{ci} is a vector of background characteristics including age, gender, immigration status, years of schooling, as well as the mother’s and father’s education levels. Here I allow for the effects of age and gender to differ between countries; age is included linearly.¹¹⁷ (See the Appendix, section A2, for graphical presentations of the age trends.)

Now, subjective health is a latent variable, only observed on a collapsed ordinal scale. I account for this by applying adjusted-POLS – an alternative

¹¹⁷ An alternative approach would be to use *common* age trends (for example in the form of age dummies). However, the age-trends differ between countries in a way that makes such an approach susceptible to bias. Country-specific polynomial age trends would be a more pertinent alternative in this regard, but adds a large penalty in terms of precision. The result-section (5.4.2) discusses the chosen specification and its robustness in more detail.

linear estimator to ordered probit. This implies that the coefficients are measured in units of standard deviations. Example: An effect size at 0.1 is an estimated increase by 0.1 standard deviations on the latent variable.¹¹⁸

Note that this model assumes that retiring creates a discontinuity in the age-trend if anything. An alternative possibility is that retiring creates a trend-break, so that the age-related decline either accelerates or slows down after retiring. In an alternative model, I allow for this possibility by instrumenting your *retirement duration* using the *number of years since you reached the normal retirement age*. The retirement duration is a variable that takes on the value 0.5 for those who left their last paid job within the last 12 months, the value 1 for those who left that job last year, and increases by 1 for each year thereafter. Those who are currently working have a value of 0. Here I further limit the sample to individuals who were still working at the age of 50 (or who left their last paid job at the age of 50). All individuals who are currently working are included.

The reduced-form equation can now be described:

$$y_{ci} = \alpha_c + \beta(normal_{ci} \times yearssince_{ci}) + \vartheta X_{ci} + \varepsilon_{ci} \quad (2)$$

where $normal \times yearssince$ is a variable that increases by 1 for each additional year that passes since the normal retirement age; an individual at the retirement age gets a value of 0.5; an individual below that age gets the value 0. Example: An individual who is 65 and who lives in a country where the retirement age is 62 years and 4 months gets a value of 3.17 ($65.5 - 62.33 = 3.17$). In some countries, the retirement age has increased over time and I account for this when calculating the number of years since individuals from different cohorts reached the retirement age. X_{ci} represents the same set of covariates as is equation (1); α_c and ε_{ci} are the country-specific fixed intercept and error term.

All estimates are weighted as to account for the country specific survey designs; on a cross-country level, all countries are weighted equally. Standard errors are estimated using jackknife replicate sampling weights. In the reduced-form and instrumental variables regressions I further correct the standard errors for the imputation variance added by using plausible values.

¹¹⁸ Adjusted-POLS works through quantifying the distances between the observed values on the ordinal variable. This requires an assumption regarding the distribution of the latent variable. Here, I assume that the latent variable – subjective health – is normally distributed. For each individual, I calculate the expected ‘standardized health score’ conditional on the observed score (using the relevant sample as reference). The new quantified variable is then divided by its variance, which adjusts for the ‘attenuation bias’ (see chapter 2; Van Praag & Ferrer-i-Carbonell, 2006).

5.4.2 Results when instrumenting *retirement*

In this section, I estimate the effect of being retired under the assumption that retiring creates a discontinuity in the age-trend if anything (see model (1) from section 4.1). Table 2 presents the first stage estimates. Reaching the ‘early’ retirement age increases your chance of being retired by 16 percentage points; reaching the ‘normal’ retirement age adds another 15 percentage points to that probability. The instruments are strong, with F-values ($df = 2, 78$) at 258.3 (overall) and at 110.7 for women and 117.1 for men. Using the slightly smaller sample – where self-reported health is known – gives similar results (not reported in the table).¹¹⁹

Table 2. First stage estimates, overall and by gender. Outcome: Retired (0/1)

	ALL	WOMEN	MEN
Reached normal ret. age	0.15*** (0.0099)	0.14*** (0.012)	0.16*** (0.015)
Reached early ret. age	0.16*** (0.013)	0.17*** (0.019)	0.17*** (0.016)
Observations	43,978	23,763	20,215
R-squared	0.283	0.277	0.256

Notes: Reached normal ret. age and Reached early ret. age are dummies for those who have reached the normal and early retirement ages, respectively. All regressions include country fixed effects and country-specific linear age trends, immigration status, years of schooling, mother’s and father’s education (three levels + unknown). The first regression (“ALL”) also includes a gender effect which is allowed to vary between countries. All regressions are weighted as to account for the country specific survey designs; on a cross-country level all countries are weighted equally. Standard errors (in parenthesis) are estimated using jackknife replicate sampling weights. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Although retirement ages have clear effects on your chance of being retired, I find little evidence of such effects on literacy or numeracy.¹²⁰ Table 3 presents the estimates from the instrumental variables regressions. Overall, retiring is estimated to increase your literacy and numeracy scores by 11 and 7 percent of a standard deviation, respectively. These effects are similar for men and women. None of the effects is significant, however.

¹¹⁹ For the smaller sample: Reaching the early retirement age increases your chance of being retired by 16 percentage points; reaching the normal retirement age adds 16 percentage points to that probability ($F = 247.8$). For women, the corresponding F-value is 100.4 and for men, 115.9.

¹²⁰ Reaching the early and normal retirement ages are estimated to increase your *literacy* score by 0.020 (SE 0.027) and 0.015 (SE 0.024) standard deviations. For *numeracy*, the corresponding effects are 0.013 (SE 0.024) and 0.0099 (SE 0.023) standard deviations.

It is worth noting that the error margins are rather wide: for both literacy and numeracy, the confidence interval (95 %) includes anything between modest harmful effects to meaningful benefits.¹²¹ Identifying retirees by their self-reported status produces similar estimates, but error margins are now smaller, backing up the conclusion that there are no important harmful cognitive effects of retiring.¹²²

I do, however, find a significant discontinuity in 'health': reaching the early and normal retirement ages are estimated to increase your health score by 0.053 and 0.068 standard deviations, respectively. This translates into a retirement-effect at 42 percent of a standard deviation. This effect is larger for women: While I find no significant effect in the male sample, I find a clear discontinuity in the female sample, suggesting that retiring increases your health score by nearly 60 percent of a standard deviation – a large effect indeed. Identifying retirees by their self-reported status, however, produces more conservative estimates.¹²³

As an alternative to adjusted-POLS, I also use a binary coding for health. This produces similar results, but the effect is now of equal magnitude for men and women: retiring is estimated to increase your chance of having a very good or excellent health by 19 percentage points overall (SE 4.6); by 16 percentage points for women (SE 6.8) and by 18 percentage points for men (SE 7.2).

¹²¹ 95 % CI (Literacy): -0.11 – 0.33. 95 % CI (Numeracy): -0.16 – 0.30.

¹²² When identifying retirees by their self-reported status, the effect on *literacy* decreases from 0.11 (SE 0.11) to 0.070 (SE 0.074); the effect on *numeracy* decreases from 0.072 (SE 0.11) to 0.045 (SE 0.075).

¹²³ When identifying retirees by their self-reported status, the effect on *health* decreases from 0.42*** (SE 0.12) to 0.25*** (SE 0.074). For women, the effect on *health* is now 0.39*** (SE 0.098).

Table 3. The effect of retiring on literacy, numeracy and subjective health (IV-estimates)

ALL	Literacy (z-score)	Numeracy (z-score)	Health (adjusted POLS)
Retired	0.11 (0.11)	0.072 (0.11)	0.42*** (0.12)
Observations	43,978	43,978	42,817
WOMEN	Literacy (z-score)	Numeracy (z-score)	Health (adjusted POLS)
Retired	0.14 (0.17)	0.13 (0.15)	0.57*** (0.18)
Observations	23,763	23,763	23,185
MEN	Literacy (z-score)	Numeracy (z-score)	Health (adjusted POLS)
Retired	0.085 (0.15)	0.073 (0.15)	0.22 (0.17)
Observations	20,215	20,215	19,632

Notes: All regressions include country fixed effects and country-specific linear age trends or country-specific quadratic age trends (for 'Health'); immigration status, years of schooling, mother's and father's education (three levels + unknown). The first regression ("ALL") also includes a gender effect which is allowed to vary between countries. For 'Health', I apply the adjusted-POLS estimator, implying that the unit of measurement is one standard deviation. All regressions are weighted as to account for the country specific survey designs; on a cross-country level all countries are weighted equally. Standard errors (in parenthesis) are estimated using jackknife replicate sampling weights and further adjusted for the use of plausible values (for literacy and numeracy). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Hence, I find no evidence of a discontinuity in cognitive performance at retirement. I do, however, find a positive health-effect for women. These conclusions are robust: the effect on literacy and numeracy are rather small and insignificant independently of specification and sample selection criteria (described in more detail below). For health, the estimate fluctuates more, varying between 0.092 and 0.42 standard deviations. The health-effect is stable for women, however, varying between 0.49 and 0.61 standard deviations.

Firstly, the estimates do not change much when I extend the number of controls, including also educational area¹²⁴, educational degree¹²⁵, having children, and if so, their numbers; having a partner, and if so, his or her labor market status. Here, I further allow for different age trends depending on educational degree and gender.¹²⁶ Likewise, the estimates are practically unaffected by dropping all controls except for country, age and gender (where the effects of age and gender are allowed to vary between countries).¹²⁷

Of special concern for the identification, however, is the possibility that the effect of aging is not captured satisfactorily by the country-specific linear trends. I test for this possibility in the following ways. First, I create fake retirement ages by moving the real retirement ages five years into the past, and estimate the reduced form equations on an otherwise identical sample of 45-60 year olds (excluding only those who already reached a real retirement age by this time). I find no significant effects of these fake instruments.¹²⁸ Secondly, I reduce the age-frame of the data to individuals with no more than seven years to or from a retirement age¹²⁹, and experiment with

¹²⁴ The educational areas include: General programmes; Teacher training and education science; Humanities, languages and arts; Social sciences, business and law; Science, mathematics and computing; Engineering, manufacturing and construction; Agriculture and veterinary; Health and welfare; Services; None of these areas (i.e. no education/low education or missing).

¹²⁵ The educational degrees include: Primary or less; Lower secondary; Upper secondary; Post-secondary, non-tertiary; Tertiary – professional degree; Tertiary – bachelor degree; Tertiary – master/research degree; Tertiary - bachelor/master/research degree without distinction; Missing. The former control variable – years of schooling – is excluded in this specification.

¹²⁶ The effect (IV) on *literacy* decreases from 0.11 (SE 0.11) to 0.081 (SE 0.12); the effect on *numeracy* decreases from 0.072 (SE 0.11) to 0.032 (SE 0.12) and the effect on *health* decreases from 0.42*** (SE 0.12) to 0.41*** (SE 0.12). For women, the effect on *health* decreases from 0.57*** (SE 0.18) to 0.49*** (SE 0.16).

¹²⁷ The effect (IV) on *literacy* decreases from 0.11 (SE 0.11) to 0.10 (SE 0.12); the effect of *numeracy* decreases from 0.072 (SE 0.11) to 0.059 (SE 0.12); the effect on *health* decreases from 0.42*** (SE 0.12) to 0.39*** (SE 0.12). For women, the effect on *health* decreases from 0.57*** (SE 0.18) to 0.51*** (SE 0.17).

¹²⁸ Reaching the *fake* early and normal retirement ages are estimated to change your *literacy* score by -0.028 (SE 0.028) and 0.023 (SE 0.027) standard deviations (jointly insignificant). For *numeracy*, the corresponding numbers are -0.037 (SE 0.029) and 0.025 (SE 0.026) standard deviations (jointly insignificant). For *health*, the *real* early and normal retirement ages are estimated to change your score by 0.053** (SE 0.025) and 0.068*** (SE 0.025) standard deviations; for the *fake* dates, the corresponding numbers are -0.015 (SE 0.030) and 0.0060 (SE 0.025) standard deviations (jointly insignificant). For women: reaching the *real* early and normal retirement dates are estimates to change your *health* score by 0.017 (SE 0.035) and 0.12*** (SE 0.027) standard deviations; for the *fake* dates, the corresponding numbers are 0.044 (SE 0.036) and 0.0083 (SE 0.030) standard deviations (jointly insignificant).

¹²⁹ The effect (IV) on *literacy* decreases from 0.11 (SE 0.11) to -0.046 (SE 0.23); the effect on *numeracy* decreases from 0.072 (SE 0.11) to 0.022 (SE 0.20); the effect on *health* decreases from 0.42*** (SE 0.12) to 0.26 (SE 0.20). For women, this effect is still large and significant at 0.57**

country-specific polynomial age-trends of the second degree¹³⁰. For both of these alternatives, estimates are reduced, and especially so for health. The positive health-effect for women remains mainly unchanged, however.

It can be noted, that the estimates are sensitive to the choice between country-specific and *common* age trends, with the latter doing a poor job at capturing the effect of aging. For example, a specification with a common linear age trend would suggest that retiring increases your literacy score by 30 percent of a standard deviation ($p < 0.01$). This shows that countries with early retirement dates observe less of a decline in literacy with age. When accounting for this heterogeneity in trends, then there is no significant discontinuity in cognitive performance related to the date of retirement.

5.4.3 Results when instrumenting *retirement duration*

The previous analysis was built on the assumption that retiring creates a discontinuity in outcomes by shifting the age-trend upwards or downwards (if at all). Here I assume that retiring creates a trend-break; the age-related decline in literacy and numeracy is assumed to either accelerate or slow down after retirement if it changes at all (see model (2) in section 4.1). Table 4 presents the first-stage estimates: For each year that passes since the retirement date, the number of years as retired increases by a half. This effect is of similar magnitude for men and women. The instrument is strong, with F-values ($df = 1, 79$) at 703.69 (overall) and at 452.24 for women and 128.28 for men. Using the smaller sample – where health is known – gives similar results (not reported in the table).¹³¹

(SE 0.26) standard deviations. The instruments are sufficiently strong, with F-values at 86.5 for the cognitive outcomes, and at 83.5 for health (45.1 among women).

¹³⁰ The effect (IV) on *literacy* decreases from 0.11 (SE 0.11) to -0.11 (SE 0.27); the effect on *numeracy* decreases from 0.072 (SE 0.11) to -0.059 (SE 0.26); the effect on *health* decreases from 0.42*** (SE 0.12) to 0.092 (SE 0.24), but increases for women to 0.61* (SE 0.32). The instruments are sufficiently strong, with F-values at 64.6 for the cognitive outcomes, and at 60.8 for health (32.9 among women). As a group, the country-specific quadratic age components are significant for literacy and numeracy, and insignificant for health.

¹³¹ For the health sample: For each year that passes since the retirement date, the number of years as retired increases by 0.49 ($F = 765.6$). For women, the corresponding F-value is 522.2 and for men, 166.8.

Table 4. First-stage regressions, overall and by gender. Outcome: Years as retired

	ALL	WOMEN	MEN
Years since retirement age	0.49*** (0.019)	0.52*** (0.024)	0.50*** (0.044)
Observations	36,996	18,625	18,371
R-squared	0.322	0.357	0.275

Notes: Years since retirement age is a variable that increases by 1 for each year since the normal retirement age; for those at the retirement age the variable takes on the value 0.5 and for those below that age the variable takes on the value 0. All regressions include country fixed effects and country-specific linear age trends; immigration status, years of schooling, mother's and father's education (three levels + unknown). The first regression ("ALL") also includes a gender effect which is allowed to vary between countries. All regressions are weighted as to account for the country specific survey designs; on a cross-country level all countries are weighted equally. Standard errors (in parenthesis) are estimated using jackknife replicate sampling. *** p<0.01, ** p<0.05, * p<0.1

As you age, your literacy score is estimated to drop by 0.014 standard deviations per year until you reach the retirement age.¹³² At that point, this changes to a modest 0.001 standard deviation decrease per year. The pattern is similar for numeracy. Table 5 presents the corresponding estimates from the instrumental variables regressions. Here, literacy is estimated to drop by 0.019 standard deviations per year among workers. As you retire, this changes to a 0.009 standard deviation increase per year. The difference in trends – 0.027 standard deviations – is significant although qualitatively rather small. For numeracy, the corresponding difference is insignificant at 0.011 standard deviations. The effect on health is similarly small at 0.024 standard deviations.

