Second Thesis Proposal

Georgios Mantas
MSc. in Economics–Public Policy track

May 8, 2017

Abstract

In general, workers possess a bundle of skills and occupations are characterized as a bundle of tasks, although technological advance has affected fundamentally the structure of labour markets and the nature of tasks performed by workers. Technology seems to complement high-educated/skilled workers who perform nonroutine complex cognitive tasks, which are not easily automated. My aim is to measure, through a Mincerian empirical framework, the returns to skill (i.e. years of schooling) and complex cognitive tasks (i.e. ICT and influencing). I use OECD PIAAC data which permit me to examine differences not only across different subgroups, but also across different countries. In my opinion, international comparison of return to skill and complex cognitive tasks intensity is important due to the fact that worker's earnings may depend on a country's unique economic structure, labour market and social institutions.

Key words: Mincerian equation; Return to skills; Education; Task approach; Labour market; Earnings; International comparison; Heterogeneity.

Information table

Student Name	Georgios Mantas
Phone number	+31 61452528640
Supervisor	dr. L.B.D. Raes
ANR	245448
SNR	2002010
E–mail	G.Mantas@uvt.nl
Submission Date	May 8, 2017

Contents

1 Introduction

In this section, I provide a brief description of my project and I will try to highlight why the measurement of the returns to skill (i.e. years of education) and sophisticated cognitive tasks, such as the information and communications technology (ICT, henceforth) related and interactive or influencing tasks, constitutes a significant subject of research. My purpose is to examine the corresponding returns based on data from various countries around the world, and not to focus my analysis explicitly on panel data of a specific country, usually the United States in the relevant literature. However, measurements of human capital restricted mainly to workers in the United States do not give a clear view of the role of education, and skills to a greater extent, across different economies. Moreover, I describe the potential (feasible) empirical methodology that I can follow to create the causal link between earnings (e.g. hourly, monthly earnings), years of education and the aforementioned cognitive tasks. The methodology I adopt is originated from the extant literature, and particularly from the fundamental contributions of Mincer (1970, 1974). My expectations about the results are also affected by the literature, but this does not necessarily mean that the results of my empirical model would be in line with my expectations.

1.1 Working Title of my Thesis

"Returns to skill[1] and complex cognitive tasks around the world: Evidence from selected OECD countries"

1.2 Motivation

Returns to investment in education based on human capital theory have been estimated since the late 1950s. Based on these estimates, various empirical reviews, in turn, tried to establish detailed patterns (Psacharopoulos and Patrinos, 2004). In the same spirit, the skills of the population are regarded, in general, as the main aspect of the knowledge-based economies (Hanushek and Woessmann, 2009). However, extant evidence on the returns to skills is markedly limited, originating almost exclusively from the earnings of early-career workers in the United States(Hanushek et al., 2013). As a result, insights about the evolution of rewards to skills and how they differ across countries are less. Thus, the investment on human capital and its contribution on people's income constitutes a perpetual topic for societies and economic science.

The measurement of these returns is based mainly on Jacob Mincer's empirical establishment. In particular, Mincer (1970, 1974) found how wage differentials could be significantly explained by school attainment by on-the-job training investments. Mincer's insight was widely approved, however this led researchers to measurement errors, ignoring that standard Mincer formulation assumes that schooling is the unique systematic source of skills differences. Therefore, this suggests that the number of years of formal education alone is not a complete measure of skill and does not explain adequately the wage inequality/dispersion within education and other societal groups (Ingram and Neumann, 2006). Indeed, there are several reasons to suspect that the value of formal education may have changed. Firstly, the rapid technological progress affected considerably the structure of the labour markets and changed the nature of many occupations. The "new economy" view suggests that the technological change has led to higher compensation for jobs requiring, for example, greater verbal, logical and other cognitive skills (Ingram and Neumann, 2006). Secondly, a heated-debate exists among economists about the causal link between labour market earnings and estimated coefficients of skills (e.g. years of education, cognitive skills) through the properties of ordinary least square estimators. Card (2001) and Angrist and Pischke (2014, chap. 6) discuss the specific endogeneity problems extensively.

It is clear then, why the measurement of returns to skills constitutes a very interesting topic and there is space for further research and contribution. My aim is to examine the return to skill, via years of schooling, within the context of a Mincerian equation across different countries. My empirical analysis will follow Mincer's sophistication because, as Mincer (1974) showed, human capital consists of "general" (i.e. years of education) and "specific" (i.e. lifetime work experience). Therefore, it is prudent to include these aspects

[1] For simplicity, I use common but imprecise language interpreting the earnings impact of additional years of schooling as the return to schooling; see (Heckman et al., 2006).

when we measure/estimate individual's earnings. However, I seek to expand the analysis by inserting the role of complex cognitive tasks in order to see whether and how they affect labour earnings. My idea is inspired by three papers because they helped to realize the role of technological change, evolution of tasks across occupations and labour market returns that vary considerably and systematically among countries. Namely, these papers are written by: Autor et al. (2003), Hanushek et al. (2013) and Yamaguchi (2016). I describe these papers with more details section 2, in my brief review of the related literature. In addition, the book of Acemoglu and Autor (2011, chap. 12) provides a diligent review of the relevant literature, while highlights the implications of technological advance on the labour market and the role of tasks which workers perform across occupations.

