Realizing full potential: Creating wage profiles for Human Capital calculations

Master Thesis, Policy Economics

Name:	Sofia Jileaeva
Student number:	571616
Supervisor:	prof. dr. (Dinand) HD Webbink
Second Assessor:	

ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Date final version: August 23rd 2024

Acknowledgements

I am grateful to my supervisor, prof. dr. Dinand H.D. Webbink from the Erasmus School of Economics, for his valuable feedback that pushed this thesis further than I thought it could go. Many thanks to my colleagues from the Centraal Planbureau—Paul Verstraten, Maria Zumbuehl, Frank van Erp, and Lily Davies—for their invaluable support and guidance, which were crucial to the development of this work. I also appreciate my friends for their encouragement and motivation. Finally, I am especially thankful to my parents for always believing in me and supporting my academic endeavors.

Abstract

Under the Lifetime Income Approach (LIA), the Human Capital (HC) Stock of a country is measured as the sum of the lifetime incomes of the population, which are equal to the present value of the earnings they will see throughout their lifetime. While HC is defined as the collection of all the knowledge and skills in a population, most national studies measure it using the labor market earnings of the working population. This means that the HC of the non-working population is not included, leading to an underestimation. I propose a new method of calculating HC stock, measuring productivity as the FTE of labor costs and including its 'potential' values for the non-working population. The latter is equal to predicted earnings using the Mincer earnings equation based on characteristics like the highest level of education, potential working experience, gender, area of residency, and migration background. My findings suggest that the conventional ('realized') measure underestimates the value of human capital stock by a range of 33 to 41 percentage points. Additionally, estimates from the Mincer equations indicate a decline in returns to education between 2015 and 2022. Despite this decrease, returns remain larger compared to the early 2000s, with annual income returns of 19% for women and 16% for men per additional year of schooling.

Contents

1 Introduction

Human capital is the collection of skills, knowledge, and experiences of the population that makes them productive members of society (OECD, 2016). It is an asset of any economy and a recognized input for economic growth (Aghion & Howitt, 1998; Lucas, 1988; Romer, 1989; World Bank, 2006). Understanding human capital is of high interest to policymakers since it could explain the drivers of growth, and performance of the labor market, and evaluate the long-term development path of an economy (UNECE, 2016). However, unlike physical and financial assets, which are easily defined, human capital is intangible and abstract, complicating its measurement.

The literature distinguishes different types of human capital: 'available' being the human capital embodied in each individual which can be potentially used in economic activities, and 'realized' which is the human capital that is actively used on the labor market. The difference between the two appears when there are individuals whose human capital is not priced on the labor market, like the unemployed. The estimate of the 'realized' human capital will then largely depend on the current economic conditions and will fluctuate with the economic cycle. Since the 'available' human capital more closely relates to the broader definition of human capital it is more useful to estimate the potential available in the economy, rather than the portion of the potential that is currently being employed. In a working paper, as part of the OECD human capital project, Liu (2011) acknowledges that due to conceptual, methodological, and data limitations commonly used methodology is limited to 'realized' human capital. In my thesis, I propose a novel method to incorporate individual potential taking account of personal working decisions and abilities.

Before addressing the technical considerations needed to estimate the economy's potential, it is important to understand the different methods of human capital stock estimation. In the Guide on Measuring Human Capital, UNECE (2016) highlights approaches based on cost, lifetime income, and indicators as different ways to measure human capital stock in an economy. The first measures the value of human capital from the education expenditures relying on the assumption that the value of outputs (human capital) is equal to the value of inputs (education). In the lifetime-income approach, income is used as a measure of productivity where the central assumption is that labor is paid according to its marginal productivity. This allows for consideration of the interplay of both labor demand and supply whereas the cost-based approach looks at the supply side only. The indicators-based approach results in a dashboard of indicators useful for policy making, though it limits compatibility across time and indicator types. In contrast, both cost-based and lifetime-income approaches arrive at a monetary estimate, which can be compared among others to different types of capital, (UNECE, 2016).

When there is a preference for a monetary estimate, the UNECE (2016) guide does not recommend one approach over the other choosing between cost-based and lifetime-income approaches since both have their advantages and disadvantages. However, several national studies and the OECD human capital project identify the lifetime-income approach as the one with the most practical implementation[1].

The Lifetime Income Approach (LIA) originally developed by Jorgenson and Fraumeni (1989, 1992), measures the value of human capital stock of a population as the discounted present value of all future income the individuals are expected to earn throughout their lifetime. Individuals are grouped by age j and education level i. Hence the total capital stock of an economy can be expressed as:

$$HC_t = \sum_i \sum_j W_{i,j,t}, \tag{1}$$

The human capital of the group of age j and education level i in year t is equal to the aggregate income of all the individuals who belong to the group:

$$W_{i,j,t} = \sum^{N_{i,j,t}} w_{i,j,t}, \tag{2}$$

where $w_{i,j,t}$ is the human capital of the representative individual of the group i, j in year t.

Under LIA, the human capital of the representative individual of the group of age j and education level i is equal to the real present value of the expected lifetime income which is the sum of their current income and the discounted value of the expected future income:

$$w_{i,j,t} = y_{i,j,t} + \theta \frac{E[w_{i,j+1,t+1}]}{1+r}, \tag{3}$$

where $y_{i,j,t}$ is the income in the current period, θ is the survival probability for the next period, $E[w_{i,j,t+1}]$ is the expected future income, and r is the discount rate. The future income flows the individual is expected to earn are equal to the present value of their human capital in the next period. Since in each period an individual has a choice to increase their

[1]Estimates of human capital based on the lifetime-income methodology are available for United States (Christian, 2010; Jorgenson & Fraumeni, 1989, 1992), Canada (Gu & Wong, 2010), UK (Jones & Chiripanhura, 2010; O'Mahony & Samek, 2021; O'Mahony & Stevens, 2004), Australia (Wei, 2008a), New Zealand (Le et al., 2006), Norway (Liu & Greaker, 2009), China (Li et al., 2009), and the Netherlands (Rensman, 2013) among others.

education level, the expected income of the next period is given by:

$$E[w_{i,j,t+1}] =(1-\beta_{i,j,t}) \times w_{i,j,t+1} + \alpha_{i,i+1,j}\beta_{i,j,t} \times w_{i+1,j,t+1} + (1-\alpha_{i,i+1,j})\beta_{i,j,t} \times w_{i,j,t+1}, \quad (4)$$

where α is the probability of switching into a higher education category and $\beta_{i,j,t}$ is the share of individuals in the group of age j and education level i currently enrolled in further education.

It follows that the most important input for the calculation of the human capital of the population is the income profiles for each of the age and education groups. As mentioned above, these income profiles are commonly based on the income currently realized in the economy ignoring the human capital which is not priced within the labor market. In this thesis, I propose ways to deal with individuals who are not (fully) active in the labor market to include their potential earnings in the income profiles.

Additionally to the observed earnings, the proposed method of estimating human capital assigns measures of productivity to the non-workers whose productivity is not priced on the market. I predict the earning potential of these individuals using a Mincer earnings equation based on their characteristics which are valued on the labor market. The predictions are assumed to be the earnings of the individuals if they had been working putting a price on their productivity. To get an indication of the potential range I consider two extreme scenarios: (1) The potential earnings of the non-working individuals are fully comparable to the realized earnings of the workers, meaning they are equal to the predicted earnings at face value. (2) A significant productivity gap exists between employees and non-workers due to selection so that the potential earnings are twice as small as the predicted earnings. Hence, I calculate two measures of the 'potential' HC where one assumes full comparability of working and non-working individuals and the other relaxes this assumption, indicating the upper and lower bounds respectively.

To better grasp the difference considering a larger population in the HC calculations makes for the resulting stock estimates, I compare the 'potential' and 'realized' measures. This comparison indicates the extent to which failing to include the broader measure of productivity in an economy leads to an underestimation of the HC stock estimate. Since the absolute numbers are highly sensitive to the discount and growth rates used, it is of no value to report these; instead, relative numbers are reported and discussed. It is also useful to examine where the gap between the 'realized' and 'available' values arises. Therefore, I compare the intermediate measures (current income and human capital of a representative individual) as well as the final estimate of HC stock between age and education groups to investigate the heterogeneity in gaps.

Given the growing interest in the topic of human capital, the findings of this research are particularly relevant. While most national studies emphasize the 'realized' measure of human capital, this thesis introduces a new perspective by calculating an alternative 'potential' measure. I compare the two approaches—specifically, the estimates based solely on observed earnings versus those that also include imputed earnings for non-workers, among other adjustments—to emphasize the importance of accounting for the skills embodied in the non-working population. My findings indicate that excluding non-working individuals from the estimation results in a value of the human capital stock lower by 41% percentage points. This supports the idea that the conventional method of calculating HC stock, which relies on market earnings, underestimates the true value of the stock.

In addition, I make use of the Mincer earnings equation (Mincer, 1974) in a different way: Since the estimates from the equation are informative on returns to schooling, I can study the results and compare them to the other estimates of returns to education in the Netherlands from previous years (Jacobs & Webbink, 2006; Webbink et al., 2013). Investment in education measured as the percentage of GDP spent on education has been growing worldwide (Patrinos, 2016). The biggest cost in education on the individual level is the foregone earnings while studying, hence knowing the returns on this investment helps make an informed decision on whether to continue education and what kind of education to follow. Returns to schooling estimates are not only useful for individual decision-making when choosing whether to study, but also for policy-making, when dealing with policies related to investment in education. I examine the returns to schooling in terms of hourly wages, hours worked, and yearly income for 2015 - 2022. I find that the returns have almost doubled for all education levels since 2009 indicating that the earnings gap between low and highly educated has grown.

The rest of the thesis is structured as follows: Section 2 gives a view on the current state of the literature in the field. The description of the data used in Section 3 is followed by an explanation of the methods in Section 4. In Section 5 I present my findings. Section 6 discusses the limitations and the findings of this research and concludes.

2 Literature Review

Using the lifetime income approach (LIA) to measure human capital (HC) stock dates back as early as 1690 (Petty, 1690). However, it is Wickens (1924) who is recognized to be the pioneer in using the capitalization of earnings method, which is the closest to the ones used today. I will evaluate the application of the approach, its advantages, and criticism, focusing on key national studies. Additionally, since the returns to education are estimated within my thesis, I discuss the related literature as well. This chapter aims to contextualize my research within the broader topic, highlight gaps, and propose new directions.

2.1 Lifetime Income Approach (LIA)

Under the lifetime income approach, the human capital stock is the aggregate human capital of all population groups, where the population is most commonly cross-classified by age, sex, and education level. The per capita human capital of the group is based on the lifetime income of the group's representative individual, which is equal to the present value of all the future income flows the individual will see throughout their lifetime.

There are many advantages to this approach which make it attractive to use for human capital stock estimation. For one, this method values human capital in market prices as seen from wages individuals receive in the market, allowing us to estimate the earning power of individuals. Since human capital is estimated using expected lifetime earnings, the approach is forward-looking compared to its counterparts based on historical costs. An economy that is interested in sustainable growth would be more interested in evaluating its future productive capacity (Graham & Webb, 1979), thus making LIA preferable. More importantly, relying on wages determined in the market, it manages to capture many factors including ability, productivity, professional qualifications, as well as the characteristics of the economy, looking at the interplay between human capital demand and supply (Dagum & Slottje, 2000). When the data needed are available, LIA is believed to produce the most reliable results (Le et al., 2003).