These estimates are somewhat larger for men than women, and especially so for literacy: While I find no significant trend-breaks in the female sample, I find a clear break for men. One year after retirement, men are predicted to score 0.078 standard deviations higher on literacy as compared to the counterfactual outcome. Another year later, the estimated advantage has grown to 0.16 standard deviations. For numeracy, the gender gap in effect sizes is smaller, and for health, there is no gap worth mentioning.

¹³² This is the simple *average* trend over all countries from the reduced-form regression.

Table 5. The effect of each additional year *as retired* on literacy and numeracy (IV-estimates)

ALL	Literacy (z-score)	Numeracy (z-score)	Health (adjusted POLS)
Years as retired	0.027** (0.014)	0.011 (0.012)	0.024* (0.014)
Observations	36,996	36,996	36,483
WOMEN	Literacy (z-score)	Numeracy (z-score)	Health (adjusted POLS)
Years as retired	0.031 (0.026)	0.015 (0.025)	0.032* (0.017)
Observations	18,625	18,625	18,517
MEN	Literacy (z-score)	Numeracy (z-score)	Health (adjusted POLS)
Years as retired	0.078*** (0.028)	0.047* (0.026)	0.044 (0.034)
Observations	18,371	18,371	17,966

Notes: All regressions include country fixed effects and country-specific linear age trends, immigration status, years of schooling, mother's and father's education (three levels + unknown). The first regression ("ALL") also includes a gender effect which is allowed to vary between countries. All regressions are weighted as to account for the country specific survey designs; on a cross-country level all countries are weighted equally. Standard errors (in parenthesis) are estimated using jackknife replicate sampling weights and further adjusted for the use of plausible values. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Hence, I find no evidence for a negative trend-break in literacy, numeracy or health at retirement. I do, however, find a positive cognitive effect for men. This conclusion is robust: Overall, the effect on literacy varies between -0.0021 and 0.060 standard deviations depending on specification and sample selection criteria (described in more detail below); the effect on numeracy varies between -0.00021 and 0.035 standard deviations. For health, the corresponding interval goes between -0.026 and 0.024. In the male sample, effect sizes vary between 0.046 and 0.14 standard deviations for literacy, and between 0.030 and 0.11 for numeracy.

Of most concern for the identification is the possibility that aging has non-linear effects on the outcomes. If, for example, the decline in literacy and numeracy is diminishing with age, then that could potentially also explain the observed effects in the data. I explore this possibility in several ways (similarly to the previous analysis, see section 4.2). First, I create fake retirement ages by moving the real retirement ages five years into the past,

and estimate the reduced form equations on an otherwise identical sample of 45-60 year olds (excluding only individuals who already reached the real retirement age by this time). I find no significant effects of this fake instrument.¹³³ Secondly, I reduce the age-frame of the data to individuals with no more than seven years to or from their retirement age.¹³⁴ This increases the positive effects on literacy and numeracy while decreasing the estimate for health.¹³⁵ Thirdly, I experiment with different variants of nonlinear age specifications. Firstly, I include age using country-specific polynomial trends of the second degree. As a group, the quadratic age components are borderline significant or significant depending on outcome, eliminating any positive effects on literacy and health in the full sample, while leaving the effect on numeracy practically unchanged.¹³⁶ In the male sample, the instrument is now lacking significant explanatory power in the first stage, ruling out any IV-based inference for this group and specification. However, extending the age frame by a couple of years allows me to use country-specific quadratic age trends in the male sample as well, which marginally increases the estimates in this group.¹³⁷ Also, the estimates for men are

¹³³ *Literacy*: The reduced form estimate decreases from 0.014** (SE 0.0067) when using the real instrument to 0.0034 (SE 0.0085) when using the fake instrument. For men, the corresponding numbers are 0.039*** (SE 0.014) for the real instrument and 0.0041 (SE 0.015) for the fake instrument. *Numeracy*: The reduced-form estimate decreases from 0.0055 (SE 0.0061) when using the real instrument to 0.0042 (SE 0.0081) for the fake instrument. For men, the corresponding numbers are 0.023* (SE 0.013) for the real instrument and -0.0024 (SE 0.014) for the fake instrument. *Health*: The reduced-form estimate increases from 0.012* (SE 0.0066) when using the real instrument to 0.015* (0.0075) for the fake instrument.

¹³⁴ For those who have not yet reached the retirement age, I make the cutoffs as if retirement ages are to stay at their current levels in the future.

¹³⁵ Sample restriction at +/- 7 years: The effect (IV) on *literacy* increases from 0.027** (SE 0.014) to 0.060** (SE 0.029) overall, and from 0.078*** (SE 0.028) to 0.14** (SE 0.057) among men. The effect on *numeracy* increases from 0.011 (SE 0.012) to 0.035 (SE 0.028) overall, and from 0.047* (SE 0.026) to 0.11** (SE 0.053) among men. The effect on *health* decreases from 0.024* (SE 0.014) to -0.0029 (0.028).

¹³⁶ The effect on *literacy* decreases from 0.027** (SE 0.014) to -0.0021 (SE 0.028); the effect on *numeracy* increases from 0.011 (SE 0.012) to 0.013 (SE 0.026); the effect on *health* decreases from 0.024* (SE 0.014) to -0.026 (0.032). The instrument is sufficiently strong for this specification (with $F > 100$ in the first stage for both samples).

¹³⁷ The sample includes men at the age of 45-65: With country-specific linear age trends, the effect on *literacy* (IV) is 0.061*** (SE 0.018) and on *numeracy* 0.041** (SE 0.018). These effects increase to 0.072 (SE 0.061) and 0.042 (SE 0.054) when turning to country-specific quadratic age trends. The sample includes men at the age of 40-65: With country-specific linear age trends, the effect on *literacy* (IV) is 0.046*** (0.014) and on *numeracy* 0.030** (0.014). These effects increase to 0.13*** (SE 0.048) and 0.094** (SE 0.045) when turning to country-specific quadratic age trends.

insensitive to using common age dummies or common quadratic age-trends.¹³⁸

Furthermore, I experiment with including the instrument using a second-degree polynomial as proposed by Dieterle & Snell (2016) as a test of validity. This leaves the estimates mainly unchanged.¹³⁹ Lastly, the estimates don't change much when I extend¹⁴⁰ or limit¹⁴¹ the set of controls.

5.4.3 Heterogeneity

It is reasonable to assume that the effects of retiring could depend on your prior career. For example, I would expect any positive effects to be larger among individuals retiring from poor work environments. Prior work conditions are not observable for retirees, however, but I do observe education. I find no evidence of heterogeneity in effects depending on education, however.¹⁴²

In the Appendix (A.3 Country-specific estimates) I present the first-stage and reduced-form regressions separately by country; table A2a and A2b use

¹³⁸ With common quadratic age-trends (males only): The effect (IV) on *literacy* increases from 0.078*** (SE 0.028) to 0.092*** (SE 0.025); the effect on *numeracy* decreases from 0.047* (SE 0.026) to 0.042* (SE 0.025). With common age-fixed effects (males only): The effect on *literacy* increases to 0.097*** (SE 0.025) and the effect on *numeracy* decreases to 0.044* (SE 0.024).

¹³⁹ Overall: The effect (IV) on *literacy* decreases from 0.027** (SE 0.014) to 0.025* (0.013); the effect on *numeracy* decreases from 0.011 (SE 0.012) to 0.0088 (SE 0.012). For *health*, the effect decreases from 0.024* (SE 0.014) to 0.016 (0.013). Men only: The effect (IV) on *literacy* and *numeracy* remain unchanged at 0.078*** (SE 0.028) and 0.047* (SE 0.026), respectively.

¹⁴⁰ Extended set of controls (includes also educational area and degree, having children, and if so, there numbers, having a partner and if so, his or her labor market status, as well as interactions between age and gender and between age and the educational degrees): Overall, the effect (IV) on *literacy* decreases from 0.027** (SE 0.014) to 0.026 (SE 0.016); the effect on *numeracy* decreases from 0.011 (SE 0.012) to 0.0069 (SE 0.014) and the effect on *health* decreases from 0.024* (SE 0.014) to 0.014 (SE 0.015). For men, the effect on *literacy* increases from 0.078*** (SE 0.028) to 0.082*** (SE 0.031); the effect on *numeracy* remains unchanged.

¹⁴¹ Limited set of controls (excludes each control except for country, age and gender, where the effects of age and gender is allowed to vary by country): Overall, the effect (IV) on *literacy* decreases from 0.027** (SE 0.014) to 0.017 (SE 0.014); the effect on *numeracy* decreases from 0.011 (SE 0.012) to -0.00021 (SE 0.014) and the effect on *health* decreases from 0.024* (SE 0.014) to 0.017 (SE 0.014). For men, the effect on *literacy* increases from 0.078*** (SE 0.028) to 0.088*** (SE 0.030); the effect on *numeracy* increases from 0.047* (SE 0.026) to 0.057** (SE 0.028).

¹⁴² By including interactions, I test whether effects differ between those holding a tertiary degree and others. I find no significant heterogeneities in effects for *literacy* or *health* when instrumenting your *retirement duration*, nor for *health* when instrumenting *retirement*. In these models, I also include interactions between holding a tertiary degree and several other covariates. Most importantly, I allow for different age trends for those with tertiary degrees and those without, and I also allow for this heterogeneity in age-trends to depend on country.

the full sample; table A2c uses the male sample. In these regressions, the variable of interest is your *retirement duration* (model (2)). The average first-stage and reduced-form estimates match quite well with the corresponding estimates from the pooled sample¹⁴³, and no single country is highly influential on its own.¹⁴⁴ (Here, the *average estimate* is calculated using inverse variance weighting.) Overall, I find no significant trend-breaks in literacy or numeracy for any country, after adjusting for multiple testing using the Bonferroni correction (this is also true for the male sample).¹⁴⁵ For health, Lithuania produces the only significant effect.

5.4.4 Validity and limitations

I find no significant discontinuity in literacy or numeracy as individuals reach the retirement age. I do, however, find a significant trend-break for men, suggesting that retiring slows down the age-related decline in cognitive functioning in this group. Also, there is evidence for a positive discontinuity in health for women. The instruments are valid, if there are no relevant differences between individuals who reached the retirement age years ago and those who reached it recently (or have yet to reach it) other than retirement alone. The fact that the estimates are insensitive to including a larger set of controls can be viewed as support for this assumption. Also, retirement ages differ from one country to the next (as well as between men and women) weakening the connection between specific cohorts and the instruments. Naturally, age is strongly correlated with the instruments in any case. Hence, the validity of the result depends critically on that the chosen specification is a sufficiently good approximation. I use country-specific linear age-trends, and find little evidence for the notion that nonlinearities can explain away the above mentioned patterns in these data.

Even if the estimates are internally valid, there are limitations to the external generalizability of the results. Firstly, the instrumental variables

¹⁴³ Overall, the effect (IV) on *literacy* decreases from 0.027** (SE 0.014) to 0.017 standard deviations when averaging the country-specific estimates using inverse-variance weighting; for *numeracy*, the corresponding effect decreases from 0.011 (SE 0.012) to 0.0030; for *health*, the effect increases from 0.024* (SE 0.014) to 0.029. For men, the effect on *literacy* decreases from 0.078*** (SE 0.028) to 0.064 standard deviations; the effect on *numeracy* decreases from 0.047* (SE 0.026) to 0.040.

¹⁴⁴ The IV-estimates for *literacy* varies between 0.0087 and 0.023 standard deviations as one country at a time is excluded (these effects are based on the country-specific regressions). For *numeracy*, the corresponding interval goes between -0.0045 and 0.011 standard deviations. For *health*, the interval goes between 0.019 and 0.038 standard deviations. For men, the intervals for *literacy* and *numeracy* are 0.053-0.077 and 0.024-0.055, respectively.

¹⁴⁵ Significance at the 5-percent level when $p < 0.0022$ or when $p < 0.0023$ depending on sample.

estimates are only identified by those affected by the instrument (Imbens & Angrist, 1994). Individuals who don't retire, or retire for unrelated reasons, don't contribute. From a policy perspective, this isn't necessarily a disadvantage; individuals who are affected by retirement ages are also the ones of most policy-concern. On the other hand, it's worth noting that I can only estimate the effects given the current systems, where the average normal retirement age is 62 years (63 among men).¹⁴⁶ Propose that effective retirement ages were to increase by two years. This model predicts that men would experience a long-run loss in literacy by 0.16 standard deviations. This estimate, however, is built on the assumption that the effect of your retirement duration is independent of your age at retirement. If, on the other hand, the advantage of being retired increases with your age at retirement, then the long-run effects could be substantially smaller.

¹⁴⁶ Note, however, that the estimates are influenced more heavily by countries with early retirement ages.

5.5 CONCLUSIONS

In this study, I estimate the effect of retiring on cognitive abilities and subjective health. The identification relies on instrumenting *retirement* as well as *years as retired* by exploiting the variation in incentives created by country-specific retirement ages. I find no evidence of a discontinuity in literacy or numeracy at retirement. However, I do find a significant trend-break for men, suggesting that retiring causes a cognitive performance advantage in literacy that grows by 0.02-0.13 standard deviations for each additional year as retired. This finding goes well together with the descriptive data: Countries where a large proportion of the labor force retire (between the ages of 50-55 and 60-65) are, on average, countries with more favorable age-trajectories in cognitive functioning. For women, I find a significant discontinuity in health, suggesting that retirement increases your subjective health score by 0.21-0.93 standard deviations.

These findings don't fit very well with the argument that individuals 'give into' cognitive deterioration as the market incentive is lost at retirement (as suggested by Mazzonna & Pereggi, 2012). These findings fit better with the idea that working life competes with healthy lifestyle activities. Such activities may include exercise, sleep, engagement in social or intellectual activities, or decreased daily stress-levels. I do not have the possibility to investigate these mechanisms using these data. However, a couple of other studies have shown that retirement indeed causes several beneficial lifestyle changes.¹⁴⁷

The findings in this study are in line with much of the recent research on retirement and health. These studies typically find that retirees have poor health as compared to workers, but when the endogeneity of the retirement decision is accounted for, this relationship turns positive. This is also what I observe. On the other hand, my findings don't fit very well with the recent studies looking specifically at cognitive functioning, which typically find a negative effect of retirement. There are several possible explanations for this, including differences in sample frame, model specification and outcome measures. However, it can also be noted that this study is not an outlier per

¹⁴⁷ For example, retiring has been found to cause changes in the form of reduced alcohol consumption, increased walking and heavy exercise, as well as increased sleep among Japanese elderly (Motegi et al., 2016) and decreased risk of 'no physical exercise' among European elderly (Celidoni & Rebba, 2017). The latter study also found retirement to cause increased number of drinking days. Both sleep patterns and exercise have been found to predict cognitive functioning among the elderly (Bherer et al., 2013; Scullin & Bliwise, 2015).

se; rather, the magnitude of effects differ sharply between studies.¹⁴⁸ This points to the importance of further investigation into the relationship, and into the potential causes for heterogeneity in effects.

¹⁴⁸ With Rohwedder & Willis (2010) finding substantial negative effects (roughly equal to -1.35 standard deviations) while, for example, Bonsang et al. (2012) find comparatively modest negative effects. On the other hand, Coe et al. (2012) and Coe & Zamarro (2011) find insignificant effects or significantly positive effects (for blue-collar workers).