As I mentioned in the my Introduction, I insert the complex cognitive task(or complex cognitive task intensity) in standard Mincer's equation, controlling for quite a few variables and other demographic characteristics. I believe this is a reasonable approach because people use their skills, which acquired from both schooling and work experience, to perform a bundle of specific tasks (Autor et al., 2003; Yamaguchi, 2016). In Subsection 1.4, I describe how I can insert the intensity of the aforementioned cognitive tasks and subsequently, how I can estimate the parameters of my interest.

In order to run my regressions, I use the Public Use Files (PUF, henceforth) of the Programme for the International Assessment of Adult Competencies (PIAAC, henceforth) developed by the Organisation for Economic Co-operation and Development (OECD, henceforth) through the Survey of Adult Skills. The specific survey measures adults' proficiency in key information-processing skills – literacy, numeracy and problem solving in technology-rich environments – and gathers information and data on how adults use their skills at home, at work and in the wider community. I will give further information about PIAAC in Subsection 1.5 where I describe other potential variables that I can use. Also, it has to be mentioned that OECD (2016) reader's companion and Stata user's guide of Pokropek and Jakubowski (2013) for PIAAC data , are quite useful to grasp the usefulness of the specific dataset and work properly with it.

1.3 Research Question

"How years of schooling (i.e. measure of skill) and complex cognitive tasks (i.e. ICT/programming and influencing/interactive tasks) affect worker's wage/earnings". PIAAC data provide the necessary variables to examine the specific research question. Using PIAAC I am in position to observe whether differences across countries and different age, education and various societal subgroups exist.

1.4 Potential Empirical Methodology

As I mentioned, my methodology is mainly based on Mincer's establishment as the vast majority of the relevant literature does. The algebraic form of my "naive" regression is the following:

$$\log y_i = \alpha_0 + \alpha_1 exp_i + \alpha_2 exp_i^2 + \alpha_3 edu_i + \alpha_4 zict_i + \alpha_5 zinfl_i + \alpha_6 age_i + \alpha_7 gender_i + \epsilon_i \tag{1}$$

- $logy_i \rightarrow$ is the gross hourly wage of individual i
- $exp \rightarrow$ is years of lifetime work experience. My regression model has a quadratic form with respect to exp similarly to Mincer's. In this way we can observe whether labour earnings are a concave function of work experience. In other words, if the marginal effect of an additional year of work experience is positive with decreasing impact.
- $edu \rightarrow$ measures the years of formal education
- $zict_i \rightarrow$ is the intensity/complexity of ICT-related tasks that individual i performs.
- $zinfl \rightarrow$ measures the intensity/complexity of influencing/interactive tasks. The higher this variable means that an individual is either a manager or holds a higher position. In other words, it is a good proxy for managerial skills.

According to OECD (2016), the definition of the specific tasks are the following:

- *ICT*: Use computer; e-mail; Internet for information; Internet to conduct monetary transactions; spreadsheets; word processing; write or prepare computer code; real-time discussions using Internet; overall level of computer use in terms of complexity.
- *Influencing*: Selling products or services; making speeches or presentations; advising; persuading or influencing others; negotiating; instructing, training or teaching others.

I add the letter z because I am planing to standardize these variables (mean zero and standard deviation one) in order to make the analytical interpretation of the empirical model easier.

- $age \rightarrow$ represents years of age
- $gender \rightarrow$ is a dummy variable that works as a gender indicator, which take the value 0 for males and 1 for females.
- $\epsilon \rightarrow$ is a stochastic term (or a error term) which represents idiosyncratic earning differences, generally assumed in empirical analyses to be orthogonal/independent to human capital, in this case years of education (edu) and work experience (exp).

Eq. (1) depicts the simplest form of my model, which can be expanded by simply including countries and country fixed-effects. The regression pooled model gains the ability to check heterogeneity of returns across countries and at the same estimate preciser the coefficients of the parameters of my interest. The algebraic form of the regression model becomes:

$$\log y_{ic} = \alpha_0 + \eta_c + \alpha_1 exp_{ic} + \alpha_2 exp_{ic}^2 + \alpha_3 edu_{ic} + \alpha_4 zict_{ic} + \alpha_5 zinfl_{ic} + \alpha_6 age_{ic} + \alpha_7 gender_{ic} + \epsilon_{ic} \quad (2)$$

Where c indicates the country and η_c country-fixed effects, respectively. Intriguingly, PIAAC data contain codes for the type industry and occupation in both one- and two-digit form. I can exploit these observations to expand the explanatory power of my model and check its robustness by including industry and occupation fixed-effects. To be precise, I can possibly include them with two ways. The first way is to include them individually, while the second one is to insert them as an interaction term.