Le et al. (2003) also discuss the shortcomings of this approach. Most crucially, to estimate human capital under LIA one has to make a hard assumption that market wages reflect the marginal productivity of an individual. In reality, differences in wages within the population may come from other factors not related to productivity, like monopsony power in labor markets, in which case wages are lower (Webber, 2015) and productivity would be underestimated, or a premium wage set by trade unions which would overvalue productivity. Counter-cyclical unemployment and institutional characteristics that drive the gap in earnings between individuals of comparable characteristics would affect the measures of realized human capital stock. Liu (2011) concludes that as a result the estimates should be interpreted with caution. Assuming that wages are determined by marginal productivity will lead to a difference in human capital stock between men and women that can be misinterpreted as a productivity differential but is actually largely dependent on the institutional characteristics, like the fact that women are more likely to work part-time (UNECE, 2016). Moreover, the absolute estimate of human capital stock is sensitive to the choice of the discount and growth rates which are used to value the present value of future income streams (Le et al., 2003).

Despite these disadvantages, "The Measurement of Human Capital" workshop organized jointly by the OECD and Fondazione Giovanni Agnelli in Turin, 2008, which gathered leading researchers from the field identified the lifetime income approach as the best practical way to measure human capital stock. The current literature that uses this approach usually departs from the baseline model developed by Jorgenson and Fraumeni (1989, 1992).

2.2 J-F Model

The original J-F model improved on the lifetime income model of Wickens (1924) by simplifying the stream of future earnings to the present value and includes both market activities (involving employment) and non-market activities (like household production and leisure time) in the calculations. While the first type of activities is easy to measure using the annual estimates of hours worked and labor compensation, the value of non-market activities is a complicated issue. By deducting working hours, schooling (assumed to be 1300 hours a year if enrolled), and maintenance (10 hours a day) from the total number of hours available to an individual yearly, they arrive at an estimate of the yearly hours spent in non-market activities. Then, under the assumption that an optimizing individual chooses to work up to the point where the marginal benefit of work equals the marginal benefit of leisure, Jorgenson and Fraumeni (1989) value this time using the marginal after-tax market wage rate reflecting the opportunity cost of leisure. Since in each period individuals have a choice of increasing their education level and hence also their future earnings, the expected lifetime income in the next period is a combination of the lifetime income of an individual of a higher education level and of an individual of the same education level, one year older, weighted by the probability of enrollment into higher education level. Then, the total annual earnings for both market and non-market activities are valued using the marginal after-tax wage rate (Jorgenson & Fraumeni, 1989, 1992).

In his attempt to measure the human capital stock in the US between 1994 and 2006, Christian (2010) closely follows the original J-F methods and defines total human capital stock as lifetime labor incomes for both market and non-market activities of the entire population. He derives all the components needed for his calculations from the Current Population

Survey. Market activities are valued using average earnings and yearly working hours by sex, age, and education level groups. The value of non-market activities is estimated using the after-tax market wage rate and the average number of yearly non-market hours, which are equal to the average yearly available hours after deducting maintenance, school, and working hours. The post-tax wage rate is measured for each population subgroup using the data on employed individuals from the group. This implies that the human capital of the unemployed is measured using the wage rate of comparable employed individuals, introducing a bias in the calculations due to the selection into work. For individuals of age below 15, earnings and hence also the value of the non-market time are set to zero, so all the human capital for these age groups comes from the expectation of the future income inflow. Assuming a discount rate of 4% and an income growth rate of 2%, he arrives at very large estimates, $738 trillion for the human capital stock in 2006, of which 70% ($526 trillion) comes from the non-market activities. To compare, the value of physical capital in the same year was estimated at $45 trillion, which is approximately 16 times smaller. He observes a growth in the real HC stock with an annual rate of 1.1% between 1994 and 2006 and attributes it to the growth in the US population size rather than to the change in demographics of the population, but this growth can also be attributed to the growth in real income.

In his 2014 and 2017 studies, Christian continued the work he initiated in 2010. Due to changes in the data sets, in the paper of 2014, he re-evaluates his calculations for the previous years and arrives at an estimate of $672.8 trillion for the human capital stock in 2006 compared to the $738 trillion in Christian (2010). Here he also highlights the main differences between his methods and the original accounts in Jorgenson and Fraumeni (1989, 1992): the oldest age is set higher in his estimation and individuals from the oldest group are allowed to earn income; he values market work at pretax wage rate rather than post-tax wage rate. Christian (2017) advocates for using pretax wage or even full pretax compensation which adds employer contributions to the pension fund, benefits, etc. to the wage, to reflect the marginal product of labor rather than the marginal return to the worker.

2.3 Limiting to Non-Market Activities and Working-Age Population

The imputation of the value of non-market activities in the original J-F model was met with controversy and did not get widely adopted in the succeeding literature with the exception of a few studies (Ahlroth et al., 1997; Christian, 2010, 2014). Criticizing the model for the valuation of non-market production Rothschild (1992) uses an example of the audience in a football game or an opera. He shows the absurdity of valuing leisure time by the marginal after-tax wage rate since the value of experience within the audience would vary with the wage rates in that case. Aulin-Ahmavaara (2002) argues that not all the non-working time should be included in the valuation since some leisure time is required to prepare for work and some is not productive.

Christian (2014) proposes a new way of valuing non-market time by using The American Time Use Survey to restrict activities that should be valued in the non-market time. He identifies categories that are part of non-market work, like volunteering and childcare, and leisure, like watching TV, since these arguably should not be classified as non-market production. When estimating the value of non-market activities using non-market work and care responsibilities only, the estimates decrease to one-third of the baseline, which would also greatly affect the estimate of the total human capital stock since the baseline value of the non-market activities amounts to 70% of the total stock.

The issue of valuing non-market activities is contentious and is far from reaching consensus which is why most of the national studies focus on the value of market activities. UNECE (2016) developed the Guide on Measuring Human Capital, aimed at presenting the best practices in constructing human capital satellite accounts using common methods. Here they restrict the definition of human capital stock to its economic benefits excluding non-market activities from the framework. The paper by Liu (2011) written within the OECD human capital project limits its scope to market work due to "a number of conceptual, methodological, and data limitations". The estimates of the human capital stocks for a number of countries presented there relate to the realized human capital rather than the broader definition of human capital stock (Liu, 2011).

Additionally, research differentiates between working-age population human capital, and marketed human capital. Liu (2018) separates the employed population from the total population of working age and calculates 'active' human capital stock next to the 'total', to match the measures with the national accounts of Norway as much as possible. In the calculations of Australian human capital Wei (2008) explains his choice of confining to the working-age population: he argues that the portion of human capital embodied in the population of working age is the most important for economic activities and has a priority in the measurement of human capital. This is similar to the case for Canada: Gu and Wong (2010) focus on the working-age population stating its high relevance for evaluating the productive capacity of the Canadian population.

2.4 Other National Studies

Several national studies used LIA to estimate the value of the human capital stock by adopting the original J-F methodology to their needs. Measuring human capital of Norway, Liu and

Greaker (2009) altered the way students are treated within the common model used to calculate the human capital stock. In their approach, the calculation of the lifetime labor income of a student depends on the number of years left to complete their current study. Additionally, instead of the current income of a comparable individual of the same education level and age, income during the time of their studies is equal to the (part-time) earnings of the students in the same study phase. By distinguishing between 'students' and 'workers' this approach recognizes that the current earning potential of the students is limited by their studies and does not overestimate them with the earnings of the non-studying individuals.

Rensman (2013) calculates the stock of human capital for the Netherlands between 1999 and 2009 using microdata gathered by Statistics Netherlands. Here the estimate is limited to the realized value of the working age (15 - 64 years old) population reflected in the market wages of the individuals and is claimed to be the lower limit of the total human capital stock. To account for decisions to study for a higher education level and earn higher income the researcher considers two models. In the baseline calculations lifetime income of an individual in the next period is equal to the expected values of incomes resulting from the probability of studying for a higher education level (proxied by the enrollment rate). In the alternative model, she follows the approach developed by the above-mentioned Liu and Greaker (2009). Considering students as a separate group results in lower nominal values of human capital but higher growth than in the baseline model results.

2.5 Attempts to Measure Human Capital Potential

Similarly to the studies described above (Liu, 2011; UNECE, 2016; Wei, 2008b), Rensman (2013) reaches the conclusion that the human capital embodied in females was lower than that of males in all the years (1999-2009) considered. She recognizes however that this is likely due to females often working part-time leading to lower annual earnings and tries to correct it by adjusting the earnings of the part-time workers to a full-time equivalent. This is the first step towards the estimation of the human capital potential which is not currently utilized in the market. The model for the human capital potential includes part-time factors - the ratio of the number of hours worked by full-time employees to the average number of hours worked in the group. These part-time fractions are rough estimates, and the researcher recognizes a need to refine them to lower aggregation levels. When including the full-time equivalent earnings of the part-timers, the estimates of the human capital stock increase by 9.0-9.5%, while the growth does not differ much. This estimate shows the economic potential if all employees worked full-time, which would not be as easy to achieve with government policies, since individuals may have personal reasons to opt for part-time work (Rensman, 2013).

By focusing on morbidity instead of mortality, O'Mahony and Samek (2021) adjust the J-F model to account for health effects on human capital through absenteeism (quantity effect) and presenteeism (quality effect). To measure these effects they construct two human capital profiles: productive and potential. Potential human capital stock aggregates the lifetime incomes over the whole population, while the productive measure multiplies these by the employment rates for each population group. To construct the potential wages of the individuals that are out of the labor market, the researchers impute the employment earnings of similar demographic groups, controlling for non-random selection into work with the Heckman (1976) two-stage procedure. They find that in 2018 the estimate of the 'realized' human capital stock of the UK would have been 54% higher if all persons of relevant ages were working and the earnings of those in poor health were equal to the earnings of the healthy individuals. Therefore it is evident that when focusing on the human capital stock put productively into use on the market, the researchers are missing the bigger picture which can be used to inform policy.

2.6 Returns to Education

Since I estimate Mincer earning equations as part of my thesis, (Mincer, 1974) it is useful to compare my estimates of the returns to schooling to other findings from the literature on returns to schooling in the Netherlands. The earnings equations provide monetary estimations of returns to schooling which are easily comparable across studies. These can be used by individuals when making a decision on studying as well as policymakers to shape policies with respect to investment in education, like the design of the student loan system. Estimating Mincer equations for different groups, for example, men and women, provides insight into the extent of labor market discrimination (Patrinos, 2016). The studies that use this method to estimate the returns to schooling (Jacobs & Webbink, 2006; Leuven & Oosterbeek, 2000; Webbink et al., 2013) note the fact that it is not used to estimate the causal effect of education on earnings but rather provides an insight into the relationship between earnings, schooling, and experience. Studying the returns to education in the Netherlands Jacobs and Webbink (2006) and Leuven and Oosterbeek (2000) find that these have been increasing in the last years of the 20$^{\text{th}}$ century. More recently, Webbink et al. (2013) found that the monetary returns of an extra year of schooling have increased even more in the period between 1999 and 2009 from 5% (10.5%) to 9% (11.5%) for men (women). Since in the same period the supply of higher-educated individuals grew, such a robust increase in returns to education indicates that the demand for skilled labor increased even more rapidly. In relative terms, the demand for higher-educated individuals compared to low-educated has been increasing more than the relative supply for these types of labor. This also leads to an increase in earnings inequality between the groups (Webbink et al., 2013).

2.7 Research Gap and New Direction

In the case in which the goal is to measure the human capital stock matching the broader definition of human capital as the collection of knowledge and skills in the population, confiding the estimates to working-age employees presents a limited view on the concept. Since the estimates of the 'realized' human capital stock largely depend on the current economic conditions and vary with employment, an estimate of human capital which is independent of these is an attractive measure of the productivity of an economy. Identifying the sources of the gap between the 'potential' and the 'realized' human capital stock is useful to understand the labor market effects (O'Mahony & Samek, 2021), although few studies attempted to include imputations for potential productivity in their measures of human capital stock. While the attempt of Rensman (2013) is limited to accounting for the human capital potential of part-time employees, I recognize the need to estimate the available human capital of the whole population which I will explore in this thesis.