References

- Atalay, K., & Barrett, G. F. (2014). The causal effect of retirement on health: New evidence from Australian pension reform. *Economics Letters*, 125(3), 392-395.
- Bherer, L., Erickson, K. I., & Liu-Ambrose, T. (2013). A review of the effects of physical activity and exercise on cognitive and brain functions in older adults. *Journal of Aging Research*, 2013.
- Bingley, P., & Martinello, A. (2013). Mental retirement and schooling. *European Economic Review*, 63, 292-298.
- Bloemen, H., Hochguertel, S., & Zweerink, J. (2013). The causal effect of retirement on mortality: evidence from targeted incentives to retire early.
- Bonsang, E., Adam, S., & Perelman, S. (2012). Does retirement affect cognitive functioning?. *Journal of Health Economics*, 31(3), 490-501.
- Celidoni, M., & Rebba, V. (2017). Healthier lifestyles after retirement in Europe? Evidence from SHARE. *The European Journal of Health Economics*, 18(7), 805-830.
- Coe, N. B., von Gaudecker, H. M., Lindeboom, M., & Maurer, J. (2012). The effect of retirement on cognitive functioning. *Health Economics*, 21(8), 913-927.
- Coe, N. B., & Zamarro, G. (2011). Retirement effects on health in Europe. *Journal of Health Economics*, 30(1), 77-86.
- Dave, D., Rashad, I., & Spasojevic, J. (2006). *The effects of retirement on physical and mental health outcomes* (No. w12123). National Bureau of Economic Research.
- de Grip, A., Dupuy, A., Jolles, J., & van Boxtel, M. (2015). Retirement and cognitive development in the Netherlands: Are the retired really inactive?. *Economics & Human Biology*, 19, 157-169.
- Dieterle, S. G., & Snell, A. (2016). A simple diagnostic to investigate instrument validity and heterogeneous effects when using a single instrument. *Labour Economics*, 42, 76-86.
- Dwyer, D. S., & Mitchell, O. S. (1999). Health problems as determinants of retirement: Are self-rated measures endogenous?. *Journal of Health Economics*, 18(2), 173-193.
- Eibich, P. (2015). Understanding the effect of retirement on health: mechanisms and heterogeneity. *Journal of Health Economics*, 43, 1-12.
- Fisher, G. G., Stachowski, A., Infurna, F. J., Faul, J. D., Grosch, J., & Tetrick, L. E. (2014). Mental work demands, retirement, and longitudinal trajectories of cognitive functioning. *Journal of Occupational Health Psychology*, 19(2), 231.
- Grossman, M. (1972). On the concept of health capital and the demand for health. *Journal of Political Economy*, 80(2), 223-255.
- Hagen, J. (2016). *What are the health effects of postponing retirement? An instrumental variable approach* (No. 2016: 11). Working Paper, IFAU-Institute for Evaluation of Labour Market and Education Policy.

- Hallberg, D., Johansson, P., & Josephson, M. (2015). Is an early retirement offer good for your health? Quasi-experimental evidence from the army. *Journal of Health Economics*, 44, 274-285.
- Hernaes, E., Markussen, S., Piggott, J., & Vestad, O. L. (2013). Does retirement age impact mortality?. *Journal of Health Economics*, 32(3), 586-598.
- Hessel, P. (2016). Does retirement (really) lead to worse health among European men and women across all educational levels?. *Social Science & Medicine*, 151, 19-26.
- Imbens, G., & Angrist, J. (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica*, 62(2), 467-475.
- Johnston, D. W., & Lee, W. S. (2009). Retiring to the good life? The short-term effects of retirement on health. *Economics Letters*, 103(1), 8-11.
- Kelly, M. E., Loughrey, D., Lawlor, B. A., Robertson, I. H., Walsh, C., & Brennan, S. (2014). The impact of cognitive training and mental stimulation on cognitive and everyday functioning of healthy older adults: a systematic review and meta-analysis. *Ageing Research Reviews*, 15, 28-43.
- Krantz-Kent, R., & Stewart, J. (2007). How do older Americans spend their time. *Monthly Lab. Rev.*, 130, 8.
- Mazzonna, F., & Peracchi, F. (2012). Ageing, cognitive abilities and retirement. *European Economic Review*, 56(4), 691-710.
- Mazzonna, F., & Peracchi, F. (2017). Unhealthy retirement?. *Journal of Human Resources*, 52(1), 128-151.
- Mein, G., Martikainen, P., Hemingway, H., Stansfeld, S., & Marmot, M. (2003). Is retirement good or bad for mental and physical health functioning? Whitehall II longitudinal study of civil servants. *Journal of Epidemiology & Community Health*, 57(1), 46-49.
- Meng, A., Nexø, M. A., & Borg, V. (2017). The impact of retirement on age related cognitive decline—a systematic review. *BMC Geriatrics*, 17(1), 160.
- Motegi, H., Nishimura, Y., & Terada, K. (2016). Does Retirement Change Lifestyle Habits?. *The Japanese Economic Review*, 67(2), 169-191.
- Neuman, K. (2008). Quit your job and get healthier? The effect of retirement on health. *Journal of Labor Research*, 29(2), 177-201.
- OECD (2013). *Pensions at a Glance 2013: OECD and G20 Indicators*, OECD Publishing, Paris.
- OECD (2015). *Pensions at a Glance 2015: OECD and G20 indicators*, OECD Publishing, Paris.
- Rohwedder, S., & Willis, R. J. (2010). Mental retirement. *The Journal of Economic Perspectives*, 24(1), 119-138.
- Salthouse, T. A. (2010). Selective review of cognitive aging. *Journal of the International Neuropsychological Society*, 16(5), 754-760.
- Scullin, M. K., & Bliwise, D. L. (2015). Sleep, cognition, and normal aging: integrating a half century of multidisciplinary research. *Perspectives on Psychological Science*, 10(1), 97-137.

- Stock, J. H., & Yogo, M. (2002). Testing for weak instruments in linear IV regression.
- van der Heide, I., van Rijn, R. M., Robroek, S. J., Burdorf, A., & Proper, K. I. (2013). Is retirement good for your health? A systematic review of longitudinal studies. *BMC Public Health*, 13(1), 1180.
- Van Praag, B., & Ferrer-i-Carbonell, A. (2006). An almost integration: Free approach to ordered response models (No. 06-047/3). Tinbergen Institute discussion paper

Appendix

A.1 Adjusting for observable heterogeneity

Table A1 describes the estimated gaps between retirees and those in the labor force, controlling for a range of background characteristics (i.e. country of residence and the background characteristics listed in table 1, where the effects of age and gender are allowed to vary between countries). The regressions reveal that the gaps in literacy and numeracy are comparatively modest after adjusting for differences in observable heterogeneity; the effect on health, on the other hand, is still large.

These estimates are quite sensitive to how retirees are defined. Indeed, individuals who *self-identify* as retirees have equally good or better cognitive performance than others (controlling for the same set of covariates as earlier). The effects on literacy and numeracy are now -0.0047 (SE 0.019) and 0.052*** standard deviations (SE 0.016); the effect on self-reported health is -0.12*** standard deviations (SE 0.020). Note, however, that the reference group – others – consists of both workers as well as unemployed individuals and others outside the labor force (such as housewives). Hence, even if retirement is exogenous in this model, these estimates are not only reflecting the effect of transitioning from *working life* to retirement. More importantly, however, retirement is probably not exogenous. For example, it seems likely that some individuals with little or no work experience do not identify as retirees at any age. Hence, I am potentially comparing former workers to current workers *and* non-workers. However, when I restrict the reference group to self-identified workers only, the effects on literacy and numeracy are still small at -0.050** (SE 0.021) and -0.013 standard deviations (SE 0.020); the effect on self-reported health is -0.32*** standard deviations (SE 0.021). In these regressions, I also exclude anyone who has never had a paid job as to further homogenize the groups. Lastly, I also experiment with restricting the sample by excluding anyone who quit their last paid job before the age of 50 (as all workers are 50+). Now, there are no differences in literacy and numeracy between the retirees and the workers, but still a significant and quite large difference in health.¹⁴⁹

A technical note (table A1): Subjective health is measured using the raw scale (1-5) as well as adjusted-POLS (see section 4.1). Here, the estimates happen to be the same regardless.

¹⁴⁹ The effect on literacy is -0.019 standard deviations (SE 0.024); the effect on numeracy is 0.028 standard deviations (SE 0.022) and the effect on self-reported health is -0.23*** standard deviations (SE 0.022).

Table A1. Adjusted differences in literacy, numeracy and subjective health between retirees and those in the labor force (ages 50-65)

ALL	Literacy (z-score)	Numeracy (z-score)	Health (5-point scale)	Health (adjusted POLS)
Retired	-0.13*** (0.016)	-0.14*** (0.015)	-0.47*** (0.013)	-0.49*** (0.013)
Observations	43,978	43,978	42,817	42,817
R-squared	0.338	0.401	0.217	0.217
WOMEN	Literacy (z-score)	Numeracy (z-score)	Health (5-point scale)	Health (adjusted POLS)
Retired	-0.11*** (0.020)	-0.12*** (0.017)	-0.43*** (0.017)	-0.43*** (0.016)
Observations	23,763	23,763	23,185	23,185
R-squared	0.346	0.410	0.227	0.226
MEN	Literacy (z-score)	Numeracy (z-score)	Health (5-point scale)	Health (adjusted POLS)
Retired	-0.16*** (0.022)	-0.17*** (0.022)	-0.53*** (0.021)	-0.53*** (0.021)
Observations	20,215	20,215	19,632	19,632
R-squared	0.333	0.385	0.212	0.212

Notes: All regressions include country dummies and country-specific linear age trends, immigration status, years of schooling, mother's and father's education, having children and if so, their numbers; having a partner and if so, his or her labor market status. The regressions in the first panel (ALL) also include a gender effect which is allowed to vary between countries. All regressions are weighted as to account for the country specific survey designs; on a cross-country level all countries are weighted equally. Standard errors (in parenthesis) are estimated using jackknife replicate sampling weights and further adjusted for the use of plausible values (for literacy and numeracy). *** p<0.01, ** p<0.05, * p<0.1

A.2 Age-trends in literacy, numeracy and self-reported health

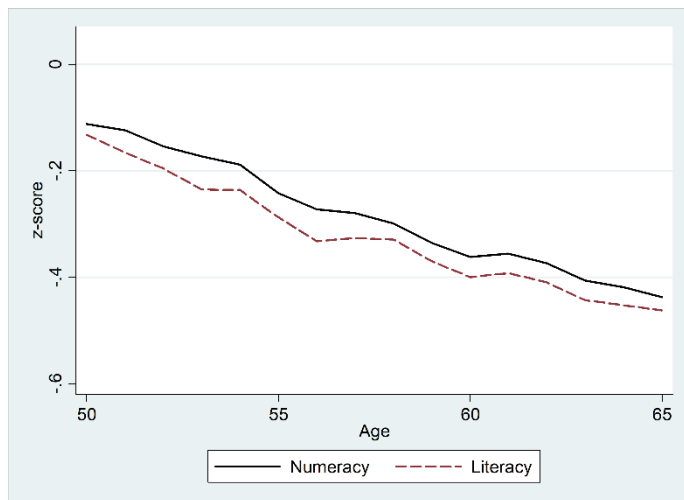


Figure A1. Average literacy and numeracy by age

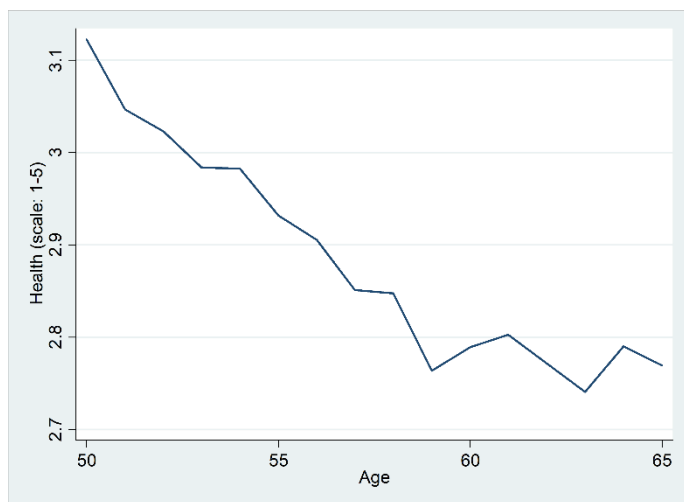


Figure A2. Average self-reported health by age

Notes: These figures exploit all available data for 50-65 year olds (25 countries for literacy and numeracy; 24 countries for self-reported health). All averages are weighted as to account for the country specific survey designs; on a cross-country level all countries are weighted equally when estimating the mean at each age.

A.3 Country-specific estimates

Table A2a. The number of years since you reached the retirement age and its effects: First-stage and reduced-form regressions separately by country. Sample: All

Outcome:	Years as retired	Literacy	Numeracy
Belgium	2.28* (1.22)	0.28 (0.29)	0.24 (0.27)
Chile	0.24* (0.14)	-0.013 (0.044)	-0.0020 (0.037)
Cyprus	2.37** (0.97)	-0.14 (0.29)	-0.25 (0.33)
Czech Republic	0.57*** (0.059)	-0.0097 (0.019)	-0.0038 (0.018)
Denmark	2.75*** (0.49)	0.049 (0.15)	0.046 (0.15)
Estonia	0.35*** (0.058)	-0.0093 (0.017)	0.0064 (0.016)
Finland	3.45*** (0.72)	0.077 (0.19)	0.24 (0.18)
France	0.54*** (0.043)	0.050** (0.020)	0.042** (0.021)
Greece	0.71*** (0.23)	-0.018 (0.072)	-0.018 (0.060)
Ireland	-	-	-
Israel	0.42** (0.17)	0.022 (0.059)	-0.0078 (0.059)
Italy	0.52*** (0.054)	0.047** (0.023)	0.013 (0.024)
Japan	0.26 (0.94)	0.044 (0.21)	0.17 (0.22)
Korea	0.17** (0.068)	-0.039* (0.023)	-0.061** (0.024)
Lithuania	0.19*** (0.064)	-0.0091 (0.027)	-0.030 (0.025)
Netherlands	2.81*** (0.88)	-0.13 (0.24)	-0.11 (0.24)
Norway	-	-	-
Poland	0.68*** (0.11)	-0.024 (0.031)	0.011 (0.029)
Russia	0.54*** (0.059)	-0.023 (0.025)	-0.044* (0.023)
Slovakia	0.49*** (0.034)	0.011 (0.012)	0.0066 (0.013)
Slovenia	0.68*** (0.052)	0.019 (0.021)	0.020 (0.021)
Spain	2.53** (0.95)	-0.094 (0.26)	-0.066 (0.24)
Sweden	0.96 (0.68)	-0.037 (0.18)	0.24 (0.18)
Turkey	0.47*** (0.17)	0.10* (0.053)	0.10* (0.061)
United Kingdom	0.33*** (0.078)	0.051* (0.026)	0.0062 (0.024)
Weighted average ^a	0.48	0.0082	0.0014

Notes: Controls included for age (linearly), gender, immigration status, years of schooling and parental education levels. Standard errors in parenthesis. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$ ^aInverse variance weighting.