$$\begin{aligned}\log y_{icdp} = \alpha_0 + \eta_c + &\mu_d + \nu_p + \alpha_1 exp_{icdp} + \alpha_2 exp_{icdp}^2 + \alpha_3 edu_{icdp} + \\ &\alpha_4 zict_{icdp} + \alpha_5 zinfl_{icdp} + \alpha_6 age_{icdp} + \alpha_7 gender_{icdp} + \epsilon_{icdp}\end{aligned} \quad (3)$$

or

$$\begin{aligned}\log y_{icdp} = \alpha_0 + \eta_c + &\mu_d * \nu_p + \alpha_1 exp_{icdp} + \alpha_2 exp_{icdp}^2 + \alpha_3 edu_{icdp} + \\ &\alpha_4 zict_{icdp} + \alpha_5 zinfl_{icdp} + \alpha_6 age_{icdp} + \alpha_7 gender_{icdp} + \epsilon_{icdp}\end{aligned} \quad (4)$$

μ_d denotes the industry fixed-effects and d industry's code. Similarly, ν_p denotes the occupation fixed-effects and p occupation's code.

Furthermore, I have to enrich further my model with more dummy variables or indicators in order to unveil heterogeneity across different subgroups. A few indicative examples are:

1. Additional (un)official job-related training: Dummy indicator, seminars and hours of training
2. Age group indicators: entry-age (25-34), prime-age (35-54) and exit-age (55-65)
3. Types of employment: i) Public sector vs. Private sector, ii) Full-time vs. Part-time
4. Societal groups: i) Natives vs. Migrants, ii) Native vs. Non-native speakers
5. Household characteristics: i) Parental highest level of education (either mother's or father's, or both of them) ii) Number of children

6. Other interesting variables that PIAAC contain: i) Perception of respondents about their health, ii) indicator if respondents needed help to clarify survey's questions, iii) place of interview, iv) if another person was around during the time of the interview (quite interesting value which can show us if peer effects existed)

   Thus, all things considered, the final form of my econometric model will be:

$$\begin{aligned} \log y_{icdp} = \alpha_0 + \eta_c + \mu_d * \nu_p + \alpha_1 exp_{icdp} + \alpha_2 exp^2_{icdp} + \alpha_3 edu_{icdp} + \\ \alpha_4 zict_{icdp} + \alpha_5 zinfl_{icdp} + \alpha_6 age_{icdp} + \alpha_7 gender_{icdp} + \beta X + \epsilon_{icdp} \end{aligned} \tag{5}$$

   Where X represents the control variables of the various subgroups, as I described above.

"Endogeneity issues": After building my regression model (Eq. (5)), I have to consider carefully whether endogeneity problems exists from omitted variables. For instance, years of education might not represent skill level accurately because the quality of education or the ability of respondents may vary across counties. For this case, I can use as instruments the skill measurements of cognitive skills in three domains(i.e. literacy, numeracy, and problem solving in technology-rich environments) that PIAAC survey contains. These instrumental variables may help me to isolate the exogenous variation of education in my model because the assessments of cognitive skills originate from representative samples and internationally harmonized background questionnaire (Hanushek et al., 2013; OECD, 2016). Plausibly another problem might be induced by the actual work experience because it might enter, in part, endogenously to skill levels. Even though, Hanushek et al. (2013) find that their results remained robust and hardly unchanged by using potential experience (i.e. age minus year of schooling minus six) instead of actual experience. I enlist the available "instruments" in Table; Subsection 1.5.

1.5 General information for PIAAC and Table with potential variables

PIAAC data description: Developed by the OECD to provide internationally comparable data on adult skills (OECD, 2013, 2016). The first round of PIAAC data, administered to 24 countries between August 2011 and March 2012, and the second round, administered to an additional nine countries between April 2014 and March 2015, provide comparable skill data for 32 countries. To date, two 'rounds' of the survey have been undertaken. The countries participating in Round 1 were: Australia, Austria, Belgium, Canada, the Czech Republic, Cyprus, Denmark, Estonia, Finland, France, Germany, Ireland, Italy, Japan, Korea, the Netherlands, Norway, Poland, the Russian Federation, the Slovak Republic, Spain, Sweden, the United Kingdom and the United States. In Belgium, data was collected in the Flanders region only. In the United Kingdom, two of the four devolved administrations participated in the survey: England and Northern Ireland. In Russian Federation, as OECD (2013, 2016) underlines, the results relate to the territory of the Russian Federation excluding the Moscow municipal area. Moscow was excluded after the data collection had been completed due to problems with a data collection in this area. In Round 2 of the survey, the participating countries were: Chile, Greece, Indonesia, Israel, Lithuania, New Zealand, Singapore, Slovenia and Turkey. In Indonesia, data was collected only in the Jakarta municipal area. PUF PIAAC data for the nine Round-2 countries and revised Public Use Files for Round-1 countries with updated data were released on 28 June 2016. The updates to the Round 1 data files include the rescaling of the skills use indices and the recalculation of the derived earning variables.