As for the topic of returns to education, the most recent study to my knowledge estimates the returns to education for the Netherlands up to 2009 (Webbink et al., 2013). My study not only provides an updated look at this measure and its development over time but also improves upon the data quality. Unlike previous research that relied on survey data, my analysis benefits from a larger and more reliable administrative dataset. This approach improves the estimation by relying on a larger sample size and minimizing the concern of the errors associated with self-reporting which can be the case in surveys.

3 Data

To construct income profiles I use multiple linkable datasets on the population of the Netherlands from Statistics Netherlands (CBS). Data collected by CBS is used for statistical research on behalf of the Dutch government and can likewise be used by researchers for their own studies. I define my full sample as working-age (16-65 years old) and registered in the Netherlands on the 1st of January of 2019 (11,228,327 observations). I use the data from 2019 since it is the last observed year in the tax information that was not affected by the COVID-19 pandemic.

I derive information on age (determined by the year of birth), gender, and level of education (the highest level of education achieved). It is important to note that it is only recently that the level of education became consistently recorded on the individual level in the Netherlands. This leads to it being more likely that older people have an unknown education level. To avoid removing them from my sample, I create a separate group for those for whom the information on the highest level of education is not available ('Unknown'). Furthermore, the others are grouped into three education categories based on the CBS classification: low, corresponds to primary education, VMBO, and MBO 1 (levels that do not give a start qualification); medium, corresponds to MBO 2/3/4, HAVO, and VWO (secondary education); high, corresponds to HBO, WO, and doctorate (higher education).

The information on the highest level of education comes from different sources: registers, population survey (EBB), and the unemployment organization (UWV). Due to the same issue as mentioned above where the level of education started to be consistently recorded in the registers only recently, the data on education for older people comes mainly from either EBB or UWV if not missing. While the population survey is representative, the information from UWV is likely to suffer from selection. I expect their earnings to be lower than those of a comparable individual of the same age and education level whose information is coming from a register or EBB. Using their earnings in the calculations might affect the estimates; this idea will be further explored in Section 6.

Individuals in formal employment make up the sample of employees[2], which is a portion of my overall sample. I retrieve the information on their taxable salary payments, benefits[3] received, and the number of hours. Individuals with an average hourly wage (determined as yearly income divided by the total hours worked in the year) which is less than 70% of the

[2]Note that this only includes formally employed and does not include self-employed (they are not observed in the dataset). Here and further I refer to them as 'employees'.

[3]Employer's contribution to pension and early retirement, holiday allowance, end-of-year bonuses, performance rewards, bonuses, profit distributions, and employer's part of the employee's insurance (in Dutch: Werkhervattingskas (Whk).

2019 Dutch age-dependent minimum wages are deleted from the sample leaving 8,148,253 unique observations.

Since the earnings need to reflect an individual's productivity, I incorporate benefits received on the job rather than looking at salary only and construct the variable Yearly Compensation. Jobs hiring individuals of similar productivity may offer compensation in different terms: primary (wage) or secondary (benefits) working conditions. More importantly, marginal productivity is assumed to be equal to the monetary compensation on the labor market, which is best reflected in the cost of the worker from the employer's side. This approach is also advocated by Christian (2017) since full compensation better captures a worker's productivity while the earnings only reflect the return to the worker. Then, to arrive at the full-time equivalent (FTE) measure of compensation I calculate the average hourly compensation and multiply it by 1720 hours, which are defined as the yearly work hours for full-time employment by the Dutch Ministry of Social Affairs and Employment.

I identify the students using the information on enrollment in education on the 1st of January 2019. For the place of residency, I create a dummy variable Randstad equal to one if an individual lives in one of the four big cities of the region known as G4[4] and zero otherwise. I record the migration background of individuals in the dummy variable Native, which equals one if the individual is a native Dutch and zero otherwise. In case the person and both of their parents are born in the Netherlands, they are considered a Dutch-native.

Therefore, I arrive at two samples of overall working-age individuals and of employees. Table 3.1 presents the descriptive statistics of the overall sample, while Table 3.2 - for the employees, both additionally split by education category. Columns (2) – (4), Table 3.2, show that mean Yearly Compensation and Yearly Hours increase with the level of education for the employees. Meanwhile, the earnings of those with an unknown education level are in between the means of Medium and High education categories, while working the most hours on average. They are also on average older than employees from the other groups, reflecting the fact that the education information is missing for older individuals mostly. Means of FTE Compensation are consistently larger for employees than for the overall sample indicating that they are on average more productive than the rest. Comparing the totals of the two groups, the employees are on average younger, are less likely to live in the Randstad area, and are more likely to be Dutch-native than the persons from the overall sample, while the share of students is similar. Females from the Low and Unknown education categories are underrepresented in employment but the difference with the overall sample is not large. Overall, the sample of employees makes up a big part of the overall sample (around 72%).

[4]Amsterdam, Rotterdam, The Hague, and Utrecht.

Table 3.1: Descriptive Statistics of the Overall Sample

	Low	Medium	High	Unknown	Total
FTE Compensation	21265.50	32694.97	62429.57	53771.15	43567.73
	(28096.38)	(32324.84)	(122582.77)	(348118.81)	(179004.08)
Age	35.91	36.64	40.69	50.75	41.11
	(16.78)	(13.90)	(11.60)	(10.40)	(14.40)
Female	0.49	0.49	0.52	0.49	0.50
Randstad	0.15	0.13	0.20	0.13	0.15
Native	0.65	0.76	0.80	0.72	0.74
Students	0.32	0.21	0.07	0.00	0.14
Observations	2,029,889	3,486,739	2,876,128	2,835,571	11,228,327

Note: The sample is split into groups based on the highest-achieved education level. For the variables FTE Compensation and Age, the mean and standard deviation (in parentheses) are shown. For the binary variables Female, Randstad, Native, and Students the fraction of the population for whom the value is equal to one is shown. FTE Compensation is measured in 2019 Euros. Age is measured in years.

Table 3.2: Descriptive Statistics of the Sample of Employees

	Low	Medium	High	Unknown	Total
Yearly Compensation	16631.43	28254.06	55306.77	48318.97	38936.95
	(18879.05)	(27581.15)	(54078.21)	(69145.67)	(49589.86)
Yearly Hours	924.54	1224.79	1420.92	1432.57	1283.62
	(640.47)	(579.93)	(477.23)	(502.25)	(572.80)
FTE Compensation	25004.20	36516.87	64214.55	56010.32	47273.12
	(28797.92)	(32100.31)	(123425.36)	(353719.13)	(183680.56)
Age	32.91	35.61	39.88	50.46	39.84
	(16.18)	(13.42)	(11.19)	(9.79)	(14.01)
Female	0.46	0.49	0.52	0.55	0.51
Randstad	0.13	0.11	0.19	0.12	0.14
Native	0.69	0.79	0.82	0.75	0.77
Students	0.40	0.21	0.07	0.00	0.14
Observations	1,240,424	2,675,560	2,367,524	1,864,745	8,148,253

Note: The sample is split into groups based on the highest-achieved education level. For the variables Yearly Compensation, Yearly Hours, FTE Compensation, and Age the mean and standard deviation (in parentheses) are shown. For the binary variables Female, Randstad, Native, and Students the fraction of the population for whom the value is equal to one is shown. Yearly Compensation and FTE Compensation are measured in 2019 Euros. Age is measured in years.

4 Methodology

4.1 Mincer Equations

4.1.1 Returns to Education

The first part of my analysis focuses on estimating returns to education using Mincer earnings equations (Mincer, 1974). These are used to explain earnings as the function of schooling and work experience. Typically, log wages are modeled as the function of education years and years of experience adding a quadratic term for experience. Since work experience is rarely known, its potential value is used instead, which is proxied by the time an individual had available to participate in the labor market. The estimated coefficient of the schooling variable indicates the financial returns to education and explains the extent to which extra education is valued by the market.

To easily compare the estimates of the returns in 2019 to the estimates found by Webbink et al. (2013) for 2009 and by Jacobs and Webbink (2006) for 2002 I closely follow the similar regression specifications (1 - returns to education levels, 2 - returns to an extra year of education) used in the papers. Webbink et al. (2013) use the yearly income as the dependent variable, while Jacobs and Webbink (2006) uses the hourly wages. Since I have data on both, I can estimate both hourly wage and annual income returns and compare them to the researchers' findings for earlier years. A regression using hourly wages estimates the returns of education on the wage level while using annual income provides insight into labor supply returns. In case both wages and hours worked increase with education, the estimated returns on annual income would be greater than on hourly wages. To investigate the association between education and labor supply I additionally estimate the hours worked returns to education.

Therefore, I adjust the two specifications used by the researchers by using three dependent variables. For this, I use data on the employees with available education data from the period of 2015 until 2022. Since I also observe whether an individual achieved a doctorate, I include it as one of the education levels dummy variables. The first measures the returns of different levels of education relative to the lowest education level, namely primary education ('Basisonderwijs' in Dutch). The other levels are grouped into five groups: VMBO for lower (for VMBO and MBO 1), MBO/HAVO/VWO for middle (MBO 2/3/4, HAVO, VWO), HBO for higher vocational education, WO for university degree, and DOC for doctorate:

$$\log(Y_i) = \alpha_0 + \sum_{k=1}^{5} \beta_k \text{Education Level}_{k,i} + \gamma_1 \text{Experience}_i + \gamma_2 \text{Experience}_i^2 + \gamma_3 \text{Woman}_i + \epsilon \tag{5}$$

where Y is the outcome variable: hourly wage, hours worked, or annual income. Each β will deliver the estimate of the returns to the relevant level of education relative to the primary education level. To translate the estimates into percent for easier interpretation I use the semi-elasticity formula for log-linear regressions[5]. Experience is measured in years an individual could potentially work which is equal to their age minus the time spent in education. The latter is defined by the necessary time needed to achieve the relevant education level by default. In this I follow Webbink et al. (2013): for primary education, it is six years, for VMBO it is 10 years, for MBO/HAVO/VWO - 12, for HBO - 15, and 17 years for WO. For DOC I equate it to 20 years. To account for diminishing returns to experience I include the quadratic term for Experience. Males and females experience differences in earnings even after accounting for education and years of experience, hence it is important to include a control variable for the sex.

The second specification uses years of education instead of education levels to estimate the earnings return to an extra year of education:

$$\log(Y_i) = \alpha_0 + \beta_1 \text{Education Years}_i + \gamma_1 \text{Experience}_i + \gamma_2 \text{Experience}_i^2 + \epsilon, \tag{6}$$

where Y is the outcome variable equal to hourly wages, hours worked, or total income. Here the estimate of interest is β_1. Converting it to percent using the before-mentioned semi-elasticity formula, shows the return of an extra year of education on the wage level, hours worked, or yearly income. Additionally, I estimate the regression separately for men and women as in Webbink et al. (2013). Since the work experience profiles likely differ for the two sexes due to time off the labor market related to pregnancy and childcare for women, this is accounted for when splitting the sample by sex. A difference in estimated coefficients would indicate a difference in returns to an extra year of schooling between men and women.

Both equations (5) and (6) are estimated on the sub-sample of employees whose education level is known and those who do not work more than 2400 hours a year (approximately 49 hours a week) (the descriptive statistics of this subsample are reported in Appendix A.2). The regressions are run separately for each year of the available data (2015 - 2022) and pooled across the whole period. The analysis over multiple years will allow me to examine the

[5] $\% = \left(e^{\beta} - 1\right) \times 100$.

changes in returns throughout time. The regression using all the years from the period studies provides average estimates within the period. As mentioned, the sample is additionally split by gender for equation (6) to account for the time profile differential.

4.1.2 Using Estimates for 'Potential' Earnings Prediction

To later calculate the 'potential' human capital, the Mincer equation's estimates found from the sample of employees are used to predict the earnings of the non-working individuals. Other than estimating the returns to schooling, Mincer earnings equations can capture hedonic price functions of the extent to which the labor market values certain individual characteristics (Heckman et al., 2003).