Table A2b. The number of years since you reached the retirement age and its effects: First-stage and reduced-form regressions separately by country. Sample: All (Health)

Outcome:	Years as retired	Health
Belgium	2.29* (1.22)	0.17 (0.26)
Chile	0.24* (0.14)	-0.020 (0.042)
Cyprus	2.37** (0.97)	0.46 (0.28)
Czech Republic	0.57*** (0.059)	0.013 (0.020)
Denmark	2.75*** (0.49)	0.42** (0.20)
Estonia	0.34*** (0.058)	-0.0011 (0.014)
Finland	3.48*** (0.74)	0.33 (0.20)
France	0.54*** (0.043)	0.0041 (0.017)
Greece	0.71*** (0.23)	0.010 (0.058)
Ireland	-	-
Israel	0.41** (0.17)	0.067 (0.064)
Italy	0.52*** (0.054)	-0.0025 (0.028)
Japan	0.26 (0.94)	0.36 (0.32)
Korea	0.17** (0.068)	-0.020 (0.028)
Lithuania	0.19*** (0.064)	0.073*** (0.022)
Netherlands	2.81*** (0.88)	-0.51* (0.27)
Norway	-	-
Poland	0.68*** (0.11)	-0.034 (0.022)
Russia	0.54*** (0.059)	-0.039 (0.025)
Slovakia	0.49*** (0.034)	0.029** (0.012)
Slovenia	0.68*** (0.052)	0.025 (0.024)
Spain	2.53** (0.95)	0.19 (0.30)
Sweden	0.96 (0.68)	0.34 (0.25)
Turkey	-	-
United Kingdom	0.33*** (0.092)	0.070*** (0.025)
Weighted average ^a	0.48	0.014

Notes: Controls included for age (linearly), gender, immigration status, years of schooling and parental education levels. Standard errors in parenthesis. *** p<0.01, ** p<0.05, * p<0.1 ^aInverse variance weighting.

Table A2c. The number of years since you reached the retirement age and its effects: First-stage and reduced-form regressions separately by country.
Sample: Males only

Outcome:	Years as retired	Literacy	Numeracy
Belgium	4.01** (1.62)	0.14 (0.34)	0.013 (0.32)
Chile	-0.40 (1.13)	0.55 (0.56)	0.28 (0.45)
Cyprus	2.40** (1.19)	0.026 (0.35)	-0.069 (0.39)
Czech Republic	0.39*** (0.13)	0.024 (0.058)	-0.0034 (0.051)
Denmark	1.68** (0.67)	-0.13 (0.21)	-0.043 (0.21)
Estonia	0.47*** (0.17)	-0.090 (0.060)	-0.012 (0.053)
Finland	3.07** (1.23)	0.014 (0.31)	0.060 (0.28)
France	0.47*** (0.065)	0.042 (0.029)	0.042 (0.028)
Greece	0.63** (0.31)	-0.042 (0.093)	-0.043 (0.081)
Ireland	-	-	-
Israel	-	-	-
Italy	0.54*** (0.075)	0.056* (0.029)	0.014 (0.031)
Japan	0.62 (0.75)	0.0018 (0.29)	0.14 (0.31)
Korea	0.20*** (0.074)	-0.027 (0.033)	-0.042 (0.033)
Lithuania	0.33 (0.27)	-0.0049 (0.083)	-0.055 (0.074)
Netherlands	1.84** (0.86)	0.0011 (0.27)	0.0024 (0.25)
Norway	-	-	-
Poland	2.21 (1.75)	-0.18 (0.48)	-0.19 (0.43)
Russia	0.58*** (0.16)	0.075 (0.053)	0.058 (0.055)
Slovakia	0.67*** (0.21)	0.099 (0.067)	0.055 (0.066)
Slovenia	0.50*** (0.073)	0.041 (0.029)	0.048* (0.027)
Spain	2.93** (1.23)	0.10 (0.35)	0.030 (0.33)
Sweden	2.22** (1.03)	-0.35 (0.26)	-0.055 (0.28)
Turkey	0.46** (0.21)	0.085 (0.053)	0.067 (0.060)
United Kingdom	0.38 (1.29)	-0.33 (0.33)	-0.34 (0.30)
Weighed average ^a	0.46	0.029	0.018

Notes: Controls included for age (linearly), gender, immigration status, years of schooling and parental education levels. Standard errors in parenthesis. *** p<0.01, ** p<0.05, * p<0.1 ^aInverse variance weighting.

CHAPTER 6

Socioeconomic Background and College Application Strategies¹⁵⁰

Abstract

In this chapter, we analyze college application behaviors of high school graduates, and how these behaviors differ depending on socioeconomic background as measured by your neighborhood education level. To this purpose, we use a large administrative data consisting of Finnish high school students who graduated in 2012 and applied to college that same spring. Using a discrete choice model, we find that students gravitate towards colleges located nearby, and that this tendency is somewhat stronger for students from poorly educated neighborhoods. Also, female students from poorly educated neighborhoods apply in a way that is consistent with having a comparatively strong aversion to risk. These differences in application strategies are, however, unlikely to cause any large wage gaps between the student groups (i.e. those from highly and poorly educated neighborhoods).

¹⁵⁰ In co-operation with Roope Uusitalo

6.1 INTRODUCTION

The relationship between parents and children's educational and financial outcomes is well-documented (see, for example, Jäntti et al., 2006; Chavelier, 2004; Hertz et al., 2007 and the overview by Solon, 1999). Whether these intergenerational correlations are problematic or not, depends heavily on the mechanism behind: What are the driving forces? There are several hypotheses, which includes inherited ability and financial constraints (see the discussion by Black & Devereux, 2010). Another possible mechanism concerns preferences regarding higher education. A couple of recent studies have looked into this by analyzing college application behaviors of high school students. These studies typically find clear evidence of differences in application strategies depending on social background. Hoxby and Avery (2013) find that high-achieving low-income students seldom apply to selective institutions, while this is not the case for their high-income counterparts. They identify a couple of potential explanations, including a lack of guidance and lack of older role models. Other authors have found similar patterns (see Dillon and Smith, 2017; Heller, 2004; Hill, 2005 and Smith et al., 2013). A couple of authors (see Do, 2004; Frenette, 2006 and Griffith & Rothstein, 2009) have further tested whether proximity to colleges (or selective institutions) can explain the differences in outcomes between student groups. The results from these studies are mixed. Although proximity seems to matter for college choices, it is not clear whether it explains the outcome gap between low- and high-income students.

In this chapter, we study the application strategies of Finnish high school students, and how these strategies differ depending on social background. We find that students gravitate towards colleges located nearby, and that this tendency is stronger among those from poorly educated neighborhoods. Furthermore, female students from poorly educated neighborhoods apply in a way that is consistent with being comparatively wage risk averse. These patterns can only partly be explained by group differences in ability.

Despite these behavioral differences, the expected wage gap between students from highly and poorly educated neighborhoods is small. Furthermore, there is little support for the notion that students from poorly educated neighborhoods are hindered by their application behavior: When students from poorly educated neighborhoods are 'given' the behavior model of students from highly educated neighborhoods, the expected wage gaps between the groups are mainly unaffected. This suggests that differences in *endowments* – geographical location and high school grades – are of more importance. The same is not true when comparing male and female applicants however: When female students are 'given' the behavior model of male students, the gender wage gap is practically eliminated.

6.2 MODELING PORTFOLIO CHOICES

The aim is to analyze application behaviors of high school graduates, and how these behaviors differ between students depending on their neighborhood education level. To this purpose we utilize a conditional logit model, where students can choose one application portfolio from a near infinite set of portfolios. An application portfolio is a possible combination of educational programs; in total there are 313 programs and students can apply to anywhere between one and thirteen alternatives.¹⁵¹ We assume that the student chooses the one portfolio which maximizes expected utility. In the following section we present the main model describing portfolio expected utility. This model treats students as having one shot at applying, so that the outcome of ‘accept’ or ‘reject’ fully determines the student’s educational path.

6.2.1 The model

The student’s expected utility from choosing a specific portfolio is given by: $E[U_{i,j}] = \sum_n p_{i,n} u_{i,n}$, where $E[U_{i,j}]$ is the expected utility of portfolio j from the perspective of student i ; $u_{i,n}$ is the utility from educational program n belonging to this portfolio, with one of the educations being “high school diploma only”. $p_{i,n}$ denotes the probability of ending up with education n if choosing this portfolio, with $p_{i,N}$ being the probability of ending up with a high school diploma only, i.e. of not being admitted to any of the programs. The student can only end up with one of these educations so that the probabilities sum to one, $\sum_n p_{i,n} = 1$. The utility of educational program n is modeled as:

$$u_{i,n} = x_{i,n}\beta + u(w_n) + \varepsilon_{i,n}$$

where $x_{i,n}$ represents characteristics of education n including the log of the geographical distance to that college from the students home, the student’s employment prospect post-graduation, the female percentage on the post-graduate labor market and this percentage interacted with the gender of the student, as well as a vector of 18 indicator variables for different educational fields, with “high school diploma only” serving as reference.¹⁵² The ‘female

¹⁵¹ Not all combinations are allowed as students can apply to maximally four polytechnical programs and nine university programs.

¹⁵² The fields are: “Pharmacy and laboratory work”, “Business administration”, “Agronomy, horticulture, forestry”, “Pedagogy”, “TV- and movie-work, acting, art”, “Music and dance”, “Social sciences”, “Engineer, builder”, “Nutrition, veterinary medicine”, “Social work, health science”, “Science”, “Medicine, Odontology”, “Cultural worker”, “Languages, History, Philosophy”, “Theology, religious studies”, “Law”, “Architecture”, “Psychology”.

percentage' is primarily included as a proxy for unobserved occupational characteristics that men and women may respond differently to. $u(w_n)$ is the utility of the post-graduate wage for education n . The portfolio expected utility is then described by:

$$E[U_{i,j}] = E[x_{i,j}]\beta + E[u(w_{i,j})] + \varepsilon_{i,j} \quad (1)$$

where $E[x_{i,j}]$ represents the student's expected outcomes if choosing portfolio j : the expected log moving distance, the employment probability, the expected female percentage on the post-graduate labor market (interacted with the gender of the student) as well as a vector of probabilities for ending up in each of the 18 fields. The expected utility of the future wage, $E[u(w_{i,j})]$, is modeled as linearly increasing in expected log wages with an added effect for wage dispersion as measured by the variance for log wages. Section 2.2 (Measuring portfolio properties) discusses the basis for this specification, as well as the estimation of the expected values mentioned above. Finally, $\varepsilon_{i,j}$ is a vector of random error components which are treated as identically and independently extreme value distributed.

For practical reasons, we use only a tiny fraction of the full choice set of application portfolios in estimating the parameters. As shown by McFadden (1978) one can consistently estimate the parameters of a conditional logit by using a random subset of alternatives. Here, this subset is obtained by, for each student, randomly drawing nine portfolios from the distribution of portfolios actually chosen by the student population, in addition to the student's actual choice. If a sampled choice set includes less than ten portfolios (as a result of one or several portfolios having been drawn repeatedly) then a new choice set is sampled until ten unique portfolios are obtained.¹⁵³ As shown by McFadden, the probability of choosing portfolio k , conditional on the sampled subset, S_i , is then given by:

$$\Pr(k|S_i) = \frac{p_{i,k}q(S_i|k)}{\sum_{j \in S_i} p_{i,j}q(S_i|j)} = \frac{\exp[V_{i,k} + \ln q(S_i|k)]}{\sum_{j \in S_i} \exp[V_{i,j} + \ln q(S_i|j)]}$$

where $p_{i,k}$ is the probability of student i choosing portfolio k among the portfolios in the full choice set: $p_{i,k} = \exp(V_{i,k}) / \sum_{j \in F} \exp(V_{i,j})$, with F denoting the full choice set and $V_{i,j}$ denoting the structural part of expected utility. This is the standard equation describing the conditional logit model. $q(S_i|k)$ is the probability of drawing subset S_i given the actual choice k and considering the selection process; $\ln q(S_i|k)$ is the natural logarithm of this

¹⁵³ An alternative would be to draw each subset randomly and uniformly from the full choice set of portfolios. This, however, would arguably produce less precise estimators, as this subset would consist of comparatively undesirable alternatives providing little information on the reasons behind a student's choice (Train, 2009).

probability, with coefficient constrained to one in estimation. A description of the calculation of $q(S_i|k)$ is given in the Appendix, section A.1.

One central issue of this chapter is assessing whether the behavioral parameters of this model differ depending on neighborhood education level. In particular, we are interested in whether students from poorly educated neighborhoods are more (or less) sensitive to the expected moving distance and the properties of the portfolio wage distribution in comparison to students from highly educated neighborhoods. To the extent effects differ, this can potentially be an important explanation for the differences in schooling outcomes between groups. If, for example, distance is a stronger barrier for students from poorly educated neighborhoods – making choices more geographically restricted for this group – then this is also likely to be reflected in schooling outcomes. In practice, we test for differences in effects by estimating a model that includes interactions between neighborhood education levels and the independent variables. In a second step we further test whether any differences in effects between the groups can be explained by high school grades. To this purpose we reweight observations for all neighborhood groups as to represent the pooled distribution of high school grades.

6.2.2 Measuring portfolio properties

We now return to discussing how the expected values entering (1) are modeled. Skipping the student index for simplicity, the expected value for some characteristic (x) of a portfolio is described by:

$$\begin{aligned}
 E[x_j] &= \sum_{n=1}^N p_{(n)} x_{(n)} \\
 &= \pi_{(1)} x_{(1)} + (1 - \pi_{(1)}) \pi_{(2)} x_{(2)} + \dots \\
 &\quad + (1 - \pi_{(1)}) (1 - \pi_{(2)}) \dots (1 - \pi_{(N-1)}) x_{(N)}
 \end{aligned} \tag{2}$$

where $\pi_{(n)}$ is the student's probability of being admitted to the program ranked as number n among the programs in that portfolio, whereas $p_{(n)}$ denotes the probability of enrolling in that program. If a student is admitted to several programs she will, in other words, accept the one she ranks the highest. As this ranking is unobservable to us, however, we rank programs according to admissions probabilities, starting with the toughest program (from the perspective of the student). Not being admitted to any of the portfolio programs is treated as being the least preferred outcome, denoted by $x_{(N)}$.

The student's probability of being admitted to a program, $\pi_{(n)}$, is estimated based on data from the previous cohort of applicants, using separate logit regressions by each diploma (33 in total). The outcome measures whether an applicant was admitted and the predictors include the national matriculation exam results and different intercepts by college. The matriculation exam results are included through several variables measuring the result from different subjects, interacted with the student having taken that subject.¹⁵⁴ This captures how educational fields weigh subjects differently (for example, math would have more weight for a future engineer than a musician). In practice, matriculation exam scores are not the sole determinant of admittance success. Entrance exams, interviews, students' stated priorities and earlier work experience may also come into play. Note, however, that our goal is not to predict the outcome of 'admitted' or 'rejected' as well as possible, but rather to estimate the admissions probabilities from the viewpoint of students. It is not likely that students have access to more or higher quality information than we in general, and if anything, the predicted probabilities are likely to understate the uncertainty as experienced by applicants.