The survey included an assessment of cognitive skills in three domains: literacy, numeracy, and problem solving in technology-rich environments. The tasks respondents had to solve were often framed as real-world problems, such as maintaining a driver's logbook (numeracy domain) or reserving a meeting room on a particular date using a reservation system (problem- solving domain). PIAAC measures each of the three skill domains on a 500-point scale. The domains, described rigorously in OECD (2013, 2016), refer to key information-processing competencies and are defined as:

1. *Literacy*: 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;

2. *Numeracy*: 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 in adult life;

3. *Problem solving in technology-rich environments*: ability to use digital technology, communication tools and networks to acquire and evaluate information, communicate with others and perform practical tasks.

Variable	Purpose	Database
Labour marker earnings: Hourly wage (+ bonus); Monthly wage (+ bonus)	Dependent variable of my empirical methodology	OECD–PIAAC
Lifetime work experience; Additional (un)official job-related training	Control variable	OECD–PIAAC
Total years of schooling	Control variable; Return to skill index	OECD–PIAAC
ICT task complexity	Control variable; Additional parameter of interest	OECD–PIAAC
Influencing task intensity/ frequency	Control variable; Additional parameter of interest	OECD–PIAAC
Task discretion in workplace	Control variable; Potential parameter of interest	OECD–PIAAC
Literacy	Potential "instrument" for years of schooling	OECD–PIAAC
Numeracy	Potential "instrument" for years of schooling	OECD–PIAAC
Problem solving in technology-rich environments	Potential "instrument" for years of schooling	OECD–PIAAC
Type of employment	Control variable; Observe heterogeneity	OECD–PIAAC
Household characteristics	Control variables; Observe heterogeneity	OECD–PIAAC
Interesting subgroups: Natives/ Migrants; Native/ Non-native speakers	Control variables; Observe heterogeneity	OECD–PIAAC
Self-reported health condition; indicator if clarification needed; place of interview; indicator if another person was present	Control variables; Observe Heterogeneity; Insights about the potential behavioural aspects of PIAAC survey	OECD–PIAAC
Trade union density	Additional control variable for robustness as interaction term	OECD–Labour Statistics
Employment Protection Legislation (EPL)	Additional control variable for robustness as interaction term	OECD Indicators of Employment Protection
Product Market Regulation (PMR)	Additional control variable for robustness as interaction term	OECD indicators of Product Market Regulation
Public sector size: share of workers employed in the public sector	Additional control variable for robustness as interaction term	OECD–PIAAC; OECD: Government at a Glance 2015 and ILO
GDP per capital growth in recent years	Additional control variable for robustness as interaction term	OECD–Productivity Statistics and World Bank *Databank*
ICT-related statistics	Additional control variable for robustness as interaction term	OECD Data–ICT
Number of patents	Additional control variable for robustness as interaction term	U.S. PATENT AND TRADEMARK OFFICE
Enterprises by size class	Additional control variable for robustness as interaction term	OECD Structural and Demographic Business Statistics (SDBS) Database

1.6 Expected Results

Insofar, the expected results are in line with the literature. Thus, the results may be the following: 1) With country fixed-effects, returns to skills (i.e. years of schooling) will be positive – not higher than 10% – and differ considerably across countries. 2) With country fixed-effects, the parameters which represents the complex cognitive skills will plausibly be statistically significant and may differ substantially across countries due to differences in economic fundamentals and labour markets' structure. 3) Including also industry and occupation fixed-effects, the estimated coefficients will remain relatively unchanged and the corresponding variables, especially the parameters of interest depicted in either Eq. (3) or Eq. (4), will remain statistically significant. 4) If I verify that my model suffers from endogeneity (potential sources: omitted variable bias, measurement error, simultaneity or reverse causality), then the return to "instrumented" skill will be higher, on average, and different between countries including all fixed-effects. 5) All fixed-effects included, when I "instrument" years of education might affect the estimated coefficients and statistical significance of the complex cognitive tasks (i.e. other parameters of interest). 6) Finally, using the parameters I enlist as "additional control variables for robustness as interaction terms" in Subsection 1.5 Table, I may find that return to skill and tasks depend also on a country's specific labour-market and social institutions.

2 Literature Review

Human capital and the impact of schooling on labour market earnings is a topic that has been studied rigorously from many researchers. In particular, starting from Mincer (1958), researchers were primarily interested in wage dispersion and inequality between different education groups. However, gradually economists fathomed that the technological advance affects the structure of labour markets, nature of occupations and wage patterns.

Tinbergen (1974) finds that the wage equality depends on the technological progress and at the same time, on the 'race' between the supply and demand of high educated/skilled workers. In particular, he suggests that the reduction in inequality found for the last century can be resumed after the stagnant period from 1950 to 1970, depending on the 'race' between demand for third-level manpower due to technological development and supply of it due to increased schooling. Tinbergen (1974) bases his findings on cross-section material for the 28 most populous American states. This is a pattern that is prominent in the relevant literature and the vast majority of researchers uses panel data by combining representative data on job task requirements from the Dictionary of Occupational Titles (DOT) with samples of employed workers from the Census and Current Population Survey (CPS) to form a consistent panel of industry and occupational task input.