I opt for using education levels rather than years of education in my estimations. This decision should not affect the estimates since the two specifications are related through the definition of work experience. Since the calculation of human capital stock is based on full-time equivalent earnings this is the dependent variable used for the prediction. Additionally, I improve on the simple Mincerian equation (5) used in the previous section by adding more controls for worker characteristics to increase the precision of the predicted earnings. The important factors to control for in addition to gender include migration background and place of residence. Since the earnings of migrants and natives largely differ, including migration background will help to better predict the earnings of unemployed migrants. A recent report by the International Labor Organisation Amo-Agyei et al. (2020) shows a large migrant pay gap in the Netherlands of 19.9%. Even after accounting for labor market characteristics, the gap remains mostly unexplained, highlighting that factors other than education level and work experience are at play. The researcher also concludes that within the same occupation, migrant workers earn less despite similar levels of education. Not including migration background in the estimation could lead to an overestimation of the earnings for unemployed migrants since these would be partially based on the larger earnings of their native counterparts.

Place of residence is another important determinant of income. For the Netherlands, Groot et al. (2014) conclude that spacial wage differences exist and are large between regions even after correcting for worker heterogeneity using a Mincer equation. They observe a large wage gap between the Randstad area and other regions. This motivates me to include a region dummy which equals one when a person is living in the Randstad area in the beginning of 2019 and zero otherwise. Hence the adjusted specification including additional characteristics that explain earnings is:

$$\log(\text{FTE compensation}_i) = \alpha_0 + \sum_{i=1}^{5} \beta_i \text{Education Level}_i + \gamma_1 \text{Experience}_i + \gamma_2 \text{Experience}_i^2 + \gamma_3 \text{Woman}_i + \gamma_4 \text{Migration Background} + \gamma_5 \text{Randstad} + \epsilon \tag{7}$$

To estimate the hedonic prices I use the sample of employees with non-negative earnings and working hours no more than 2400 in 2019. These estimates are then used to predict the earnings potential of the non-working individuals (including the self-employed) if their earnings are not observed in 2018 or 2020 either. The result is assumed to be equal to their market earnings if they were to be formally employed.

For the group with an unknown education level, I estimate a separate hedonic price function on the sub-sample of employees from the same group and use their characteristics for prediction. Here, the work experience cannot be calculated so it is assumed to be 20 years. Since the average age of the group is equal to 50 years old, it is a fair approximation.

4.2 Human Capital Stock

4.2.1 'Realized' HC

The 'realized' HC stock measure is derived using the observed earnings of individuals. I start with calculating the HC of the representative individual for all age and education groups. As seen from (3) this is equal to the income in the current period and the present value of all the future income flows weighted by survival probability. The income in the current period is calculated as the average income in the group[6].

An important assumption needs to be made: The lifetime income of the representative individual of age j in the next period is equal to the lifetime income of the representative individual one year older (of age $j+1$) in the current period, adjusted for income growth. This allows me to calculate lifetime incomes for each age and education group starting with the oldest (here, 65 y.o.). Their lifetime income is assumed to be equal to their current earnings due to retirement and inactivity in the labor market in the next period ($E[w_{i,j,t+1}] = 0$). Then the lifetime income of a 64-year-old of education group i is equal to their current income and the present value of the lifetime income of a 65-year-old of the same education. This exercise continues until reaching the 16-year-old groups.

[6] To calculate the average income, I use the group sizes from the overall sample to isolate the effects of different productivity calculation methods when I compare the resulting measures. This ensures that variations are attributed to the difference between the methods rather than differences in group sizes.

To calculate the expected future human capital, the shift probabilities between education levels, as defined in (4), are needed. Since the main purpose of the total HC calculation is to retrieve the proportion of the 'realized' value relative to the 'potential', such a complication can be avoided with no harm to the relative estimates. Hence, for the sake of simplicity, I disregard the probability of switching to a higher education level for all individuals, so the education level is fixed at the currently observed one. Under this assumption the present value of expected future income flows for a representative individual is equal to the human capital of the individual one year older of the same education group multiplied by the real income growth rate, g:

$$E[w_{i,j+1,t+1}] = w_{i,j+1,t} \times (1+g). \tag{8}$$

I use the estimates of real growth and discount rates from Adema and van Tilburg (2019) which are 1% and 2.5% respectively. Survival probabilities are calculated based on CBS Statline data from 2023 (can be found in Appendix A.1). I aggregate the resulting measures of HC of a representative individual to arrive at the values of HC for all groups using equation (2). Next, the total stock can be calculated as the sum of all group HC values. It represents the value of the human capital based on wages derived from employment in the Dutch economy in 2019. The absolute estimates of HC stock largely depend on the choice of the discount and growth rates, and I argue that the value of the estimation lies in comparing the measure of 'realized' HC to the 'potential' in relative terms.

4.2.2 'Potential' HC

'Potential' HC stock is an alternative measure of HC. Compared to the 'realized' value it includes the productivity of the whole working-age population, rather than employees only. Since the earnings and hence the human capital of non-working individuals are not observed on the market, the challenge lies in imputing values that would represent their earnings if they were working on the labor market.

To incorporate potential earnings in my calculations of income profiles I identify groups that should be considered separately: part-time employees, students, persons with disability, self-employed, and other non-working. The potential individual income is equal to their earnings as if they were working full time unless time/ability-constrained. For part-time workers, this means that I convert their income to full-time by multiplying their hourly wage and the average yearly full-time working hours (1720 hours). The earnings of all the individuals working more than a full-time equivalent (FTE) of hours in a year, 1720 hours, are converted to the FTE income using their hourly wages to keep the earnings measure consistent. For students, I assume no income while they are enrolled in schooling since during that time their human capital is not available for productive use during the studying period.

The largest difference in creating wage profiles that are not sensitive to business cycle fluctuations comes in the way the non-working persons are treated. Unemployment increases in the periods of economic recession so if the wages of unemployed persons are not included in the calculation of the income profiles the results will be lower during those periods. This does not however mean that the human capital is lost during recession; it is still available in the economy. So to estimate the wage profiles based on the potential earnings of the group it is important to account for the potential earnings of those who are currently out of the labor market. Due to data limitations, where the earnings of self-employed are not available in the data, they are treated as part of the non-working group.

If in the current period (t) an individual does not have a reported income, it is first possible to check whether they had an income in the previous period ($t-1$) or the period following ($t+1$). I can then derive the income in the current period from one of these figures accounting for average earnings growth between the periods for 378,555 individuals. The result is the closest estimate of that person's income since historical wages are most likely equal to the current income earning capacity of the individual. If the wages in closest periods are not available either, predictions from the Mincer equation (7) based on individual characteristics like education level, age, gender, place of residence, and migration background are used.

Since instead of using observed earnings the predicted values are used instead, the value of assigned earnings is likely biased. Participation in the labor market is an individual choice likely indicating the differences between individuals who are employed and those out of the labor market, including skills and productivity. I use Mincer predictions to impute earnings for the latter group, however since the estimates in the equation are developed from the sample of employees this productivity difference will not appear in the the predicted earnings, leading to an overestimation of the 'potential' HC stock measure. To relax this assumption, I develop an alternative measure of imputed earnings. I assume a productivity gap of 50% between a non-worker and a comparable employee, hence imputed earnings are equal to half the earnings predicted by the non-worlers' characteristics. Since the true productivity difference between workers and non-workers is unknown, I will report both measures. Assuming full comparability between individuals in and out of the labor market likely offers an upper bound of the range, while the measure including a productivity gap of 50% - a lower bound.

Another limitation of the earnings imputations lies in the fact that I am not able to observe disability status of individuals. This issue has several implications: The earnings of partially disabled, which are most likely not comparable to the earnings of an employee with no disability, are included in the estimation of hedonic prices in (7) (only if they work no more than 2400 hours a year) and hence affect the predicted earnings of the non-working sub-sample. Additionally, persons with a disability out of the labor market get assigned a wage in the same manner as the rest of the non-working individuals while their disability status is not taken into account. Since disability reduces productive working capacity not considering it leads to an overestimation of the 'potential' earnings of those persons. This idea will be further discussed in Section 6.

4.2.3 Comparison of the 'Realized' and 'Potential' HC Measures

The interest in developing both 'realized' and 'potential' measures of human capital stock lies in comparing the two in relative terms rather than absolute since the absolute values highly depend on the choice of discount and growth rates and survival probabilities. Since to arrive at the 'potential' measure I make several adjustments, I break down the effect these have on the result. The first step shows the effect of converting the earnings of the employees to a full-time equivalent, and the second shows the additional effect of imputing earnings for the non-workers (or interpolating from historical wages if available) and setting the earnings of students to zero.

To analyze the gap between the 'realized' and the 'potential' measures, next to comparing the total stocks, I look at differences in the intermediate measures. Firstly, I compare the lifetime income of a representative individual (individual HC) for each education and age group. This will show how the value used in computing the 'realized' HC stock changes relative to the value used in the 'potential' HC stock calculations. Differences between education groups can appear due to the difference in employment rates. The values of the group HC are calculated based on the values of individual HC and result in the same shares. For both, the share of the 'realized' HC in the 'potential' is likely to be less than 100% since the latter measure accounts for the broader measure of productivity. I expect the shares to decrease with age as individuals exit the labor force while still getting assigned some earnings until the official retirement age under the 'potential' method.

Next, I compare the HC stock measures separately for each education category. Differences here could appear due to differences between employment rates in each education category. Overall, since the 'potential' measure considers the productivity of a larger sample of the population compared to the 'realized' which takes into account only formally employed, I expect the former to be much larger.

5 Results

5.1 Returns to Education

This sub-section presents the results from Mincer earnings equations (5) and (6) to estimate the returns to education in the Netherlands in 2015 - 2022, and compares them to the estimates found in Webbink et al. (2013) for annual income returns and Jacobs and Webbink (2006) for hourly wage returns.

Figure 5.1 shows the results of estimating equation (5) pooled over the years 2015-2022, reported in Table 5.1. There are notable differences in the coefficients between education levels indicating differences in the pay and hours worked between individuals of different education levels, conditional on characteristics like experience and gender. Both hourly wages and hours worked increase with education. Since yearly income is a function of the two, it also increases with education.

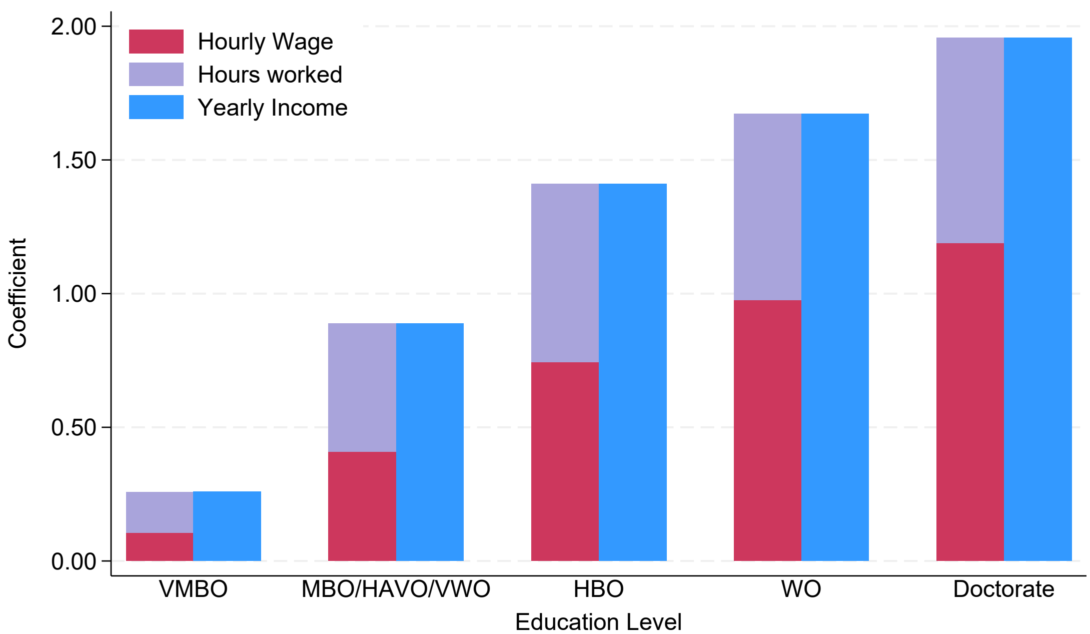


Figure 5.1: Returns per education level, 2015–2022

Note: The bars show the coefficients of the log-linear regressions with hourly wage, hours worked, and annual income as the dependent variables with levels of education as independent variables (primary schooling being the reference category).