Using (2) with estimated probabilities, $\hat{\pi}_{(n)}$, we calculate the expected log moving distance, the employment probability and the probability of ending up in a specific field for each student-portfolio combination. The expected utility of portfolio wages, $E[u(w_j)]$, however, cannot be calculated directly using (2) as this expectancy also depends on the specification of the utility function, $u(w_j)$. Here, we assume that students exhibit constant relative risk aversion, described by:

$$u(w) = \frac{w^{1-\gamma}}{1-\gamma}$$

where w denotes wages and γ is the relative risk aversion parameter. The expected utility is then given by $E[u(w)] = \frac{1}{1-\gamma} E[w^{1-\gamma}]$. In order to calculate this expectancy, we need a distributional model for portfolio wages. Letting w_j denote the wage if choosing portfolio j , then w_j is described as a sum of random variables as follows:

¹⁵⁴ The student has two mandatory subjects – her mother tongue (or one of the native languages if the student has a foreign mother tongue) and a second language (chosen by the student, usually English or Swedish). Furthermore, the student selects at least another two subjects from the following three categories: math, languages and "general studies". Within each category there are further choices to make. For example, if math is chosen then the student can write either the short or extended math curriculum; if "general studies" is chosen then the student can choose between an extended list of subjects including, for example, physics, chemistry, history, psychology and religion.

$$w_j = \sum_{n=1}^N I(z = n)w_n$$

where $I(z = n)$ is an indicator variable taking on the value 1 if z equals n and zero otherwise; z equals n if the student enrolls in program n . If we assume that w_n follows a lognormal distribution with location and shape parameters denoted by μ_n and σ_n^2 , then w_j is approximately log normal for the portfolios in the data.¹⁵⁵ This is shown by example in the Appendix, section A.2. The location and shape parameters of w_j are given by $\mu_j = E[\mu_n]$ and $\sigma_j^2 = \text{Var}[\mu_n] + E[\sigma_n^2]$, respectively. Treating w_j as log normally distributed implies that students maximize:

$$E[u(w_j)] = \frac{1}{1-\gamma} E[w_j^{1-\gamma}] = \frac{1}{1-\gamma} \exp[(1-\gamma)\mu_j + 0.5(1-\gamma)^2\sigma_j^2]$$

which, by monotonicity, is equivalent to maximizing:

$$u(\mu_j, \sigma_j^2) = \mu_j - 0.5(\gamma - 1)\sigma_j^2 \quad (3)$$

From (3) we observe that students with a risk aversion parameter above one, $\gamma > 1$, are penalized by increases in σ_j^2 while the opposite is true for students with a risk aversion parameter below one.

6.2.3 Marginal effects and error margins

Unlike in linear regression, the coefficients in a conditional logit are biased towards zero if relevant predictors are omitted, also when these are uncorrelated with the remaining predictors, i.e. the coefficients are inherently linked to the error variance. The practical implication of this is that coefficients for different neighborhood education levels are not directly comparable, only ratios between coefficients are (such as the relative risk aversion). However, average marginal effects are not affected by this kind of attenuation bias (see Mood, 2010). For this reason, we present effects as average marginal effects in addition to the raw coefficients. The marginal effects are calculated separately for each observation (i.e. for each student-portfolio combination) and averaged over all observations.¹⁵⁶ Note, however,

¹⁵⁵ The location parameter is the expected value of log wages; the shape parameter is the variance for log wages.

¹⁵⁶ The own marginal effect for an observation: $\frac{dp}{dx} = p(1-p)\beta$, where p is the probability of choosing this portfolio and β the coefficient of interest.

that the marginal effects are conditional on the sampled subset of portfolios meaning that absolute magnitudes should be viewed in that context. Error margins for average marginal effects (and risk aversion parameters) are calculated by simulating the relevant sampling distribution by repeatedly drawing average marginal effects (risk aversion coefficients) using the empirical multivariate normal distribution for the underlying coefficient vector. The matrix of covariances describing this distribution is estimated using clustering on the high school level.

6.2.4 Model selection and validity

An alternative approach for analyzing application behavior is to study program choices instead of portfolio choices, i.e. a model where the choice set consists of the full set of programs with positive outcomes for all programs applied to. This is the approach chosen by Hoxby and Avery (2013). Such a model assumes that students pick each program they prefer to having a high school diploma only, and that the value of adding another program to the portfolio can be assessed independently of any programs already included. For example, the value of adding a program may then depend on it being, say, medical, but not on how that changes your chances of becoming a medical doctor – something determined by the portfolio composition as a whole. Our assumption is that students actually care about the latter, why this is the modeling strategy chosen here.

The random expected utility approach is not without weaknesses, however. The well-known limitations of conditional logit estimators and observational data analysis in general are also present in this setting. The conditional logit estimator relies on the assumption of unobservable factors being uncorrelated over portfolios; if one portfolio becomes more attractive the other ones loses in attractiveness to the same degree (see Train, 2009). In other words, all relevant portfolio characteristics are assumed to be captured by the model, leaving only white noise in the error term. This is a strong assumption and estimates should be interpreted with caution in light of this limitation.

Moreover, as with observational data in general, the estimated effects describe statistical associations, not causal relationships. If we, for example, observe that students are more likely to choose portfolios with high expected wages, this could reflect both a preference for high wages and/or preferences regarding other correlated portfolio characteristics. Even if the model would be causal, measurement errors may still bias the estimates. For example, students may systematically over- or underestimate their chances of being admitted to different programs. Also, students may not view their future wage as a random draw from the same wage distribution we observe; they

could potentially have either better or worse information available. Both of these factors could contribute to the students' observed behavior differing from the intended one.

6.3 DATA DESCRIPTION

This chapter presents the sampling procedure and describes the main variables of interest. A descriptive comparison of the chosen portfolios of students from different educational backgrounds is also presented. In summary, we observe clear differences in mean portfolio characteristics depending on neighborhood education level. Students from highly educated neighborhoods choose portfolios giving them a comparatively small chance of admittance to an educational program, but a comparatively high expected wage and wage dispersion. Students from poorly educated neighborhoods, on the other hand, choose portfolios giving them a comparatively high expected moving distance from home.

6.3.1 Sampling and variable description

The sample is based on a large administrative data on individual level applications to higher education. The sample covers roughly 18,000 students or 80 percent of the total population of Finnish students born 1991 or later, who graduated from high school in 2012 and applied to higher education that same spring. The 20-percent fallout is due to missing values on neighborhood education level.

For each student, we observe the full set of applications sent to educational programs in 2012. Besides from these applications, the sample also contains information regarding a number of student- and program-level characteristics. The student-level variables include gender, matriculation exam results, place of residence and the percentage of highly educated adults in the student's local neighborhood (250 x 250 meters). The bottom 25 percent on this scale are defined as 'students from poorly educated neighborhood'; the middle 50 percent as 'students from neighborhoods with moderate education levels' and the top 25 percent as 'students from highly educated neighborhoods'. In moderately educated neighborhoods, the percentage of highly educated adults are 13-31 (average 20 percent); in poorly and highly educated neighborhoods these numbers are, on average, 7 and 44 percent, respectively. The data on neighborhoods is collected in 2008, ensuring that the student herself is not part of the adult population (18 years and above) at the time of measurement.

A program is a combination of a college and a diploma (for example: Helsinki University, Class teacher). Program-level variables include the geographical location of the program as well as some post-graduate labor market properties: the expected wage, the wage quartiles, employment rate and the gender distribution. The employment rate and wage properties are measured in 2008 using 30-34 year old individuals with the relevant

diploma. In other words, several programs share the same labor market properties if they award the same diploma. If a student is not admitted to any program, her future labor market outcomes is measured using individuals with a high school diploma only. Using the expected post-graduate wage and the wage quartiles, we calculate the expected log wage and log wage variance under the assumption of wages being log normally distributed. The gender distribution on the post-graduate labor market is measured as the percentage of women who enrolled in that program among the previous cohort of applicants.¹⁵⁷ The labor market for students with a high school diploma only is treated as gender neutral.

By combining the student- and program-level information we further create two variables that are specific to each student-program combination. This includes the moving distance to the program and the admittance probability. The moving distance is measured as the straight-line distance between the student's home and the program, or set to zero for having a high school diploma only. In the regressions, we use the log of the moving distance and add one kilometer to each distance. The admittance probabilities are estimated based on matriculation exam results as described in section 2.2. The following two sections give a more detailed description of the distributions for these variables.

6.3.2 The admittance probability

Overall, the average admittance probability is 23 percent, meaning that we expect a randomly chosen student to have a 23 percent chance of being admitted to a randomly chosen program. For 90 percent of students, the average admittance probability (over all programs) lies somewhere between 9 and 42 percent. Most of the variation in admittance probabilities is, however, found between programs rather than students; generally students can choose between programs with admittance probabilities that practically cover the whole scale.

Students from poorly educated neighborhoods tend to have lower admittance probabilities than those from highly, the former group averaging at 22 percent compared to 25 percent for students from highly educated neighborhoods; this difference represents roughly one third of a standard deviation. The difference in admittance probabilities between male and female applicants is small, but marginally favoring males.

¹⁵⁷ If this cohort is small (which is common), we use a weighted mean that combines information regarding the percentage of women in that specific program with the percentage of women in the larger group of programs awarding the same diploma.

6.3.3 Moving distance

Overall, mean distance from a student's home to a program is 259 kilometers. Students from highly educated neighborhoods generally live somewhat closer to the programs; for this group the mean distance is 249 kilometers compared to 270 kilometers for students from poorly educated neighborhoods. The sharpest difference between the groups, however, is that students from highly educated neighborhoods tend to live in close proximity to at least one educational program; of these students, 97 percent have an educational program within 20 kilometers from home, whereas this is the case for only 67 percent of students from poorly educated neighborhoods. This geographical pattern is also illustrated graphically in the Appendix, see section A.3.

6.3.4 Portfolio characteristics

In the regressions, we use information regarding portfolios rather than programs, such as the student's expected log moving distance or expected log wage if choosing a particular portfolio. This section describes the average portfolio properties of the chosen portfolios. These statistics are presented in table 1, where the column "chosen" gives the average portfolio characteristics for the chosen portfolio; the column "control" presents means for a random sample of nine portfolios per student, drawn from the distribution of portfolios chosen by the student population. The means are presented for the whole sample and separately by neighborhood education levels.

Table 1. Mean portfolio characteristics

<i>Sample: All</i>	Chosen	Control
Pr(admitted)	0.403	0.358
Expected moving distance (km)	48	87
Expected wage (€/month)	2552	2459
Wage standard deviation (€/month)	973	931
Expected employment probability	0.874	0.872
<i>Sample: Poorly educated neighborhood</i>	Chosen	Control
Pr(admitted)	0.416	0.332
Expected moving distance (km)	58	85
Expected wage (€/month)	2522	2438
Wage standard deviation (€/month)	939	921
Expected employment probability	0.873	0.872
<i>Sample: Moderately educated neighborhood</i>	Chosen	Control
Pr(admitted)	0.409	0.359
Expected moving distance (km)	49	87
Expected wage (€/month)	2554	2461
Wage standard deviation (€/month)	970	932
Expected employment probability	0.874	0.872
<i>Sample: Highly educated neighborhood</i>	Chosen	Control
Pr(admitted)	0.378	0.382
Expected moving distance (km)	36	87
Expected wage (€/month)	2578	2477
Wage standard deviation (€/month)	1012	940
Expected employment probability	0.874	0.871

Notes: 17,999 students in total (chosen: 17,999 observations; control: 161,991 observations).

The table shows that the average student chooses a portfolio giving her a 40 percent chance of being admitted to at least one program. If, on the other hand, students would choose the same portfolios as other applicants, they would have a 36 percent chance of admittance. This discrepancy is reflecting the fact that a, say, linguistically gifted student is more prone to apply to linguistic educations than other students. The table also reveals that students from highly educated neighborhoods choose portfolios giving them a comparatively small chance of being admitted, 38 percent, which can be compared to the 42 percent admittance probability for students from poorly educated neighborhoods. In practice, the difference in actual admissions percentages is smaller (just under 1 %-point) which could be a result of students from highly educated neighborhoods putting more effort on their

applications.¹⁵⁸ Nevertheless, they are admitted and eventually enroll to a somewhat lesser degree than students from poorly educated neighborhoods, in spite of having higher grades. This pickiness of students from highly educated neighborhoods is not shared by highly gifted students overall, who tend to choose portfolios giving them a high chance of admittance (also in comparison to the control portfolios).

Table 1 also shows that the expected moving distance for the chosen portfolios is roughly half of that of the control, implying that students gravitate towards colleges located nearby. Students from poorly educated neighborhoods have the longest expected moving distance; those from highly have the shortest. This is no surprise given the fact that these students have a comparatively large supply of programs nearby, and also take a higher risk of not being admitted to any program.

Furthermore, students generally choose portfolios with a comparatively high expected wage. For the chosen portfolios the expected wage is approximately 90€ higher per month than for the controls. This is partly due to the fact that students choose portfolios giving them a relatively high admittance probability. But also after accounting for this, students tend to choose portfolios with high expected wages. This is more so for students from highly educated neighborhoods. On the other hand, students would get portfolios with lower wage dispersion if they applied to the same programs as others. This is especially so for students from highly educated neighborhoods.

¹⁵⁸ By actual admittance percentages we mean the percentage of students who eventually were admitted to at least one program.

6.4 EVIDENCE ON DIFFERENCES IN APPLICANT BEHAVIOR BY NEIGHBORHOOD EDUCATION LEVEL

For most portfolio characteristics, the hypotheses of equality of parameters between neighborhood education levels can be strongly rejected. We find that students from poorly educated neighborhoods are relatively sensitive to the log moving distance as compared to those from highly educated neighborhoods. Also, students from poorly educated neighborhoods – girls in particular – apply in a way that is consistent with having a stronger aversion to risk. The main results are presented in table 2 (for the whole sample and separately by gender) and in table 3 (separately by neighborhood education levels). The tables report both average marginal effects as well as the underlying coefficients.

6.4.1 Moving distance

Overall, the probability of choosing a portfolio decreases with the expected moving distance; when increasing all distances in a portfolio by ten percent the probability of choosing that portfolio decreases by 1.6 percentage points or 21 percent on average.¹⁵⁹ This effect is similar for boys and girls and is presented in table 2 for the whole sample and separately by gender.

There are also significant differences in sensitivity to distance between students depending on neighborhood education level, so that students from poorly educated neighborhoods are estimated to be more sensitive to distance than those from highly.¹⁶⁰ These results are presented in table 3. Overall, a ten percent increase in all distances corresponds to a 1.9 percentage point (30 percent) decrease in the probability of choosing that portfolio for students from poorly educated neighborhoods. For students from moderately and highly educated neighborhoods, the corresponding effects are 1.7 percentage points (23 percent) and 1.4 percentage points (19 percent), respectively. This pattern is similar for girls and boys; see section A.4 in the Appendix.

Can this pattern be accounted for by differences in academic ability between the groups? We measure a student's academic ability as her expected probability of being admitted to a randomly chosen program. This measu-re

¹⁵⁹ The marginal effect of a 10 percent increase is calculated as $ME \cdot \ln(1.1)$ where ME is the marginal effect when distance is measured in log-units. The corresponding elasticity is calculated as $\frac{ME}{p} \cdot \ln(1.1)$ where p is the probability of this student choosing this portfolio.

¹⁶⁰ The estimates have been adjusted for girls being slightly overrepresented among students from poorly educated neighborhoods by reweighting the observations in each subsample as to represent the pooled distribution.

Table 2. Conditional logit estimates for portfolio choice, overall and separately by gender

	All		Women		Men		Difference in coefficients z-value (p-value)
	Coeff.	AME	Coeff.	AME	Coeff.	AME	
Expected log distance	-2.464** (0.066)	-0.164** (0.004)	-2.444** (0.078)	-0.171** (0.005)	-2.505** (0.078)	-0.160** (0.004)	-0.72 (0.474)
Expected log wage	11.964** (1.018)	0.797** (0.065)	8.670** (2.113)	0.606** (0.143)	13.521** (1.115)	0.863** (0.076)	2.18 (0.030)
Log wage variance	2.895** (0.737)	0.193** (0.052)	-1.167 (2.741)	-0.082 (0.189)	8.039** (0.882)	0.513** (0.058)	3.13 (0.002)
Employment probability (unit: 10 %-points)	-1.426** (0.244)	-0.095** (0.016)	-1.201** (0.442)	-0.084** (0.029)	-1.021* (0.106)	-0.065 (0.032)	0.32 (0.747)
Expected female percentage (unit: 10 %-points)	-0.410** (0.067)	-0.027** (0.005)	1.096** (0.087)	0.077** (0.006)	-0.599** (0.106)	-0.038* (0.007)	-13.19 (0.000)
Expected female percentage *female student	1.428** (0.035)	0.095** (0.002)	-	-	-	-	
Observations	179,990		107,320		72,670		
Students	17,999		10,732		7,267		

Notes: Standard errors in parenthesis. ** p<0.01; * p<0.05. AME is short for the average marginal effect. The regressions also include controls for the probability of ending up in each of 18 fields.