Ingram and Neumann (2006) based on the fact that, since 1975, increases in the return to skill (measured by years of education), in the percentage of the labour force that is skilled, and in the variance of wage income within skill categories have characterized the U.S. labour market; argue that education per se does not measure skill adequately, and subsequently suggest an alternative measure based on the observed skill characteristics of the job. After analysing the return to various dimensions of skill, including formal education and accounting for other elements of skill, they find that the return to years of education has been constant since 1970. Furthermore, variations in direct measures of skill account for a considerable fraction of the increased dispersion in income among the college educated, and some of the augment in wage dispersion among those who have not earned a college degree. Another interesting fact of the specific paper is that explains why the conventional rate of return on skill is, interpreted as the regression coefficient on years of schooling, is biased. At first, schooling is endogenously determined–that is schooling and stochastic term (ϵ) are not orthogonal)– and, secondly, some other factors or parameters (i.e. control variables) are functions of school.

The creation of a credible causal link between years of schooling and labour market earnings constitutes a heated-topic for Econometric's society. The paper written by Card (2001) provides an insightful review of a set of studies that have attempted to measure the causal effect of education on labour market earnings by using institutional features of the supply side of the education system as exogenous determinants of schooling outcomes. In other words, the basic idea underlying this new thrust of research is that institutional features of the education system can be used to form credible instrumental variables for individual schooling outcomes that can cut through the Gordian Knot of endogenous schooling and unobserved ability. Card (2001) finds that instrumental variables for completed education reveal that the resulting estimates of the return to schooling are typically as big or bigger than the corresponding ordinary least squares estimates. One potential interpretation of this finding is that marginal returns to education among the low-education subgroups generally affected by supply-side innovations tend to be relatively high, reflecting their high marginal costs of schooling, rather than low ability that limits their return to education.

The debate around the causality issues between years of schooling and earnings is prominent and it can be observed to contemporary papers that try to approximate the returns to skills with alternative ways. Still, the research interests of some economists remain on the wage dispersion/ inequality between different subgroups, such as men and women (i.e. gender wage gap). Beaudry and Lewis (2012) focus on the wage differentials between men and women in the United States. In particular, they observe that over the 1980s and 1990s the wage differentials between men and women–intriguingly with similar observable characteristics–declined significantly while at the same time, the returns to education increased. The literature suggests that these two trends may reflect a common change in the relative price of a skill which is more abundant in both women and more educated workers. The authors test the relevance of this hypothesis by examining the cross-city co-movement in both male-female wage differentials and returns to education over the 1980-2000 period. They show male-female wage differentials at the city levels moved in opposite direction to the changes in the return to education. Furthermore, they find this relationship to be particularly strong when they isolate data variation which most likely depicts the effect of technological change on relative prices. Interestingly, they carefully control for potential selection issues which could bias both their estimations and interpretation, to a greater extend. All in all, their cross-city estimates suggest that most of the aggregate reduction in the male-female wage differential observed over the 1980-2000 period was more likely due to a change in the relative price of skill that both females and educated workers have in profuse.

In the same spirit of wage inequality, Autor (2014) argues that public debate focus singularly on the "top 1 percent" of households and neglects the components of earnings inequality that is arguably most consequential for the "other 99 percent" of citizens. More briefly, the dramatic growth in the wage premium associates with higher education and cognitive ability. Autor (2014) documents the fundamental role of both the supply and demand for skills in shaping inequality, discusses why skill demands have persistently risen in industrialized countries, and considers the economic value of inequality alongside its potential social costs. The author concludes by underlining the constructive role for public policy in fostering skills formation and preserving economic mobility. In particular, Autor (2014) suggests that most effective policies in augmenting prosperity and reducing inequality are those that cultivate the skills of successive generations. For instance, excellent preschool through high school education; broad access to post-secondary education; and good nutrition, good public health, and high-quality home environments. Such policies could address inequality from two directions: (i) enabling a larger fraction of adults to attain high productivity, and well-paid jobs to make a reasonable living; and (ii) raising the total supply of skills available to the economy, which in turn eases the skill premium and mitigate inequality. In my opinion, this Review is quite interesting because presents a few public policies on how to decrease wage inequality, while other papers just analyse the subject without proving any practical solution or actual policies.

Deming (2015), taking into consideration the massive computerisation and technological advance, expands his analysis from the context of wage inequality and measures how the labour market rewards specific skills. Deming's motivation comes for the fact that the slow growth of high-paying jobs in the U.S. since 2000 and rapid evolution of computer technology have sparked fears that human labour will eventually be considered obsolete. Yet while computers perform demanding cognitive tasks of rapidly increasing complexity, simple human interaction has proven difficult to automate. Deming (2015) shows that employment and wage growth has been strongest in jobs that require high levels of both cognitive and social skill. Additionally, based on data from the NLSY79[2], he finds that the female hold a relative advantage in social skills and this may have

[2] National Longitudinal Survey of Youth 1979 (NLSY79). The author study changes in the the task content of work using data from the Occupational Information Network (O*NET). O*NET is a survey administered by the U.S. Department of Labor to a random sample of U.S. workers in each occupation. O*NET data are used also by Autor et al. (2003).

played some role in the narrowing of gender gaps in labour market outcomes since 1980.