I convert the coefficients from the log-linear regression into percentages using the semi-elasticity formula for easier interpretation and analyze the evolution of the returns over the

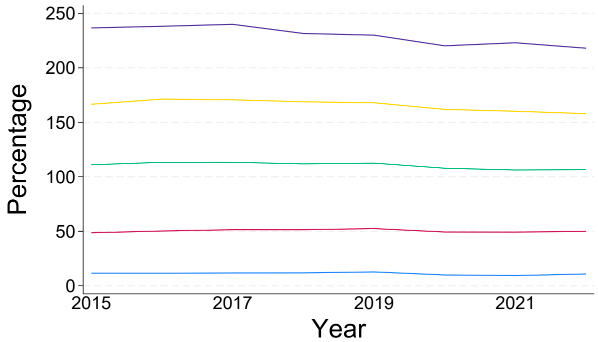
Table 5.1: Returns to Education (Levels)

Education Level	log(Hourly Wage)	log(Hours Worked)	log(Yearly Income)
VMBO	0.11	0.15	0.26
	(0.00)***	(0.00)***	(0.00)***
MBO/HAVO/VWO	0.41	0.48	0.89
	(0.00)***	(0.00)***	(0.00)***
HBO	0.74	0.67	1.41
	(0.00)***	(0.00)***	(0.00)***
WO	0.98	0,70	1.67
	(0.00)***	(0.00)***	(0.00)***
Doctorate	1.19	0.77	1.96
	(0.00)***	(0.00)***	(0.00)***
Controls	Yes	Yes	Yes
R-squared	0.66	0.23	0.47
Years	2015–2022	2015–2022	2015–2022
Observations	48,772,748	48,772,748	48,772,748

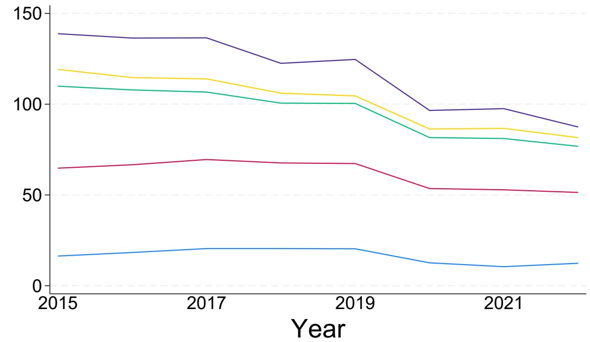
*Note: The analysis is performed on the sample of employees with a known education level pooled over the years 2015-2022. The set of controls includes Experience, Experience-squared, and a binary variable for sex. The robust standard errors are given in parentheses. * 10%, ** 5%, *** 1%.*

period studied, 2015–2022. The results are reported in Appendices A.3, A.4, and A.5 for hourly wages, hours worked, and annual income, respectively. Figure 5.2, Panel A, plots the relative wage returns for each education level. The profiles look mostly flat with small decreases for higher levels of education. A different pattern is observed in Panel B of the same figure: The relative returns on hours worked declined, with the decrease being more pronounced where the initial returns were higher. The combination of these is observed in Panel C. The relative returns on annual income have decreased, with higher levels of education experiencing the most significant declines, though not as steep as the reduction for hours worked. Overall, this indicates that the income gap between higher and lower education levels decreased in the years 2015-2022, which is largely attributed to the narrowed gap in hours worked. Nonetheless, the disparities remain high: The return on annual income for VMBO degree holders compared to those with only primary education is around 24%, while for Doctorate holders the figure is as large as approximately 496%.

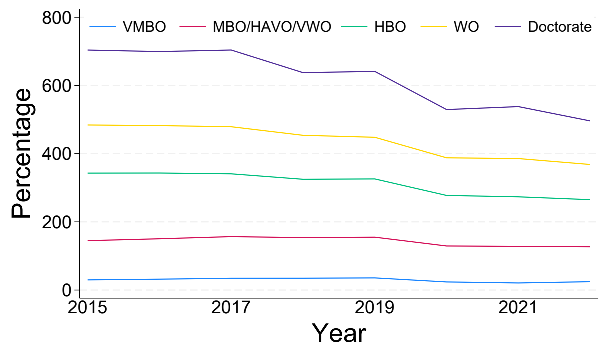
Contrary to the findings of Webbink et al. (2013) for the years 1999-2009 which show an increase in returns on annual income, I observe a decrease in the more recent years. Irrespective of this decline, the estimated returns for 2022 remain much larger than those in 2009 and the differences are more pronounced for higher education levels: 25% vs 11% for



Panel A: Hourly Wage



Panel B: Hours Worked



Panel C: Yearly Income

Figure 5.2: Returns per level of education

VMBO; 127% vs 40% for MBO/HAVO/VWO; 265% vs 75% for HBO; and 368% vs 96% for WO[7]. Even though the income gap between lower and higher education levels narrowed in the years 2015–2022, it remains larger than the gap observed in the early 2000s.

To compare the hourly wage returns, I refer to the study by Jacobs and Webbink (2006). While the researchers report an increase in relative returns on hourly wages in the last years of the period they study (1979–2002), my findings indicate that these were stable in the years 2015-2022, except for the higher education groups experiencing slight decreases. Once more the returns in my paper exceed the returns estimated by Jacobs and Webbink (2006) and are more than twice as large: 11% vs 5% for VMBO (in Jacobs and Webbink (2006): VBO and MAVO), 50% vs 23% for MBO/HAVO/VWO, 107% vs 48% for HBO, and 158% vs 70% for WO[8].

The large differences between the findings of the researchers and mine could be attributed to the differences in the data used, survey-based panels vs administrative data, or point

[7]The returns for Doctorate are not available in Webbink et al. (2013), hence could not be compared.

[8]I compare the returns for 2022 from my paper and 2002 from Jacobs and Webbink (2006). Returns for Doctorate level are not available in their paper so could not be compared.

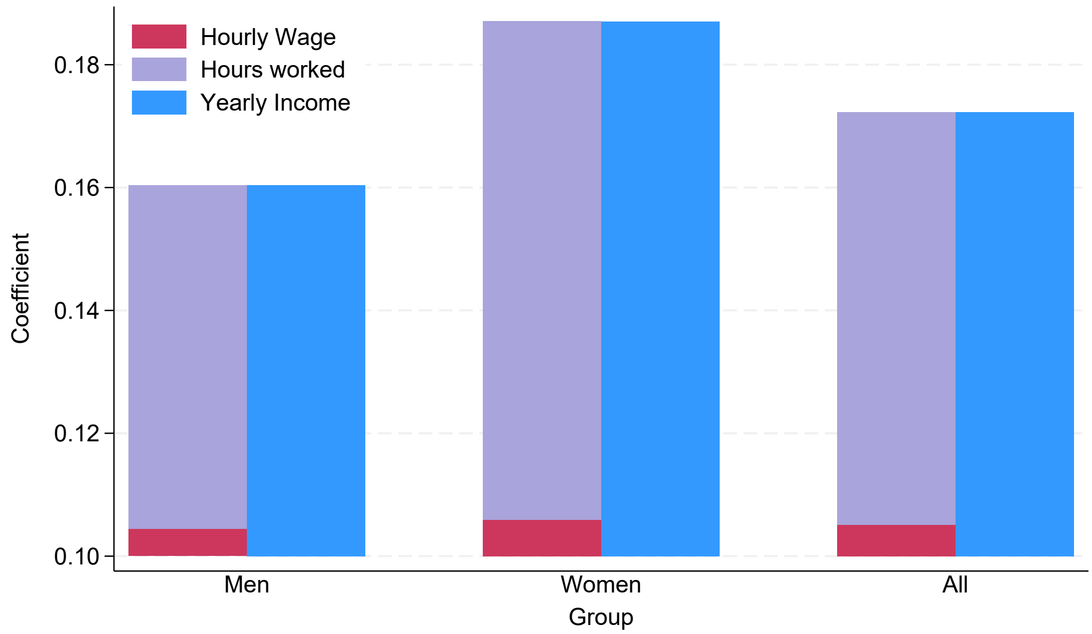


Figure 5.3: Returns to an extra year of education by gender, 2015–2022

towards a substantial increase in returns in the periods that were not analyzed (2010-2014 for annual income and 2003-2014 for hourly wages). An alternative explanation, for such large differences could be specific to the comparison of the education levels to the reference group of individuals with no more than primary schooling. The dropout rates before reaching the starting qualification (here equivalent to MBO/HAVO/VWO) are decreasing with years, meaning that the group of primary education becomes more selective with time. The large returns could be the result of such selection.

Hence, to generalize the returns, I investigate returns to an extra year of schooling estimated from equation (6). The coefficients from the regression with observation pooled over the years 2015-2022 are recorded in Table 5.2 additionally separated by sex; Figure 5.3 presents a visualization of these. Overall, the coefficients with hourly wage as the dependent variable are larger than when hours worked are used as the dependent variable. When splitting the sample by sex, the beta estimates for the log-linear regression of hourly wages differ for men (0.11) and women (0.10) as do the estimates from the regression of hours worked: 0.08 for women compared to 0.06 for men. This translates into higher annual income estimates for women.

I exponentially transform the log-linear coefficients for easier interpretation and analyze each year from the period 2015-2022 separately. The results are reported in Appendices A.6, A.7, and A.8 for hourly wages, hours worked, and annual income, respectively. Figure

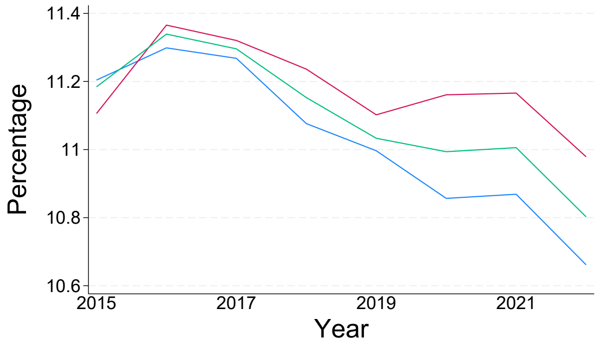
Table 5.2: Returns to Education (Years)

Dependent Variable	Women	Men	All
log(Hourly Wage)	0.11	0.10	0.11
	(0.00)***	(0.00)***	(0.00)***
	[0.65]	[0.65]	[0.64]
log(Hours Worked)	0.08	0.06	0.07
	(0.00)***	(0.00)***	(0.00)***
	[0.18]	[0.24]	[0.20]
log(Yearly Income)	0.19	0.16	0.17
	(0.00)***	(0.00)***	(0.00)***
	[0.42]	[0.49]	[0.44]
Controls	Yes	Yes	Yes
Years	2015–2022	2015–2022	2015–2022
Observations	24,037,438	24,735,312	48,772,748

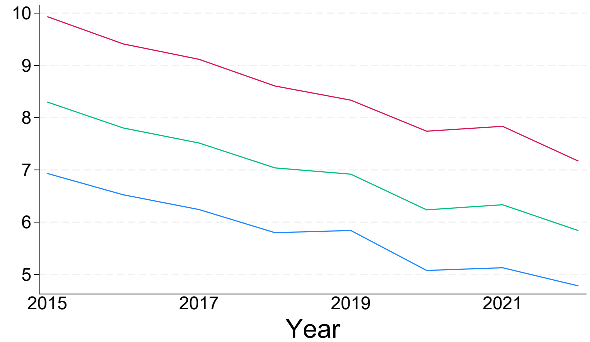
*Note: The sample of employees with a known education level was split by a binary sex indicator. The set of controls includes Experience, Experience-squared, and a binary variable for sex. The robust standard errors are given in parentheses; the R-squared are given in square brackets. * 10%, ** 5%, *** 1%.*

5.4, Panel A, shows that in the earlier years, the wage returns were initially higher for men than for women. The returns increased more significantly for women than for men resulting in higher returns for women. Although the returns declined for everyone after 2016, the gap between the sexes widened over time. Panel B of the same figure illustrates that the returns on hours worked were consistently higher for women between 2015 and 2022. This result suggests that an extra year of schooling is associated with a larger increase in hours worked for women than it does for men. Since hours worked increase with education for women, whereas, for men hours worked remain relatively constant regardless of educational attainment (Webbink et al., 2013), this outcome is substantiated. Nevertheless, returns on hours worked had been declining for both sexes indicating that the association of an extra year of schooling and hours worked was losing its significance in the studied period. The same pattern emerges for returns on annual income (Panel C): The returns are generally larger for women, although they had been on a decline for everyone.