Table 3. Conditional logit estimates for portfolio choice, separately by neighborhood education level (low, medium & high)

	Low		Medium		High		Difference in coefficients Chi2 (p-value)
	Coeff.	AME	Coeff.	AME	Coeff.	AME	
Expected log distance	-3.443** (0.138)	-0.199** (0.005)	-2.708** (0.070)	-0.175** (0.003)	-2.159** (0.075)	-0.149** (0.004)	82.39 (0.000)
Expected log wage	10.308** (2.457)	0.596** (0.146)	12.783** (1.408)	0.828** (0.090)	15.520** (1.887)	1.073** (0.114)	3.04 (0.218)
Log wage variance	-2.410 (3.538)	-0.139 (0.204)	3.043** (0.962)	0.197** (0.061)	5.107** (0.892)	0.353** (0.063)	6.12 (0.047)
Employment probability (unit: 10 %-points)	-1.656** (0.575)	-0.096** (0.035)	-1.605** (0.314)	-0.104** (0.021)	-1.160* (0.494)	-0.080* (0.031)	0.74 (0.692)
Expected female percentage (unit: 10 %-points)	-0.424** (0.108)	-0.025** (0.007)	-0.402** (0.095)	-0.026** (0.006)	-0.108 (0.158)	-0.007 (0.011)	3.32 (0.190)
Expected female percentage *female student	1.401** (0.068)	0.081** (0.004)	1.443** (0.049)	0.094** (0.003)	1.427** (0.077)	0.099** (0.005)	0.24 (0.886)
Observations	45,270		89,340		45,380		
Students	4,527		8,934		4,538		

Notes: Standard errors in parenthesis. **p<0.01; *p<0.05. AME is short for the average marginal effect. The regressions also include controls for the probability of ending up in each of 18 fields. The estimates have been adjusted for girls being slightly overrepresented among students from poorly educated neighborhoods by reweighting the observations in each subsample as to represent the pooled distribution.

Table 4. Ability-weighted conditional logit estimates for portfolio choice, separately by neighborhood education level (low, medium & high)

	Low		Medium		High		Difference in coefficients Chi2 (p-value)
	Coeff.	AME	Coeff.	AME	Coeff.	AME	
Expected log distance	-3.267** (0.140)	-0.193** (0.005)	-2.709** (0.070)	-0.176** (0.004)	-2.220** (0.077)	-0.153** (0.005)	55.17 (0.000)
Expected log wage	8.447** (2.178)	0.500** (0.129)	12.676** (1.401)	0.821** (0.085)	15.108** (2.042)	1.041** (0.117)	4.99 (0.082)
Log wage variance	0.317 (2.941)	0.019 (0.185)	2.992** (0.958)	0.194** (0.062)	4.711** (1.008)	0.325** (0.068)	3.00 (0.223)
Employment probability (unit: 10 %-points)	-1.287** (0.511)	-0.076** (0.030)	-1.600** (0.312)	-0.104** (0.019)	-1.057* (0.532)	-0.073* (0.036)	1.02 (0.601)
Expected female percentage (unit: 10 %-points)	-0.433** (0.112)	-0.026** (0.007)	-0.408** (0.095)	-0.026** (0.007)	-0.174 (0.163)	-0.012 (0.011)	2.01 (0.365)
Expected female percentage *female student	1.400** (0.070)	0.083** (0.004)	1.447** (0.049)	0.094** (0.003)	1.467** (0.081)	0.101** (0.006)	0.45 (0.800)
Observations	45,270		89,340		45,380		
Students	4,527		8,934		4,538		

Notes: Standard errors in parenthesis. **p<0.01; *p<0.05. AME is short for the average marginal effect. The regressions also include controls for the probability of ending up in each of 18 fields. The estimates have been adjusted as to represent the pooled two-dimensional distribution of academic ability and gender.

6.4.2 Properties of the portfolio wage distribution

Overall, the probability of choosing an application portfolio is increasing in both expected log wages and the log wage variance, the coefficient for expected log wages being roughly four times the coefficient for the variance (see table 2). This translates into a modest risk aversion coefficient at 0.5, with a 95-percent confidence interval (CI) allowing for a parameter value in the range of 0.3 to 0.8. In order to put this number into perspective, a risk aversion coefficient at 0.5 corresponds to being indifferent between a certain wage at 2914 euro and a fair bet between a 2000- and 4000-euro wage.¹⁶² This estimate differs between girls and boys; for girls, the estimated risk coefficient is 1.3 (CI: -0.6; 2.1) which corresponds to a certain wage at 2778 euros in the example above, while boys apply in a way that is consistent with being practically risk neutral with a coefficient at -0.2 (CI: -0.5; 0.1). The underlying coefficients for these estimates are given in table 2, whereas table 5 compiles the risk aversion estimates.

Table 5. Relative risk aversion estimates

<i>Neighborhood education level</i>				
<i>Gender</i>	All	Low	Medium	High
All	0.5 (0.3-0.8)	1.5 (-0.4-2.5)	0.5 (0.2-0.8)	0.3 (0.1-0.5)
Women	1.3 (-0.6-2.1)	3.4 (0.8-5.1)*	1.0 (-0.7-1.9)	0.7 (0.0-1.2)
Men	-0.2 (-0.5-0.1)	-0.3 (-1.6-0.5)	-0.1 (-0.5-0.2)	-0.2 (-0.7-0.2)

Notes: 95-percent confidence intervals in parentheses. *1.4 percent of the draws delivered nonsensical estimates – a distaste for higher wages – these draws were disregarded so that the final sampling distribution only contains possible risk aversion estimates. The lower limit is then given by the $[100/(100 - q)] \cdot 2.5^{\text{th}}$ percentile and the upper limit by the $100 - [100/(100 - q)] \cdot 2.5^{\text{th}}$ percentile; q is the percentage of nonsensical estimates on repeated draws.

There are also differences in responsiveness to the portfolio wage properties depending on neighborhood education level; students from poorly educated neighborhoods choose portfolios in a way that is consistent with being relatively risk averse in comparison to students from highly educated neighborhoods (see table 5). The estimated risk aversion coefficient for students from poorly educated neighborhoods is 1.5 (CI: -0.4; 2.5); the corresponding figures for students from moderately and highly educated neighborhoods are

¹⁶² $x = 2914$ solves: $0.5 \cdot \frac{2000^{1-\gamma}}{1-\gamma} + 0.5 \cdot \frac{4000^{1-\gamma}}{1-\gamma} = \frac{x^{1-\gamma}}{1-\gamma}$ where $\gamma = 0.5$

0.5 (CI: 0.2; 0.8) and 0.3 (CI: 0.1; 0.5), respectively. These estimates have been adjusted for the slight overrepresentation of girls among students from poorly educated neighborhoods.

Splitting the data into a male and female subsample shows that this pattern over neighborhood education levels is driven by girls while all male groups are estimated to be more or less risk neutral (see table 5). For female students from poorly educated neighborhoods, the coefficient of risk aversion is estimated to 3.4 (CI: 0.8; 5.1); for the moderately and highly educated neighborhoods the coefficients are 1.0 (CI: -0.7; 1.9) and 0.7 (CI: 0.0; 1.2), respectively. For male students, the risk aversion coefficient varies between -0.3 and -0.1 depending on group, with highly overlapping confidence intervals. The only significant difference is observed between female students from poorly and highly educated neighborhoods.

Recent research has suggested an association between intelligence and risk-taking behavior (see the discussion by Dohmen et al., 2018). We test whether the estimates are sensitive to adjusting for differences in academic ability by reweighting the observations in each subgroup as to represent the pooled (two-dimensional) distribution of academic ability and gender. In particular, we would expect group differences in risk aversion to decrease as a consequence. This is also what we observe; the risk coefficient for students from poorly educated neighborhoods decreases from 1.5 to 0.9 and for students from highly educated neighborhoods the coefficient increases from 0.3 to 0.4.¹⁶³

One interpretation of these estimates is that they reflect a comparatively high aversion to risk among students from poorly educated neighborhoods and women in particular. Another possibility is that they reflect differences in unobservable occupational preferences. There is suggestive evidence supporting both of these interpretations.

Firstly, for a subsample of students we have access to independent measures of risk attitude, and we find these to match well with actual application behavior. The independent measures come from a representative survey targeting high school graduates, and was conducted a couple of months prior to the application deadlines. One of the survey questions asks students to make a choice between two hypothetical educational programs. These are equal in all respects except for the post-graduate labor markets, with one having a higher mean wage and wage dispersion. The wage

¹⁶³ Highly able students are estimated to be less risk averse than moderately able students; the first group having an estimated risk coefficient at 0.2 and the second group 1.5. For poorly able students, the coefficient of risk aversion is undefined, as the coefficient for expected log wages is negative (and insignificant). Here, you are a highly able student if you belong to the top quartile group and a poorly able student if you belong to the bottom quartile group; the middle 50 percent are 'moderately able students'.

distributions are constructed so that students with a risk aversion coefficient at unity would be indifferent between the programs. Approximately half of the students claim to prefer the 'risky labor market'. Turning to their actual application behavior, we estimate their risk coefficient to 0.2 (CI: -0.2; 0.6). Students who picked the 'safe labor market' have an estimated risk coefficient at 2.7 (CI: 1.1; 4.9). This represents a highly significant difference, and lends credibility to the idea that the estimated risk coefficients do, in fact, reflect actual risk attitudes. We develop the discussion on these results in the Appendix, section A.5.

It is, however, also possible that the risk coefficients reflect differences in unobservable preferences between groups. For example, we observe a significantly negative effect of increasing the employment probability. This is likely to reflect unobserved occupational preferences rather than preferences regarding employment prospects, *per se*.¹⁶⁴

6.4.3 Alternative specifications and robustness

In this section, we study the robustness of the results when using alternative models, and when measuring some independent variables in alternative ways. Overall, we find that the trends in the data are relatively robust, while the absolute magnitudes for the estimated risk coefficients are sensitive to some alterations. This is shown in table A3 in the Appendix (see section A6. Robustness). The first column presents the main estimates for the coefficient of risk aversion in different subsamples; the middle column presents the lowest estimates when excluding one set of control variables at a time; the last column presents the highest estimates. Most notably, the risk coefficient for women is highly sensitive to controlling for the expected gender composition. Excluding this control brings the female estimate from 1.3 to 11.8. To see why, it is worth noting that the wage distribution for a particular portfolio is largely determined by the occupations that make up that portfolio. Hence, covariates that capture relevant occupational characteristics are expected to set the level for the risk aversion coefficient.

We further experiment with three alternative ways of measuring certain independent variables. Firstly, we recode the moving distance if ending up with a high school diploma from zero to 50 kilometers. It is not clear how one appropriately measures moving distance for the outside option, why using an alternative measurement is useful as a comparison. Similarly, we recode the percentage of women on the post-graduate labor market if not admitted

¹⁶⁴ It is not clear what these occupational characteristics are. One possibility is that high employment rates correlate with poor parental-leave opportunities. This interpretation is also in line with the finding that the negative employment-effect is only significant for women.

to any program from 50 to 70 (for girls) and from 50 to 30 (for boys). Again, these are (more or less) arbitrarily chosen numbers, but nevertheless informative as a comparison. Thirdly, we estimate the expected values and variance entering the value function by randomly ranking programs. As discussed in section 2.2, the main model assumes that students accept the toughest program if given several choices. Now we let students pick that program at random. The results are presented in table A4 (risk aversion coefficients) and A.5 (average marginal effects for the expected log distance). As shown in the tables, the estimates are relatively insensitive to these changes.

6.4.4 Distributional implications – a descriptive analysis

Students from poorly educated neighborhoods are estimated to be relatively sensitive to moving distance and apply in a way that is consistent with being comparatively risk averse. Also, these students live further from colleges and have lower admissions probabilities on average. In this section, we examine the implications of these differences in application behaviors and endowments on the distribution of students over portfolios. To this end, we focus specifically on one portfolio property – the expected wage. We carry out the following thought experiment. Students from poorly educated neighborhoods choose portfolios giving them an expected wage at ~2750 euros per month: What would their expected wage be if they had the application behavior of students from highly educated neighborhoods? Or conversely, what would their expected wage be if they had the endowments (place of residence and high school grades) of students from highly educated neighborhoods?

Table 6 shows expected wages for different combinations of endowments and application behaviors. For example, the first cell in the upper left corner shows that women from poorly educated neighborhoods are predicted to choose portfolios giving them an expected wage at 2638 euros/month. When given the application behavior, i.e. vector of coefficients, of girls from highly educated neighborhoods, this expectancy rises to 2656 euros. This value is calculated by predicting probabilities over portfolios using the sample of girls from poorly educated neighborhoods, but the coefficients of girls from highly educated neighborhoods. In the predictions, we exclude the portfolios actually chosen by the students, so that the choice set is comparable across groups.¹⁶⁵ Expected wages in other cells are calculated accordingly, so that endowments vary within rows while application behavior is held constant.

¹⁶⁵ The predicted probabilities are adjusted by an appropriate factor so that they always sum to unity.

Similarly, application behavior varies within columns while endowments are held constant.

Table 6. Predicted expected wages depending on application behavior and endowments (low/high indicate neighborhood education level)

<i>Coefficients of female + low</i>				
<i>Endowments of:</i>	Female + low 2638	Female + high 2699	Male + low 2654	Male + high 2736
<i>Coefficients of female + high</i>				
<i>Endowments of:</i>	Female + low 2656	Female + high 2708	Male + low 2661	Male + high 2721
<i>Coefficients of male + low</i>				
<i>Endowments of:</i>	Female + low 2798	Female + high 2841	Male + low 2842	Male + high 2892
<i>Coefficients of male + high</i>				
<i>Endowments of:</i>	Female + low 2760	Female + high 2834	Male + low 2791	Male + high 2868

Notes: The own expected wage in bold.

Table 6 reveals two facts. Firstly, the gender gap in expected wages is largely explainable by gender differences in *application behavior* (as opposed to endowments). For example, women from poorly educated neighborhoods have an expected wage at 2638 euros; men from poorly educated neighbourhoods have an expected wage at 2842 euros. That is a 204-euro difference. This gap reduces to 44 euros when the women are given the application behavior of the men. The pattern is similar when reversing the roles – the men are given the application behavior of the women – or when comparing men and women from highly educated neighborhoods. This result is not very surprising, since male and female applicants have similar endowments, i.e. they have roughly the same geographical distribution and overall admissions probability.¹⁶⁶ Also, the application strategy of men is favoring higher wages. For example, they are estimated to be practically risk neutral.

¹⁶⁶ However, men have a somewhat higher chance of being admitted to a top paid program; women have a somewhat higher chance of being admitted to a poorly paid program. Here, 'poorly paid' means an education with an expected wage at 2500 euros/month or less; 'top paid' means an education with an expected wage of at least 3000 euros/month. Approximately one

Table 6 also suggests that the gap between students from highly and poorly educated neighborhoods is largely explainable by differences in *endowments* (as opposed to application behavior). For women, there is a 70-euro gap between applicants from highly and poorly educated neighborhoods. This gap is little reduced when the application behavior is forced to be the same in both groups (either equal to that of women from highly or poorly educated neighborhoods). For men, the gap widens as students from poorly educated neighborhoods are given the application strategy of those from highly (or when those from highly are given the application behavior of those from poorly). Hence, there is nothing in the data that would suggest that male students from poorly educated neighborhoods are hindered by their application strategy in this regard. Also, there is only a 26-euro gap to begin with.