Rapid technological development and computerisation affected structurally labour markets and changed tasks performed by workers radically. Based on that fact, researchers try to investigate wage differential (or wage premium) and returns to skills through a task-based approach. One influential paper that follows a task approach and focuses on skill content of recent technological change, is the one by Autor et al. (2003). The authors try to grasp how computerization alters job skill demands. They argue that computer capital (1) substitutes for workers in performing cognitive and manual tasks that can be accomplished by following explicit rules which can be easily automated; and (2) complements workers in performing nonroutine problem-solving and complex communications tasks. Given the assumption that these tasks are imperfect substitutes, their model implies measurable changes in the composition of job tasks, which they explore using representative data on task input for 1960 to 1998. Autor et al. (2003) show that within industries, occupations, and education groups, computerization is associated with reduced labour input of routine manual and routine cognitive tasks and increased labour input of nonroutine cognitive tasks. When the authors translate task shifts into education demand, the model can explain 60 percent of the estimated relative demand shift favouring college labour during 1970 to 1998. In addition, task changes within identical occupations account for almost half of this impact.

Yamaguchi (2016) also uses a task approach to study how skilled-biased technology reduces wage inequality between men and women. More precisely, to answer that question the author constructs a task-based Roy model in which workers possess a bundle of basic skills, whilst occupations are characterized as a bundle of basic tasks. The model is estimated using the task data from the Dictionary of Occupational Titles and the PSID[3] (Panel Study of Income Dynamics). The main empirical finding is that men have more physical skills than women, but the returns to physical skills have dropped significantly leading to substantial decline gender wage gap from 1980 to 2000. During the period of 2000-2010, the returns to physical skills were stagnated, but the estimates suggest that the faster growth of women's cognitive skills compared to men's, the lower the gender wage is.

Apparently, the relevant literature is replete with studies about returns to skills, wage and task distribution for the United States. This is insightful to delineate wage patterns with respect to human capital, although this provides a distorted picture of the role of skills across different economies because U.S. labour market cannot be representative for other countries around the world. Hence, using cross-sectional data with respondents from a variety of countries will give us a more transparent view on how skills and tasks are rewarded through international comparison.

Perhaps, someone might think that even rich cross-sectional data are not useful due to the fact that we are not able to derive any conclusion about the lifetime earnings or return to skills, as we can with panel data. However, this not entirely correct. Björklund (1993) and Haider and Solon (2006) prove that current earnings constitute a good proxy for lifetime earnings when observed for individuals aged around 35 to 45.

Hanushek et al. (2013) exploit the first round PIAAC data of OECD for 22 countries within a Mincerian empirical context. Initially, they find that the concentration on U.S. data leads to underestimating the lifetime returns to skills by about one quarter. Particularly, the authors estimate that, on average, a one-standard-deviation increase in numeracy skills is associated with an 18 percent wage increase among prime-age[4] workers. However, substantial heterogeneity exists across countries. Namely, eight countries, including all Nordic countries, have returns between 12 and 15 percent, while six are above 21 percent with the highest return being 28 percent in the United States. Authors' estimates are markedly robust to different earnings and skill measures, additional controls, and various subgroups. Interestingly, returns to skills are most likely lower in countries with higher union density, stricter employment protection, and larger public-sector shares.

At last, another paper by Pipien and Roszkowska (2015) proceeds with an empirical analysis using cross-section series taken from the European Union Structure of Earnings Survey (SES), which is a large representative enterprise sample survey. The SES provides comparable information on the level of remuner-

[3]The PSID is a nationally representative household panel survey that began in 1968. To study the evolution of tasks, skills, and wages, I draw a sample of household heads and wives. The PSID contains information on hours of work and labour income in the last year, occupation, demographic variables, years of education, and an indicator for whether the job is covered by a union contract or not.

[4]35–54 years old

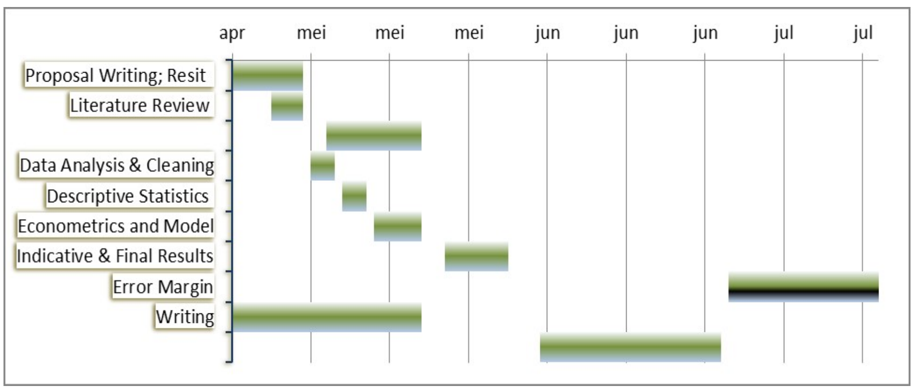
ation and characteristics of employees such as sex, age and occupation. The SES data are representative and contain information taken from enterprises with at least 10 employees operating in all areas of the economy except public administration. They find that estimated returns to education rate vary from 17% in Scandinavian countries to 40% and more in Southern Europe countries. The authors conclude to a generalization that countries with low estimated returns to education can be characterized by higher labour force participation rates, better educated population, higher public expenditures on education and lower dispersion of wages.