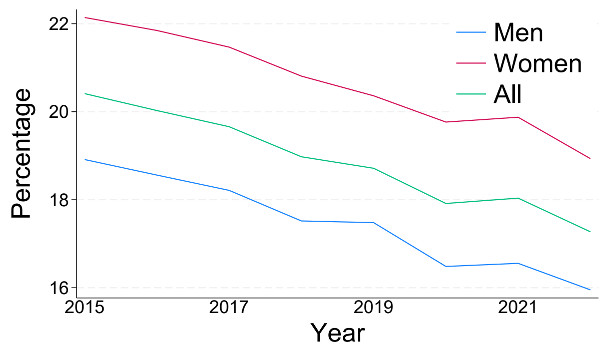
Despite this decline in recent years, the returns on annual income remain larger than those estimated in Webbink et al. (2013): 19% (16%) in 2022 versus 12% (8%) in 2009 for women (men). In terms of wage returns, my findings indicate higher returns for both men and women of approximately 11% in 2022, while Jacobs and Webbink (2006) report these at 7% in 2002.



Panel A: Hourly Wage



Panel B: Hours Worked



Panel C: Yearly Income

Figure 5.4: Returns to an extra year of education

Overall, these findings imply that there is a large association between financial returns and education level, coming from both higher hourly wages and annual hours worked the higher the education level. In the years 2015-2022, the returns on hours worked had decreased leading to a decrease in annual income returns as well. The differences in hourly wage returns remained stable. This implies that while the difference in hourly pay persisted between lower- and higher-educated individuals, the gap in working hours narrowed, leading to a narrowing annual income gap as well. An extra year of schooling is valued similarly for both men and women, however, hours worked are disproportionately larger for women the higher the degree. This trend was on a decline in the period between 2015 and 2022, but the gap persists. The differences with the returns estimated in Webbink et al. (2013) and Jacobs and Webbink (2006) imply that the association between the wage/annual income and education has increased substantially between the period studied in this thesis and the periods studied by them.

5.2 'Realized' vs 'Potential' HC Measures

In this sub-section, I present the results of comparing the measure of human capital stock using the conventional method from the literature ('realized') and the measures using my proposed method ('potential'). As laid out in Section 4, I compare several intermediate measures as well as the end estimate of HC stock.

First, Figure 5.5 plots income profiles averaged across all education categories for different approaches. The profiles are based on the current income of the representative individual of the age and education group. For the 'realized' value, this corresponds to the average compensation of the group in the relevant period. 'FTE' shows the income profiles when the compensation is converted to a full-time equivalent. For 'potential' two alternative measures are shown: One that takes into account a productivity gap of 50% between a non-worker and a comparable employee and another that assumes the productivity of individuals is equal conditional on their characteristics. The corresponding income profiles are reported in Appendix A.11. The estimates used for the prediction of unobserved productivity can be found in Appendices A.9 and A.10.

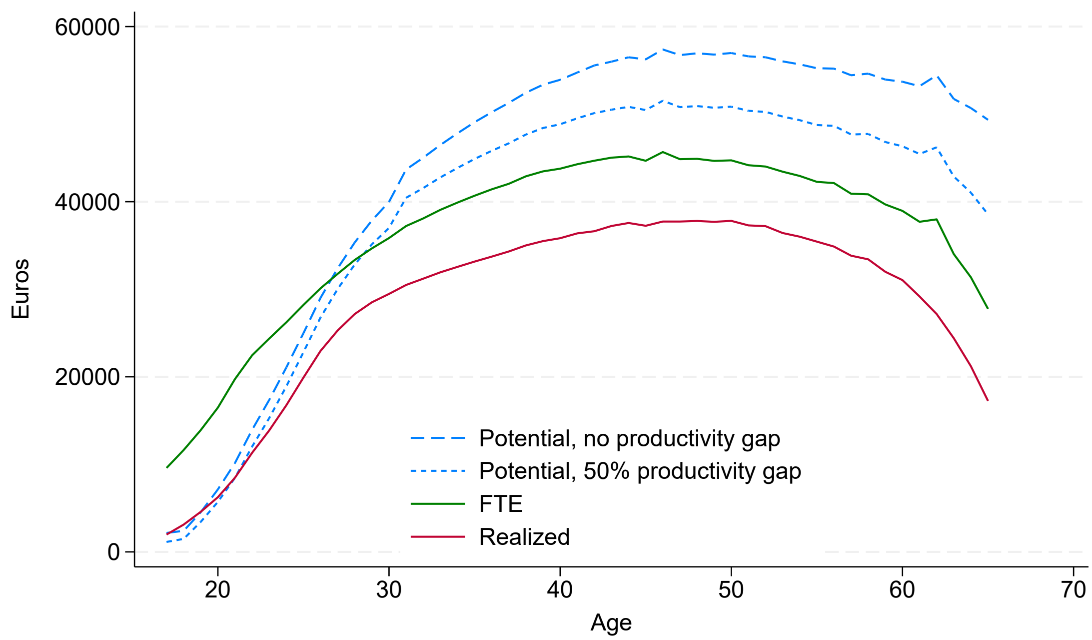


Figure 5.5: Income profiles per HC measure

As expected, for most of the ages the average income is larger in the 'potential' measures than in the 'realized' since the former measures both convert the earnings of the employees to an FTE and assign a non-zero income to individuals out of the labor market. When only including converted FTE earnings, the average incomes increase almost proportionally for all ages, significantly narrowing the gap between the 'realized' and 'potential' measures. The opposite is observed for the younger age groups where the FTE income profiles exceed the 'potential' ones. This is the result of assigning a zero income for students in the 'potential' measures while keeping the observed earnings and converting them to an equivalent of a full-time job in the 'FTE' one.

The difference between the two 'potential' measures shows how the assumption about the non-workers' productivity affects the measure of average income. The gap widens over time since the pool of non-workers increases with age and the assumption made about their earnings' imputations gains more significance for the total measure.

For all, the average annual income peaks at around 45 years old and then experiences a decline. The decline is sharper for the measures relying on observed income, which can be explained by older age individuals exiting the labor market before reaching the official retirement age while getting assigned an income under the 'potential' method.

Using the income profiles for each age and education category group I calculate human capital measures of representative individuals using equation (3). I report the findings in relative terms indicating the proportions of 'realized' HC compared to the two measures of 'potential' HC: Appendix A.12 for no assumed productivity gap and Appendix A.13 for the measure with a 50% gap. The graphical illustration of these is in Figure 5.6 by education category and in Figure 5.7 as a weighted average.

Figure 5.6 shows that for all education categories, the shares gradually decline with age. This can be explained by individuals exiting the labor market nearing retirement. The difference in shares between groups is large and occurs due to the difference in employment rates. Individuals from the Low education category (Panel A) are the least likely to be employed, while those who are employed work lower hours on average, hence converting the observed earnings to an FTE and imputing earnings for the non-workers significantly increases the HC of this group. Shares increase with education, reflecting an increase in employment rates and hours worked.

The category 'Unknown' stands out for the younger ages: While for other groups the shares are larger for younger groups and then decrease, this is the opposite for those with unknown education. A possible explanation could be that the group composition changes over time: Since the education data has been collected consistently since recently, the group of young people with an unknown education level is very selective. This could signal that they are international degree holders or migrants. Therefore, they would be more likely to resemble the Low Education group, hence the lower share. With age, the group becomes more representative since the education data is missing for the majority of older age individuals and the shares grow.

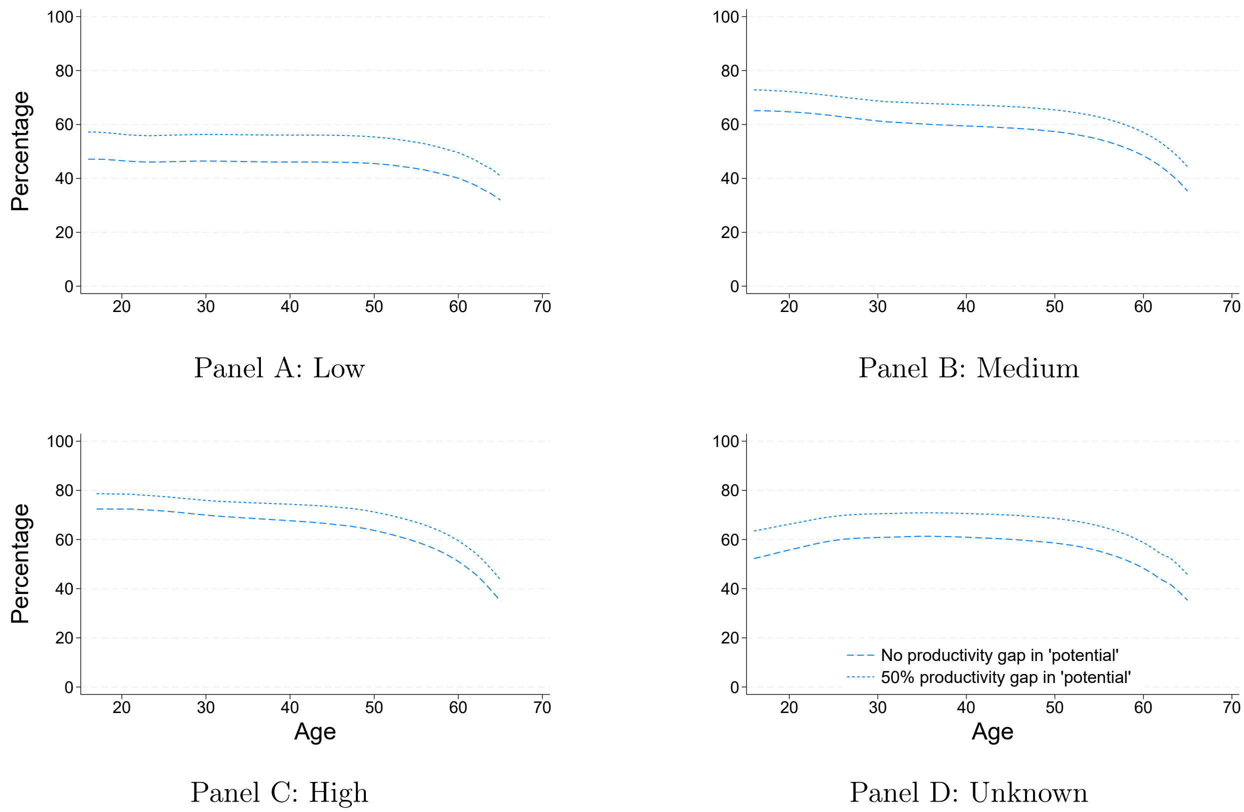


Figure 5.6: Share of ‘realized’ vs ‘potential’ individual HC by education group

When combining all education categories in the weighted average measure the shares are small for younger ages when most individuals are lower educated and the low values of shares of the Low Education group are given more weight. When individuals switch to higher education levels the averages increase since more weight is given to the categories with larger values of shares. Then, as for all groups the shares decrease with age, the averages also experience a sharp decline.