What kinds of endowments make students from poorly educated neighborhoods have lower expected wages than those from highly? A comparison of the geographical location and grades between the groups show that both factors may come into play. For students from poorly educated neighborhoods, the geographically closest college pays an average post-graduate wage at 2706 euros/month.¹⁶⁷ The corresponding figure for students from highly educated neighborhoods is 3239 euros. For the ten closest colleges, the figures are 3020 and 3160 euros, respectively.¹⁶⁸ Since students respond negatively to increasing moving distances, this suggests that the geographical location of students from poorly educated neighborhoods may steer them away from top paid programs. Overall, however, there is little correlation between distance to a program and the expected post-graduate wage in either group.¹⁶⁹ Nevertheless, students from poorly educated neighborhoods

third of programs fit into each category. Girls have a 22.3 percent chance of being admitted to a highly paid program on average; for boys, the corresponding number is 25.9 percent. Similarly, girls have a 21.7 percent chance of being admitted to a poorly paid program; for boys, this number is 20.6 percent chance. For programs paying between 2500 and 3000 euros, the average admissions probability is of similar magnitude for girls and boys. Looking at the underlying grade distributions show that men are significantly more likely to have a grade in the extended math curriculum and science, while women are significantly more likely to have a grade in social sciences/humanities and they score higher on the Finnish language exam.

¹⁶⁷ This is a *weighted* average over averages, i.e. your closest college is often rewarding several diplomas, each with their own mean wage. For each individual, we weigh these mean wages according to the size of that program.

¹⁶⁸ Overall, however, there is little correlation between distance to a program and the expected post-graduate wage in either group. Here we calculate the rank correlation (Spearman) between distance to a program and its average post-graduate wage. This correlation is calculated separately for each individual. Among students from highly educated neighborhoods, this correlation is 0.02 on average; among those from poorly educated neighborhoods, the correlation is 0.03 on average. When weighting programs by their size, these correlations change to -0.04 and 0.02 for applicants from highly and poorly educated neighborhoods, respectively.

¹⁶⁹ Here we calculate the rank correlation (Spearman) between distance to a program and its average post-graduate wage. This correlation is calculated separately for each individual. Among

are predicted to add somewhat more to their expected wage than students from highly educated neighborhoods, when equalizing the expected log distance to each application portfolio.¹⁷⁰ Furthermore, we observe students from poorly educated neighborhoods having significantly lower grades as measured by their average admissions probabilities, which also lowers their expected wages.

In summary, this analysis suggests that students from poorly educated neighborhoods are hindered by their endowments – rather than their application behavior – when the portfolio outcome of interest is expected wages. However, there are only small differences in expected wages to begin with. Data also suggests that girls choose portfolios giving them lower expected wages than boys, mainly due to their application strategy. For example, girls apply in a way that is consistent with being relatively risk averse.

students from highly educated neighborhoods, the correlation is 0.02 on average; among those from poorly educated neighborhoods, the correlation is 0.03 on average.

¹⁷⁰ Example: The wage gap between female students from highly and poorly educated neighborhoods is 61 euros when both are given the application behavior of students from poorly educated neighborhoods. When equalizing all expected log distances, this gap decreases to -11 euros. Similarly, the wage gap between male students from highly and poorly educated neighborhoods is 50 euros when both are given the application behavior of students from poorly educated neighborhoods. When equalizing all expected log distances, this gap decreases to 35 euros.

6.5 CONCLUSIONS

This chapter studies the college application strategies of Finnish high school graduates, and how these strategies differ between students depending on socioeconomic background. We find that students from poorly educated neighborhoods apply in a way that is consistent with having a stronger sensitivity to the moving distance, and a stronger aversion to risk. Among students from poorly educated neighborhoods, the relative risk aversion coefficient is estimated to 1.5; among those from highly, the corresponding estimate is 0.3. This pattern is driven by women; female applicants from poorly educated neighborhoods apply in a way that is consistent with being especially wage risk averse, while there is little difference in risk aversion coefficients among male applicants. These differences in application strategies can only partly be explained by differences in academic ability between the groups.

Despite these differences in application strategies, there are only quite small gaps in expected wages between students from poorly and highly educated neighborhoods. Furthermore, there is little support for the notion that students from poorly educated neighborhoods would have much to gain regarding expected wages, by taking on the application strategy of students from highly educated neighborhoods. The same is not true for the gender wage gap: When female students are 'given' the behavior model of male students, the gender wage gap is practically eliminated.

It is worth noting, however, that the adolescents under study here are a selected group; all of them have graduated from high school and all of them applied to college. Hence, the sample individuals are not representative of the full population of adolescents. It is quite possible that there are important differences in educational choices occurring at earlier ages, or at the stage when students decide on whether to apply.

References

- Black, S. E., & Devereux, P. J. (2010). *Recent developments in intergenerational mobility* (No. w15889). National Bureau of Economic Research.
- Chevalier, A. (2004). Parental education and child's education: A natural experiment.
- Dillon, E. W., & Smith, J. A. (2017). Determinants of the match between student ability and college quality. *Journal of Labor Economics*, 35(1), 45-66.
- Do, C. (2004). The effects of local colleges on the quality of college attended. *Economics of Education Review*, 23(3), 249-257.
- Dohmen, T., Falk, A., Huffman, D., & Sunde, U. (2018). On the relationship between cognitive ability and risk preference. *Journal of Economic Perspectives*, 32(2), 115-34.
- Frenette, M. (2006). Too far to go on? Distance to school and university participation. *Education Economics*, 14(1), 31-58.
- Griffith, A. L., & Rothstein, D. S. (2009). Can't get there from here: The decision to apply to a selective college. *Economics of Education Review*, 28(5), 620-628.
- Heller, D. E. (2004). Pell Grant recipients in selective colleges and universities. *America's Untapped Resource: Low-income Students in Higher Education*, 157-166.
- Hertz, T., Jayasundera, T., Piraino, P., Selcuk, S., Smith, N., & Verashchagina, A. (2007). The inheritance of educational inequality: International comparisons and fifty-year trends. *The BE Journal of Economic Analysis & Policy*, 7(2).
- Hill, C. B., Winston, G. C., & Boyd, S. A. (2005). Affordability family incomes and net prices at highly selective private colleges and universities. *Journal of Human Resources*, 40(4), 769-790.
- Hoxby, C., & Avery, C. (2013). The Missing "One-Offs": The Hidden Supply of High-Achieving, Low-Income Students. *Brookings Papers on Economic Activity*, 44(1 (Spring)), 1-65.
- Jäntti, M., Bratsberg, B., Roed, K., Raaum, O., Naylor, R., Österbacka, E., Björklund, A. & Eriksson, T. (2006). American exceptionalism in a new light: a comparison of intergenerational earnings mobility in the Nordic countries, the United Kingdom and the United States.
- McFadden, D. (1978). Modeling the choice of residential location. *Transportation Research Record*, (673).
- Mood, C. (2010). Logistic regression: Why we cannot do what we think we can do, and what we can do about it. *European Sociological Review*, 26(1), 67-82.
- Smith, J., Pender, M., & Howell, J. (2013). The full extent of student-college academic undermatch. *Economics of Education Review*, 32, 247-261.
- Solon, G. (1999). Intergenerational mobility in the labor market. In *Handbook of Labor Economics* (Vol. 3, pp. 1761-1800). Elsevier.
- Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge University Press.

Appendix

A.1 Measuring $q(S_i|k)$.

The probability of drawing subset S_i when taking portfolio k as given, and considering the selection process, is described by:

$$q(S_i|k) = \frac{q(C_i|k)}{Pr(unique|k)} = \frac{9! \cdot \prod_{\substack{j \in C_i \\ j \neq k}} f_j}{\pi_k \cdot (1 - f_k)^9}$$

where the numerator on the right-hand side is the probability of drawing the observed combination of portfolios, $C_i = S_i$, taking portfolio k as given and considering a selection process where repeated draws of the same portfolio is allowed. f_j is the relative frequency of students having chosen portfolio j in the student population.

The denominator is the probability of drawing a combination of ten unique portfolios when taking portfolio k as given, and considering a selection process where repeated draws of the same portfolio is allowed. This probability can be described as the product of two probabilities; the probability of drawing a set of nine unique portfolios conditional on portfolio k not being one of these (π_k), and the probability of portfolio k not being sampled on nine repeated draws, $(1 - f_k)^9$.

Here, we assume that π_k/π_j is sufficiently close to unity so that we can ignore π_k in the calculations without inducing any noticeable bias. (Note that, for our purposes, information regarding $q(S_i|k)/q(S_i|j)$ is sufficient.) Hence, $q(S_i|k)$ is approximated by:

$$\widehat{q(S_i|k)} = \frac{9! \cdot \prod_{\substack{j \in S_i \\ j \neq k}} f_j}{(1 - f_k)^9}$$

A.2 The distribution for portfolio wages

If post-graduate wages (w_n) are lognormally distributed, then portfolio wages (w_j) are approximately lognormal for the portfolios in the data. This is exemplified below (figure A1) showing the empirical distribution for w_j after 3,000,000 random draws of wages from four randomly chosen student portfolios, together with the corresponding lognormal approximation.

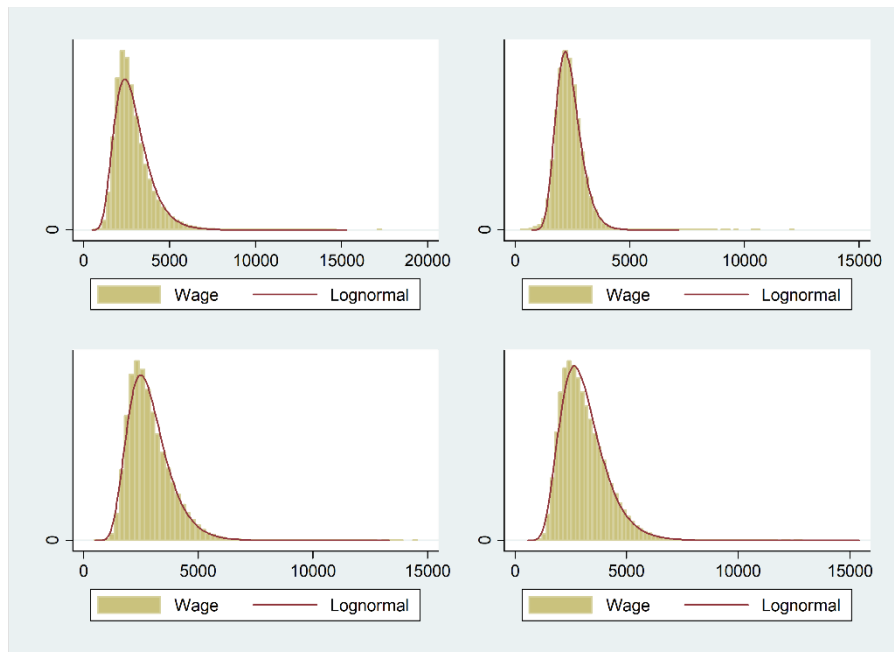


Figure A1. The portfolio wage distribution, four simulations

A.3 Geographical distributions

Figure A2 presents the distance distribution for students from poorly and highly educated neighborhoods, where the observations are weighted according to the size of that program (enrollments). Each student is represented 313 times, i.e. one observation per student-program-combination. The figures hint that students from highly educated neighborhoods are living in certain geographical 'islands', whereas the poorly educated neighborhoods are more scattered.

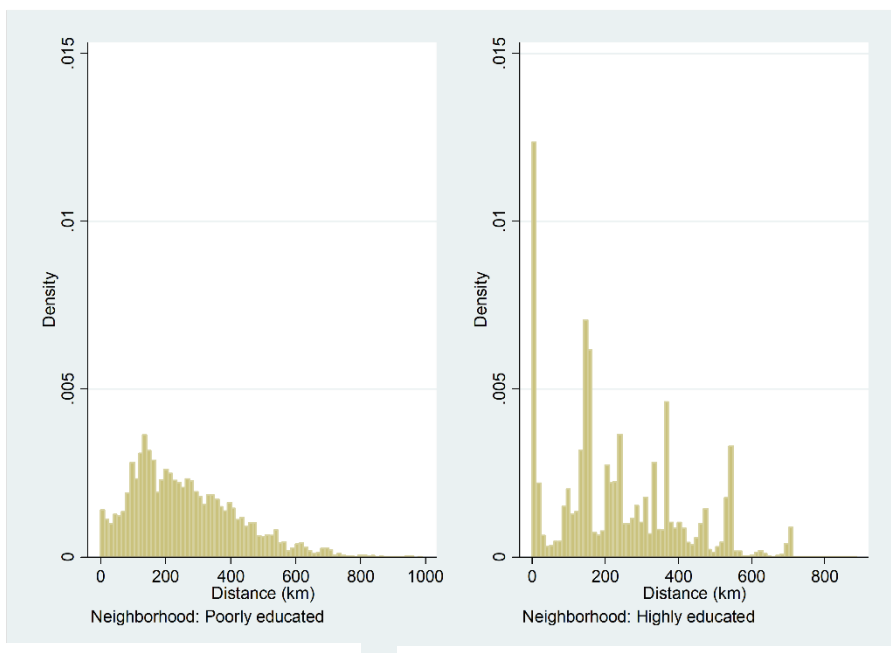


Figure A2. Weighted distance distributions, separately by neighbourhood education level

Note: An observation is one student-program combination. Weights proportional to the size of that program.

This pattern becomes clearer in figure A3 a) and b), which maps the number of students from poorly (a) and highly (b) educated neighborhoods per applicant in that municipality. The location of a program is depicted as a circle with size proportional to the number of students who enrolled. The figures show a distinct geographical division between the groups; of all applicants, students from highly educated neighborhoods make up a majority in only a handful of municipalities including Helsinki, while students from poorly educated neighborhoods make up a majority in roughly half of the municipalities of which practically all are scarcely populated and many lack an institution for higher education.

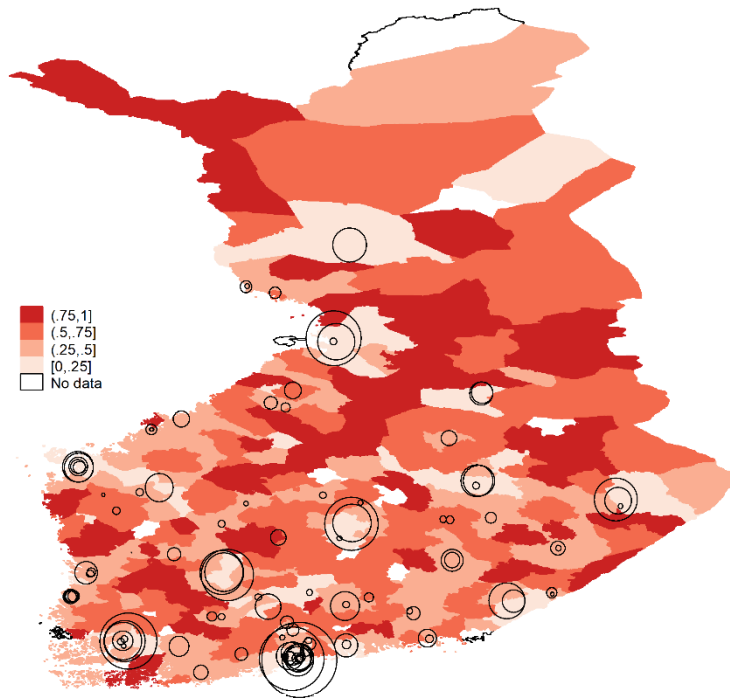


Figure A3a. The geographical distribution of students from poorly educated neighborhoods.