2.1 Gantt Chart

Important Deadlines:

- Proposal Resit: 8/5/2017
- Exams; Resit Period: 8–15/05
- Thesis Submission: 30/6/2017
- Defence: Before 15/7/2017
- Thesis Submission; Resit: Before 31/8
- Meetings with supervisor: Not known a priori, although they can be adjusted with punctual communication with my supervisor, as we have already discussed.

Gantt chart



My Gantt chart depicts approximately how many days I will invest on each task for the purpose of completion of my thesis. However, this Gantt chart is not definite, but it is based on my current expectations and it might change. In this case, I should inform my supervisor in advance. The **"error margin"** that I included, represents the additional days I may need to complete my thesis, beyond the initial submission deadline. I presume that I will possibly use these days because of the fact that I lost substantial amount of time to write a second proposal on a different topic. In particular, the exact dates and the duration (in days) for each task are the following:

- Proposal Writing; Resit: 29/4–8/5; Duration: 9 days
- Literature Review: 4–8/5 & 11–23/5; Duration: 16 days
- Data Analysis & Cleaning: 9–12/5; Duration: 3 days
- Descriptive Statistics: 13–16/5; Duration: 3 days
- Econometrics and Model: 17–23/5; Duration: 6 days
- Writing: 29/4–23/5 & 7–30/6: Duration: 47 days
- Error Margin: 1–20/7; Duration: 19 days

2.2 Work Plan & Remarks

1. Organise better the papers and books that I need for my thesis. In this way, I will be able to write my literature review quicker and complete the materials I need for my topic.

2. Download and clean the PUF PIAAC data as soon as possible from OECD website. Fortunately, I have been elaborated with PIAAC data of Round 1. I found and downloaded the corresponding dataset in an organized Stata form here. Thus, I had the opportunity to grasp the meaning of the multiple variables and examine the regression models I described in Subsection 1.4.

   However, on 28 June 2016 OECD published the PUF PIAAC for the 9 Round-2 countries and revised the relevant data for Round-1 countries. In addition, OECD provides – individually for countries participating in the Survey of Adult Skills– PUF PIAAC in SAS, SPSS and Excel format, but the Excel form contains all the information in the first cell and I am not able to insert them in Stata. Converting the data from SPSS to Stata form will take some time.

3. After converting the data to Stata form, I have to check which variables I need and after start to merge the data of each country to a common dataset. The advantage of my thesis is that PUF PIAAC of Round 2 are available and I extend the analysis of Hanushek et al. (2013) for other countries.

4. Round 1 PUF PIAAC do not contain the continuous labour earnings data for Austria, Canada, Germany, Sweden and the United States. I must find whether the data of the countries participated in Round 2 suffer from the same problem. I may need to contact with the Statistical Authorities of the corresponding nations in order to get access to their PIAAC data.

5. I may need to use average annual exchange rates and convert all the earnings variables to a single currency, because the worker's wages are reported in their national currency. Perhaps, I should do that because I am not sure if these differences are absorbed by country fixed-effects. If not, this may distort the size of my estimations.

6. In Subsection 1.4, I mentioned that my model might suffer from endogeneity. For that reason, I have to be prudent which variables I use, how I am including them in my model and which "instruments to exploit" (see Table, Subsection 1.4). Thus, I should consider other ways to insert the complex cognitive tasks. For instance, I can add *zict* and *zinfl* and after standardise them for each country (mean zero and standard deviation one). The new variable can be named as $zCCT$. In addition, I can also insert the standardised task discretion (see Table, Subsection 1.4). Unfortunately, I am not able to insert an indicator for physical tasks' complexity because OECD removed the specific variable from the revised version of PUF PIAAC. Then, Eq. (5) will become:

$$\begin{aligned} \log y_{icdp} = \alpha_0 + \eta_c + \mu_d * \nu_p + \alpha_1 exp_{icdp} + \alpha_2 exp^2_{icdp} + \alpha_3 edu_{icdp} + \\ \alpha_4 zCCT_{icdp} + \alpha_5 zdisc_{icdp} + \alpha_6 age_{icdp} + \alpha_7 gender_{icdp} + \beta X + \epsilon_{icdp} \end{aligned} \tag{6}$$

7. Finally, in order to check for various labour market institution and economic fundamentals I have to develop a regression model with interaction terms. I enlist a few potential variables in Table; Subsection 1.4. Hence, the algebraic form of Eq. (5) can become:

$$\begin{aligned} \log y_{icdp} = \alpha_0 + \eta_c + \mu_d * \nu_p + \alpha_1 exp_{icdp} + \alpha_2 exp^2_{icdp} + \alpha_3 edu_{icdp} + \gamma edu_{icdp} * E_c + \alpha_4 zict_{icdp} + \\ \delta zict_{icdp} * Z_c + \alpha_5 zinfl_{icdp} + \xi zinfl_{icdp} * F_c + \alpha_6 age_{icdp} + \alpha_7 gender_{icdp} + \beta X + \epsilon_{icdp} \end{aligned} \tag{7}$$

8. Except of endogeneity, my model may suffer from heteroskedasticity. Thus, I should consider carefully whether robust standard erros are needed in my empirical analysis.