Finally, I compare the measures of Human Capital stock resulting from the ‘realized’ and ‘potential’ methods. Figure 5.8 shows the relative shares of the ‘realized’ measure to the ‘potential’ measure which assumes no productivity gap for the non-workers by education level. A similar figure where instead the 50% productivity gap is assumed for the earnings imputations can be found in Appendix A.1.

The share of the ‘realized’ HC stock increases with education, consistent with the differences in the individual HC shares. These reflect an increase in employment with education. The contribution of converting observed earnings to an FTE to the total ‘potential’ is the largest for lower education levels since more people are working less than full-time. This is not the case for the higher education groups so converting their earnings to an FTE does not make as large a difference for the HC stock. The extra HC coming from assigning earnings

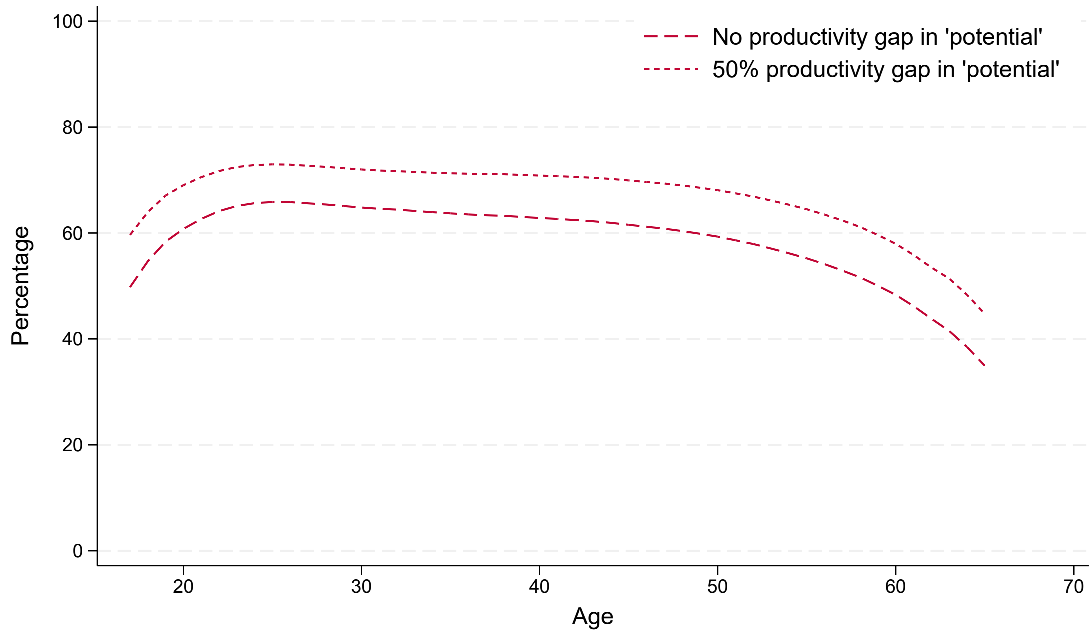


Figure 5.7: Weighted average of the share of 'realized' vs 'potential' individual HC

to the non-workers decreases with education, which is also related to employment rates increasing with education. There are fewer non-workers in the groups with higher education hence the contribution of the imputations of their earnings is low relative to the total stock.

The relative share of 'realized' HC stock is lower for the Unknown Education than for the Medium and High Education categories due to low employment within this group. Those in employment, however, work more than employees from the highest education group. Hence, the conversion of their earnings into full-time equivalent does not significantly increase the measure of HC stock; the difference between the 'potential' and 'realized' measures comes mostly from including the productivity of the non-workers in the total measure.

Figure 5.9 compares the shares of the 'realized' HC stock depending on the assumption regarding the productivity gap in earnings imputations for the 'potential' measure. This allows me to approximate a range for the underestimation of HC stock when only the employees are considered. I find that the gap between the conventional method of measuring the HC stock and the measure that includes the broader definition of productivity ranges between 33 and 41 percentage points (Appendix A.14), subject to the assumption taken about the productivity of the non-workers. Approximately 16 to 19 percentage points of the gap are closed by converting the observed earnings to FTE, and 13 to 24 percentage points by assigning a productivity measure to non-workers. These indicate that both features of the 'potential' calculations are important for the resulting measure.

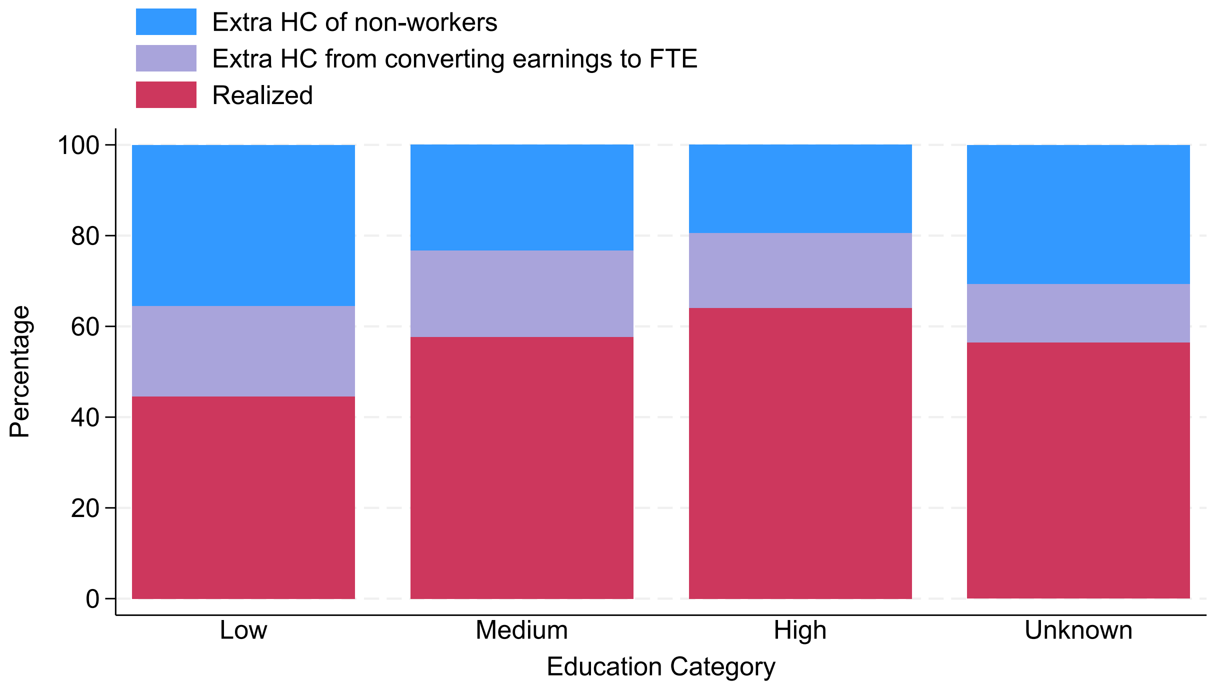


Figure 5.8: Share of ‘realized’ vs ‘potential’ (no productivity gap assumed) HC stock by education group

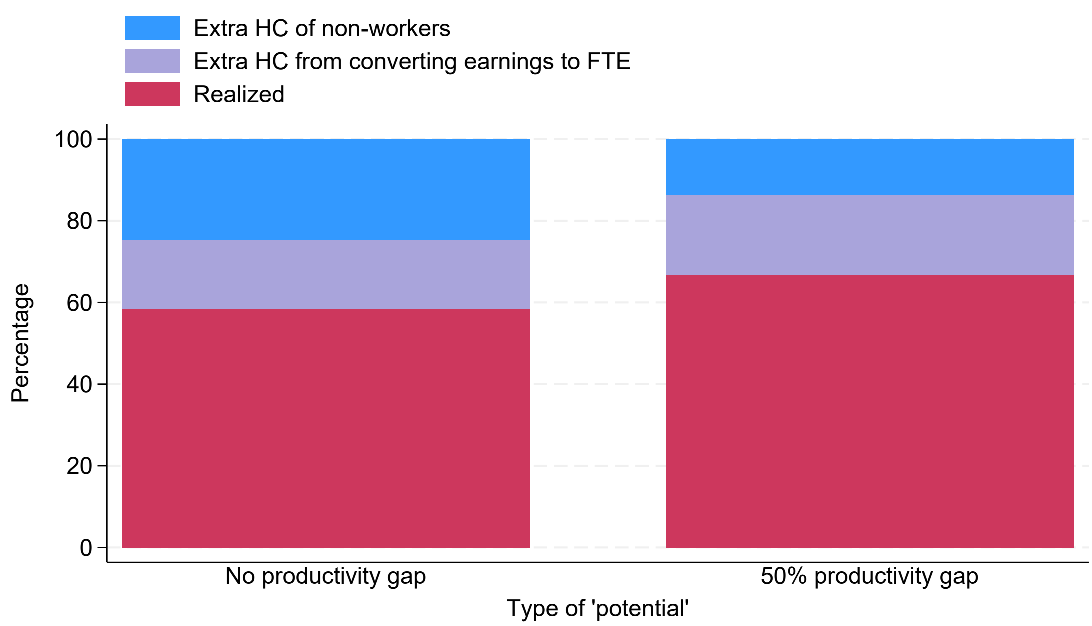


Figure 5.9: Share of ‘realized’ HC stock vs ‘potential’

6 Discussion and Conclusion

I acknowledge the potential limitations of this research. The biggest one lies in imputing the earning potential of the non-workers based on their characteristics or, if available, interpolating this based on the closest earned wage. In this, I assume that an individual out of the labor market is as productive as a comparable worker. This assumption ignores the inherent differences between employees and non-workers. The fact that the former are employed in the labor market may signal superior skill and productivity. Since individual ability and effort are not observed, I have to rely on characteristics like education level, age, gender, migration background, and place of residence to predict the earnings of the non-workers as if they had been working. Using the data on workers for the earnings prediction of the non-workers probably overestimates their potential earnings inflating the 'potential' HC measure.

This is why I develop an alternative measure of 'potential' human capital, where the assumption on absolute comparability between non-workers and employees (conditional on their characteristics) is relaxed. I introduce the productivity gap into the calculations, assuming that a non-worker is half as productive as a comparable worker. This conservative view provides a lower bound for the range in which the true measure of HC 'potential' is. Still, since the productivity gap is unknown, the uncertainty about the resulting measure remains. Hence, it is important to consider the possibility of both.

Furthermore, to simplify the calculations I omit the possibility of switching to a higher education level and assume that individuals stay in the respective education category over their lifetime. This likely results in the underestimation of HC of the younger groups since they are concentrated in the Low Education group. Additionally, when calculating the 'potential' measure I impute full-time earnings equivalent to all individuals (except students) of unobserved productivity (if earnings in the closest periods are not available). This also applies to those with a disability with limited working capacity. Since I do not have the data on disability status, under my method these individuals are assigned an earnings potential equivalent to a 40-hour week while in reality their productivity in the labor force is reduced. This again overestimates the 'potential' measure and hence the resulting proportion of the 'realized' HC is downward biased.