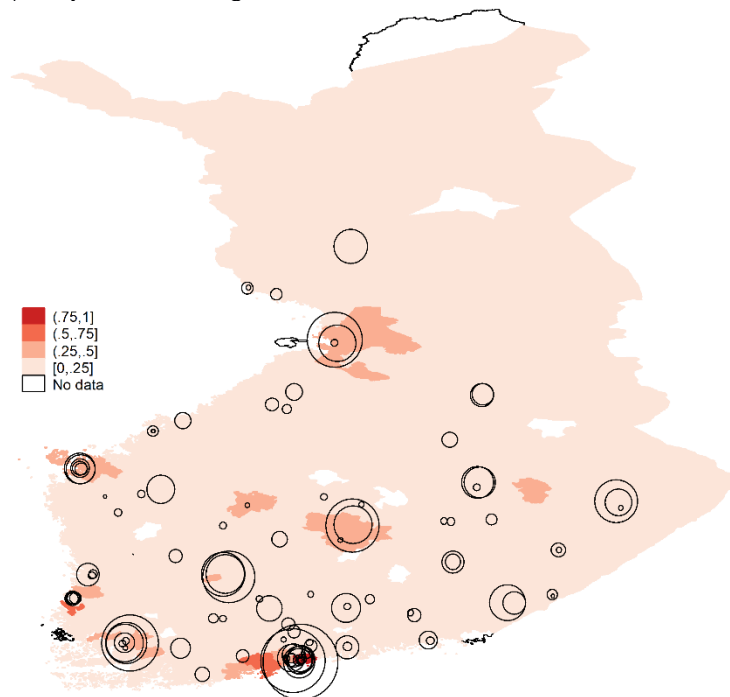


Figure A3b. The geographical distribution of students from highly educated neighborhoods.

Notes: A circle depicts an educational program, with size proportional to the number of students who enrolled.

A.4 Gender heterogeneity in portfolio choice

Table A1a. Conditional logit estimates for portfolio choice, separately by neighborhood education level. Sample: Men.

	Low		Medium		High		Difference in coefficients Chi2 (p-value)
	Coeff.	AME	Coeff.	AME	Coeff.	AME	
Expected log distance	-3.366** (0.179)	-0.194** (0.008)	-2.794** (0.097)	-0.173** (0.005)	-2.221** (0.097)	-0.145** (0.006)	37.30 (0.000)
Expected log wage	9.587** (2.040)	0.551** (0.127)	14.866** (1.649)	0.919** (0.101)	18.459** (2.284)	1.206** (0.136)	8.48 (0.014)
Log wage variance	6.239** (1.956)	0.359** (0.121)	8.235** (1.154)	0.509** (0.076)	10.773** (1.775)	0.704** (0.117)	3.19 (0.203)
Employment probability (unit: 10 %-points)	-0.525 (0.669)	-0.030 (0.042)	-1.576* (0.646)	-0.097* (0.040)	-0.734 (0.784)	-0.048 (0.055)	1.62 (0.444)
Expected female percentage (unit: 10 %-points)	-0.663** (0.181)	-0.038** (0.010)	-0.565** (0.150)	-0.035** (0.010)	-0.350 (0.197)	-0.023 (0.013)	1.45 (0.484)
Observations	17,290		35,790		19,590		
Students	1,729		3,579		1,959		

Notes: Standard errors in parenthesis. **p<0.01; *p<0.05. AME is short for the average marginal effect. The regressions also include controls for the probability of ending up in each of 18 fields.

Table A1b. Conditional logit estimates for portfolio choice, separately by neighborhood education level. Sample: Women.

	Low		Medium		High		Difference in coefficients Chi2 (p-value)
	Coeff.	AME	Coeff.	AME	Coeff.	AME	
Expected log distance	-3.531** (0.166)	-0.212** (0.012)	-2.663** (0.083)	-0.182** (0.004)	-2.125** (0.099)	-0.153** (0.006)	62.22 (0.000)
Expected log wage	14.686* (6.683)	0.882* (0.346)	7.923** (2.307)	0.541** (0.155)	11.093** (2.428)	0.797** (0.153)	1.53 (0.465)
Log wage variance	-17.957 (10.195)	-1.078 (0.531)	0.042 (2.061)	0.003 (0.145)	1.916 (1.604)	0.138 (0.109)	4.07 (0.131)
Employment probability (unit: 10 %-points)	-3.232* (1.494)	-0.194* (0.078)	-1.007* (0.424)	-0.069* (0.029)	-0.785 (0.549)	-0.056 (0.037)	2.35 (0.309)
Expected female percentage (unit: 10 %-points)	1.149** (0.150)	0.069** (0.008)	1.088** (0.119)	0.074** (0.007)	1.398** (0.194)	0.100** (0.013)	2.13 (0.346)
Observations	27,980		53,550		25,790		
Students	2,798		5,355		2,579		

Notes: Standard errors in parenthesis. **p<0.01; *p<0.05. AME is short for the average marginal effect. The regressions also include controls for the probability of ending up in each of 18 educational fields.

Table A2a. Ability-weighted conditional logit estimates for portfolio choice, separately by neighborhood education level.
Sample: Men

	Low		Medium		High		Difference in coefficients Chi2 (p-value)
	Coeff.	AME	Coeff.	AME	Coeff.	AME	
Expected log distance	-3.215** (0.183)	-0.188** (0.008)	-2.783** (0.099)	-0.172** (0.005)	-2.252** (0.105)	-0.147** (0.006)	25.39 (0.000)
Expected log wage	8.478** (2.024)	0.497** (0.132)	14.799** (1.644)	0.915** (0.106)	18.703** (2.434)	1.222** (0.146)	11.07 (0.004)
Log wage variance	5.661** (2.237)	0.332** (0.140)	8.303** (1.145)	0.514** (0.075)	10.372** (1.808)	0.678** (0.117)	2.73 (0.255)
Employment probability (unit: 10 %-points)	-0.303 (0.664)	-0.018 (0.041)	-1.577* (0.641)	-0.098* (0.046)	-0.569 (0.758)	-0.037 (0.052)	2.43 (0.297)
Expected female percentage (unit: 10 %-points)	-0.629** (0.183)	-0.037** (0.011)	-0.560** (0.149)	-0.035** (0.010)	-0.433* (0.199)	-0.028 (0.013)	0.54 (0.762)
Observations	17,290		35,790		19,590		
Students	1,729		3,579		1,959		

Notes: Standard errors in parenthesis. **p<0.01; *p<0.05. AME is short for the average marginal effect. The regressions also include controls for the probability of ending up in each of 18 fields.

Table A2b. Ability-weighted conditional logit estimates for portfolio choice, separately by neighborhood education level.
Sample: Women

	Low		Medium		High		Difference in coefficients Chi2 (p-value)
	Coeff.	AME	Coeff.	AME	Coeff.	AME	
Expected log distance	-3.337** (0.165)	-0.205** (0.010)	-2.671** (0.083)	-0.182** (0.005)	-2.219** (0.100)	-0.158** (0.006)	40.58 (0.000)
Expected log wage	9.742 (8.140)	0.600 (0.432)	7.683** (2.446)	0.524** (0.153)	10.429** (2.668)	0.745** (0.169)	0.61 (0.738)
Log wage variance	-7.638 (12.390)	-0.470 (0.680)	-0.201 (2.268)	-0.014 (0.146)	1.372 (1.952)	0.098 (0.140)	0.77 (0.679)
Employment probability (unit: 10 %-points)	-2.051 (1.821)	-0.126 (0.098)	-0.994* (0.450)	-0.068* (0.027)	-0.688 (0.594)	-0.049 (0.041)	0.56 (0.756)
Expected female percentage (unit: 10 %-points)	1.157** (0.151)	0.071** (0.009)	1.078** (0.121)	0.074** (0.008)	1.419** (0.212)	0.101** (0.014)	2.22 (0.330)
Observations	27,980		53,550		25,790		
Students	2,798		5,355		2,579		

Notes: Standard errors in parenthesis. **p<0.01; *p <0.05. AME is short for the average marginal effect. The regressions also include controls for the probability of ending up in each of 18 fields.

A.5 Different measures of risk attitudes and application behavior

For a smaller subsample we have access to student-level survey responses covering risk attitudes. The survey was conducted a couple of months before the actual application deadlines. 59 high schools were sampled and practically all graduate students participated, in total 3,437 students. In the current sample, we have data for 1,124 to 1,354 students depending on risk measure. The fallout is due to students not agreeing to being linked to the administrative data; students not being linked successfully, or missing values.

In this section we study how the survey responses regarding risk attitudes correspond to actual application behavior. The actual application behavior is estimated using the standard model, with one exception; we now include 50 (instead of ten) portfolios in each student's choice set as the survey sample is small enough to allow for large choice sets without inducing any computational difficulties.

Using the survey we construct three measures of risk attitude. The first one, labeled "labor market preferences", is the response to a choice between two hypothetical educational programs. The programs are equal in all respects except for the post-graduate labor market, with one having a higher mean wage as well as wage dispersion. The wage distributions are constructed so that students having a relative risk aversion of one would be indifferent between the choices. Roughly half of the students claim to prefer the 'risky labor market'. Students from highly educated neighborhoods and boys in particular are overrepresented in this group.

The second measure, labeled "risk attitude", measures students' subjective willingness to take risks on a scale from 1 to 10. Students at or below the median (7) are labeled 'safe'; those with higher values are labeled 'risky'. Girls and students with poorly educated parents are marginally overrepresented among the 'risky' according to this measure. The third measure, 'application strategy', is the response to a choice between four application strategies. The students respond to whether they plan to apply to their dream education; whether they plan to apply to the safest option; whether they plan to apply to both kinds of programs or if they have an alternative strategy. We label students choosing the first option (dream education) as having a 'risky' strategy and others as having a 'safe' strategy, with 40 percent of students having a risky strategy. Students with highly educated parents and boys are overrepresented in this group.

Overall, we find a clear pattern where students choosing 'risky options' in the survey also show an application behavior consistent with being less risk averse. Students who prefer a 'risky' labor market have an estimated risk coefficient at 0.2; the corresponding figure for those preferring a 'safe' labor

market is 2.7. Similarly, students having a below-median risk attitude have an estimated risk coefficient at 1.5 while those laying above the median have a coefficient at 0.3. These represent significant differences. Also, students who plan on choosing a risky application strategy have an estimated risk coefficient at 0.4 while the other group has a coefficient at 1.6. This, however, does not represent a significant difference.

A.6 Robustness

Table A3. Relative risk aversion estimates using alternative models

Sample	Main	Estimate Lowest	Highest
All	0.5 (0.3-0.8)	0.2 ^b (-0.0-0.4)	0.9 ^a (0.3-1.4)
<i>Neighborhood level</i>			
Low	1.5 (-0.4-2.5)	0.5 ^b (-0.5-1.4)	2.3 ^c (0.9-3.2)
Medium	0.5 (0.2-0.8)	0.2 ^b (-0.0-0.4)	0.8 ^a (0.2-1.3)
High	0.3 (0.1-0.5)	0.2 ^b (-0.1-0.4)	0.5 ^a (0.2-0.9)
<i>Gender</i>			
Women	1.3 (-0.6-2.1)	1.0 ^b (0.2-2.0)	11.8 ^c (2.1-73.9)*
Men	-0.2 (-0.5-0.1)	-0.4 ^b (-0.8--0.1)	-0.1 ^c (-0.4-0.1)

Notes: 95-percent confidence intervals in parentheses. *37.1 percent of the draws delivered nonsensical estimates – a distaste for higher wages – these draws were disregarded so that the final sampling distribution only contains possible risk aversion estimates. The lower limit is then given by the $[100/(100 - q)] \cdot 2.5^{\text{th}}$ percentile and the upper limit by the $100 - [100/(100 - q)] \cdot 2.5^{\text{th}}$ percentile; q is the percentage of nonsensical estimates on repeated draws. ^aExcludes the vector of variables describing the probability of ending up in a specific field (leaving only the probability of being admitted). ^bExcludes the employment probability. ^cExcludes the expected percentage of women on the post-graduate labor market (interacted with gender).

Table A4. Relative risk aversion estimates using alternative measures for the independent variables

Sample	<i>Alternative ways of measuring independent variables</i>			
	Main	Distance (50 km) ^a	Gender comp. ^b	Ranking of programs ^c
All	0.5 (0.3-0.8)	0.5 (0.2-0.8)	0.5 (0.2-0.7)	0.6 (0.2-0.9)
<i>Neighborhood level</i>				
Low	1.5 (-0.4-2.5)	1.5 (-0.4-2.5)	1.4 (-0.6-2.5)	1.4 (0.1-2.4)
Medium	0.5 (0.2-0.8)	0.5 (0.2-0.8)	0.5 (0.2-0.8)	0.6 (0.1-1.0)
High	0.3 (0.1-0.5)	0.3 (0.1-0.5)	0.3 (0.1-0.5)	0.4 (-0.0-0.8)
<i>Gender</i>				
Women	1.3 (-0.6-2.1)	1.3 (-0.6-2.2)	1.3 (-0.6-2.2)	2.4 (-2.7-4.6)*
Men	-0.2 (-0.5-0.1)	-0.2 (-0.5-0.1)	-0.2 (-0.5-0.1)	-0.2 (-0.5-0.1)

Notes: 95-percent confidence intervals in parentheses. *2.8 percent of the draws delivered nonsensical estimates – a distaste for higher wages – these draws were disregarded so that the final sampling distribution only contains possible risk aversion estimates. The lower limit is then given by the $[100/(100 - q)] \cdot 2.5^{\text{th}}$ percentile and the upper limit by the $100 - [100/(100 - q)] \cdot 2.5^{\text{th}}$ percentile; q is the percentage of nonsensical estimates on repeated draws. ^aThe distance to the ‘outside option’ (high school diploma only) is set to 50 km. ^bThe gender composition for those with a high school diploma only is set to 70 percent (for women) and 30 percent (for men). ^cThe preference order of the programs is randomized.

Table A5. Average marginal effects for the expected *log distance* using alternative measures for the independent variables (estimates only)

Sample	<i>Alternative ways of measuring independent variables</i>			
	Main	Distance (50 km) ^a	Gender comp. ^b	Ranking of programs ^c
All	-0.164	-0.155	-0.164	-0.154
<i>Neighborhood level</i>				
Low	-0.199	-0.191	-0.199	-0.185
Medium	-0.175	-0.167	-0.175	-0.164
High	-0.149	-0.138	-0.149	-0.140
<i>Gender</i>				
Women	-0.171	-0.160	-0.164	-0.160
Men	-0.160	-0.150	-0.155	-0.150

^aThe distance to the ‘outside option’ (high school diploma only) is set to 50 km. ^bThe gender composition for those with a high school diploma only is set to 70 percent (for women) and 30 percent (for men). ^cThe preference order of the programs is randomized.

Angela Djupsjöbacka

Human Capital

Formation, Maintenance and Transmission

Human capital, as measured by your text-based and numerical problem-solving skills, varies between countries as well as between cohorts of individuals from the same country. The forces driving these patterns are poorly understood, however. This thesis deals with topics related to the development and maintenance of human capital at different stages of life: schooling, labor market entry and, lastly, retirement. I argue that post-primary schooling has positive effects on cognitive performance later in life; that graduating in a bad economy may have some – although economically small – negative impacts on your future cognitive performance; I find no negative effects of retirement, but possibly a positive cognitive effect among men. Lastly, this thesis also deals with the transmission of human capital by studying college application strategies of high school students, and how these differ depending on socioeconomic background. The arguments I lay out in this thesis rely heavily on databased inference using econometric tools, and I also make a contribution to that toolbox by suggesting ways of dealing with ordinal data in a regression setting.