2.3 Contingency Table & Risks

Risk	Likelihood	Severity	Solution
Not enough data	Low	High	PIAAC provide sufficient amount of data, but, in case of emergency, I can change the econometric model, use an alternative design, keep more countries.
Not capable to model returns to education and tasks "as good as exogenous"	Moderate	Moderate	Revise my notes from Econometrics and check the theory again, try to exploit variables of PIAAC as "instruments", read carefully the relevant literature, change modelling of the parameters interest or remove parameters from the analysis and check again control variables and fixed-effects.
Not capable to design the econometric model	Low	Moderate	Consult colleagues in Master's; PhDs from the Econometric courses; professors and supervisor.
Results are not statistically significant	Moderate	Low	Check again the relevant literature. It is not uncommon to end up with insignificant results even if the econometrics behind the model are correct.
Writer's block	High	Low	Just relax and take a brake. Brainstorming with other people will be useful.
Problems with writing structure	Moderate	Low	Check the course materials of Applied Economic Analysis 2, imitate how other authors write in the economic literature and try to develop further my *repertoire*.

References

Acemoglu, D. and Autor, D. (2011). Skills, tasks and technologies: Implications for employment and earnings. *Handbook of labor economics*, 4:1043–1171.

Angrist, J. D. and Pischke, J.-S. (2014). *Mastering'metrics: The path from cause to effect*. Princeton University Press.

Autor, D., Levy, F., and Murnane, R. J. (2003). The skill content of recent technological change: An empirical exploration. *The Quarterly Journal of Economics*, 118(4):1279–1333.

Autor, D. H. (2014). Skills, education, and the rise of earnings inequality among the "other 99 percent". *Science*, 344(6186):843–851.

Beaudry, P. and Lewis, E. (2012). Do male-female wage differentials reflect differences in the return to skill? Cross-city evidence from 1980-2000. Technical report, National Bureau of Economic Research.

Björklund, A. (1993). A comparison between actual distributions of annual and lifetime income: Sweden 1951–89. *Review of Income and Wealth*, 39(4):377–386.

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

Deming, D. J. (2015). The growing importance of social skills in the labor market. Technical report, National Bureau of Economic Research.

Haider, S. and Solon, G. (2006). Life-cycle variation in the association between current and lifetime earnings. *The American Economic Review*, 96(4):1308–1320.

Hanushek, E. A., Schwerdt, G., Wiederhold, S., and Woessmann, L. (2013). Returns to Skills around the world: Evidence from PIAAC. Technical report, National Bureau of Economic Research.

Hanushek, E. A. and Woessmann, L. (2009). Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation. Technical report, National Bureau of Economic Research.

Heckman, J. J., Lochner, L. J., and Todd, P. E. (2006). Earnings functions, rates of return and treatment effects: The mincer equation and beyond. *Handbook of the Economics of Education*, 1:307–458.

Ingram, B. F. and Neumann, G. R. (2006). The returns to skill. *Labour Economics*, 13(1):35 – 59.

Mincer, J. (1958). Investment in human capital and personal income distribution. *Journal of political economy*, 66(4):281–302.

Mincer, J. (1970). The distribution of labor incomes: A survey with special reference to the human capital approach. *Journal of Economic Literature*, 8(1):1–26.

Mincer, J. (1974). *Schooling, Experience, and Earnings*. National Bureau of Economic Research, Inc.

OECD (2013). OECD skills outlook 2013: First results from the survey of adult skills. *OECD Publishing, Paris.*

OECD (2016). The survey of adult skills: Reader's companion, second edition. *OECD Publishing, Paris.*

Pipien, M. and Roszkowska, S. (2015). Returns to skills in europe–same or different? The empirical importance of the systems of regressions approach. Technical report, National Bank of Poland, Working Paper No. 226.

Pokropek, A. and Jakubowski, M. (2013). Piaactools: Stata® programs for statistical computing using PIAAC data.

Psacharopoulos, G. and Patrinos, H. A. (2004). Returns to investment in education: a further update. *Education economics*, 12(2):111–134.

Tinbergen, J. (1974). Substitution of graduate by other labour. *Kyklos*, 27(2):217–226.

Yamaguchi, S. (2016). Changes in returns to task-specific skills and gender wage gap. *Journal of Human Resources*, pages 1214–6813R2.