Another limitation of the data that influences the results of this paper arises from features of data collection in the Netherlands. As mentioned earlier, the recording of the education level has not been consistent over the years. This results in a large group of individuals of an unknown education level, especially for older ages. Hence, the most important characteristic for predicting an earnings potential is missing for a large share of the sample. I try to combat this issue by predicting the earnings of these individuals based on the earnings of the workers with an unknown education level using other available characteristics. While this method is not perfect, it presents a good approach in trying to limit the effect of this limitation without leaving a significant portion of the population out of the estimations.

For those with an education level recorded, this information can come from different sources, one of them being the unemployment organization, UWV. Here appear the individuals who were unemployed and looking for a job at some point in their life. This is a selected group of individuals whose earnings are not representative of the broader sample and are likely to be lower. I see several implications emerging from using them in human capital estimation: First, the income profiles of the representative individuals will be negatively affected. Given that these individuals have unrecorded education levels and low chances in the labor market, it is likely that they are older and less educated. Consequently, these groups are the most affected. Secondly, since their income profiles are used to calculate the human capital for both younger and their respective groups, the measures of individual HC and the resulting HC stock are underestimated. Thirdly, the predicted earnings of the non-workers are affected as well since the earnings of the previously unemployed are used in the estimation of the Mincer equation. Due to selection, their earnings are lower than those of a representative individual for whom the educational information comes from registers or EBB, hence the resulting predictions are downward biased. Alternatively, one could disregard the information on education if it comes from UWV and assign those to the group of Unknown Education; however, the potential benefits of this approach are uncertain.

To conclude, I proposed a novel method of estimating the human capital stock of an economy. It improves on the Lifetime Income Approach used in the literature by capturing the productivity of individuals outside of the labor market. I do so using the estimates from the Mincer earnings equations including education level among other personal characteristics to predict the earnings 'potential' of the non-workers as the measure of their productivity. Since the comparability between employees and non-workers is arguable I calculate two measures of productivity for non-workers: (1) based on the assumption that their productivity is similar to that of employees, (2) including a 50% productivity gap for the non-workers, both conditional on other characteristics. Furthermore, I scale all the observed earnings to their full-time equivalent which increases the HC measure by 16 to 19 percentage points, depending on the above-mentioned definition of non-workers' productivity. Students' earnings are set to zero since their 'potential' is unavailable for productive use in the economy during their studies.

These being the main features of the method to estimate the 'potential' HC measure, I compare it to the 'realized' estimate which uses the regular practices from the literature. The resulting relative shares vary from 44% to 71% across education categories and depending on the ‘potential’ measure. These average at 58% when I assume full comparability, and at 66% when a productivity gap of 50% is included. Hence, my findings suggest that the ‘realized’ method underestimates the productive potential of the economy by a range of 33 to 41 percentage points. The gap results from both converting observed earnings to a full-time equivalent and assigning earnings for the individuals of unobserved productivity, with the latter having a more substantial impact. Failing to take account of the non-working population and excluding their productive capacity from the estimations results in drastically different figures, which in the case of HC stock amounts to trillions of Euros. I acknowledge the limitations behind my proposed methodology since the true productivity of non-workers is unobserved and has to be predicted. Hence, the results should be interpreted as falling within a range depending on the hypothesized productivity of non-workers. Most importantly, this research highlights the consequences of measuring human capital while relying on observed earnings only. I believe the large differences between the two measures highlight the shortcomings of the previous literature.

Additionally, Mincer earnings equations are used to estimate returns to education in the years 2015-2022. Comparing these estimates to the findings of Webbink et al. (2013) for the annual income returns and Jacobs and Webbink (2006) for the hourly wage returns, I observe that the returns are substantially larger since last estimated. Additionally, I observe a decline in both returns in the observed period, mostly driven by the decrease in the returns to hours worked. A smaller magnitude of returns to an extra year of schooling compared to the returns to education levels relative to primary schooling supports the idea that the latter returns are exaggerated due to the selection in the reference group. Hence, I argue that the returns to an extra year of schooling are more reliable than the relative ones estimated in this paper. As for yearly income returns to an extra year of education, although decreasing for both sexes, they remain larger for women. Both returns on hourly wage and hours worked are higher for women than for men. The latter difference points towards that working hours increase for women with education more than they do for men. This is reasonable considering that higher-educated women work more hours on average than lower-educated women, while for men the hours worked are less dependent on the education level.

Overall, these findings suggest that higher education became even more rewarded in the labor market compared to the extent it was in the first decade of the 21$^{\text{st}}$ century. This indicates a broader earnings gap between the lower and higher-educated individuals. Even though the returns decreased between 2015 and 2022, they remain much higher compared to the earlier years, so the large gap remains.

Future research could build upon the approach to estimating ’potential’ human capital proposed in this thesis by refining the methods used. Additionally, it would be beneficial to compare 'realized' and 'potential' human capital measures over time. Such a dynamic view would help determine whether the 'potential' measure is truly resistant to economic cycle fluctuations and how the relationship between the two measures evolves. Furthermore, considering data availability, the returns to education should be studied for the currently un-analyzed period (2009-2014) to provide insight into how the returns changed during this gap. It would be also important to re-estimate the returns for the periods explored in Webbink et al. (2013) and Jacobs and Webbink (2006) to ensure comparability of the estimates.

References

Adema, Y., & van Tilburg, I. V. (2019). Zorgen om morgen. *Centraal Planbureau (CPB)*.

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

Ahlroth, S., Björklund, A., & Forslund, A. (1997). The output of the swedish education sector. *Review of Income and Wealth*, *43*(1), 89–104.

Amo-Agyei, S., et al. (2020). *The migrant pay gap: Understanding wage differences between migrants and nationals*. International Labour Organisation (ILO).

Aulin-Ahmavaara, P. (2002). Human capital as a produced asset.

Christian, M. S. (2010). *Human capital accounting in the united states: 1994 to 2006*. BEA.

Christian, M. S. (2014). Human capital accounting in the united states: Context, measurement, and application. In *Measuring economic sustainability and progress* (pp. 461–491). University of Chicago Press.

Christian, M. S. (2017). *Net investment and stocks of human capital in the united states, 1975-2013*. CIGI.

Dagum, C., & Slottje, D. J. (2000). A new method to estimate the level and distribution of household human capital with application. *Structural Change and Economic Dynamics*, *11*(2), 67–94.

Graham, J. W., & Webb, R. H. (1979). Stocks and depreciation of human capital: New evidence from a present-value perspective. *Review of Income and Wealth*, *25*(2), 209–224.

Groot, S. P., de Groot, H. L., & Smit, M. J. (2014). Regional wage differences in the netherlands: Micro evidence on agglomeration externalities. *Journal of Regional Science*, *54*(3), 503–523.

Gu, W., & Wong, A. (2010). Estimates of human capital in canada: The lifetime income approach. *Available at SSRN 1711935*.

Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. In *Annals of economic and social measurement, volume 5, number 4* (pp. 475–492). NBER.

Heckman, J. J., Lochner, L., & Todd, P. E. (2003). Fifty years of mincer earnings regressions.

Jacobs, B., & Webbink, D. (2006). Het rendement op onderwijs blijft stijgen. *Economisch-Statistische Berichten*, *91*(Jubileumnummer), 33–36.

Jones, R., & Chiripanhura, B. (2010). Measuring the uk's human capital stock. *Economic & Labour Market Review*, *4*, 36–63.

Jorgenson, D. W., & Fraumeni, B. M. (1989). The Accumulation of Human and Nonhuman Capital, 1948-84. *NBER Chapters*, 227–286. https://ideas.repec.org/h/nbr/nberch/8121.html

Jorgenson, D. W., & Fraumeni, B. M. (1992). Investment in education and u.s. economic growth. *The Scandinavian Journal of Economics*, *94*, 51–70. https://doi.org/10.2307/3440246

Le, T., Gibson, J., & Oxley, L. (2003). Cost-and income-based measures of human capital. *Journal of economic surveys*, *17*(3), 271–307.

Le, T., Gibson, J., & Oxley, L. (2006). A forward-looking measure of the stock of human capital in new zealand. *The Manchester School*, *74*(5), 593–609.

Leuven, E., & Oosterbeek, H. (2000). Rendement van onderwijs stijgt. *ESB*, *85*(4262), 523–524.

Li, H., Fraumeni, B. M., Liu, Z., & Wang, X. (2009). *Human capital in china* (tech. rep.). National Bureau of Economic Research.

Liu, G. (2011). Measuring the stock of human capital for comparative analysis: An application of the lifetime income approach to selected countries.

Liu, G. (2018). Measuring human capital for norway 2007-2014.

Liu, G., & Greaker, M. (2009). *Measuring the stock of human capital for norway. a lifetime labour income approach* (tech. rep. No. 2009/12). Statistics Norway. Statistics Norway.

Lucas, R. E. (1988). On the mechanics of economic development. *Journal of Monetary Economics*, *22*(1), 3–42.

Mincer, J. A. (1974). Schooling and earnings. In *Schooling, experience, and earnings* (pp. 41–63). NBER.

OECD. (2016). Human capital [Retrieved from https://www.oecd.org/economy/human-capital/].

O'Mahony, M., & Samek, L. (2021). *Incorporating health status into human capital stocks: An analysis for the uk* (tech. rep.). Economic Statistics Centre of Excellence (ESCoE).

O'Mahony, M., & Stevens, P. (2004). International comparisons of performance in the provision of public services: Outcome based measures for education. *London: National Institute of Economic and Social Research.*

Patrinos, H. A. (2016). Estimating the return to schooling using the mincer equation. *IZA World of Labor.*

Petty, W. (1690). *Political arithmetik* [Reprinted in C. H. Hull (Ed.), *The Economic Writings of Sir William Petty* (1899). Cambridge: Cambridge University Press].

Rensman, M. (2013). *Human capital in the netherlands.* Statistics Netherlands.

Romer, P. M. (1989). *Human capital and growth: Theory and evidence* (Working Paper No. 3173). National Bureau of Economic Research.

Rothschild, M. (1992). Comment on 'output of the education sector'. In Z. Griliches (Ed.), *Output measurement in the services sector* (pp. 339–341). The University of Chicago Press.

UNECE. (2016). *Guide on measuring human capital* (tech. rep.) (Retrieved June 19, 2024, from https://unece.org/statistics/publications/guide-measuring-human-capital). United Nations Economic Commission for Europe.

Webber, D. A. (2015). Firm market power and the earnings distribution. *Labour Economics*, *35*, 123–134.

Webbink, D., Gerritsen, S., & van der Steeg, M. (2013). Financiële opbrengsten onderwijs verder omhoog. *Economisch-Statistische Berichten*, *98*, 13–15.

Wei, H. (2008a). *Developments in the estimation of the value of human capital for australia*. International Association for Official Statistics.

Wei, H. (2008b). *Developments in the estimation of the value of human capital for australia*. International Association for Official Statistics.

Wickens, C. H. (1924). Human capital [Cited in Treadgold (2000)]. *Report of the Sixteenth Meeting of the Australasian Association for the Advancement of Science*, 526–554.

World Bank. (2006). *Where is the wealth of nations?*

A Appendix

Table A.1: Survival Probabilities by Age

Age	P(Survival)
16	0,9998
17	0,9998
18	0,9998
19	0,9997
20	0,9998
21	0,9997
22	0,9998
23	0,9997
24	0,9997
25	0,9997
26	0,9997
27	0,9996
28	0,9996
29	0,9996
30	0,9995
31	0,9995
32	0,9996
33	0,9996
34	0,9995
35	0,9995
36	0,9994
37	0,9994
38	0,9993
39	0,9993
40	0,9992
41	0,9990
42	0,9991
43	0,9989

Continued on next page

Age	P(Survival)
44	0,9988
45	0,9987
46	0,9985
47	0,9983
48	0,9983
49	0,9982
50	0,9978
51	0,9974
52	0,9975
53	0,9969
54	0,9965
55	0,9964
56	0,9956
57	0,9953
58	0,9949
59	0,9941
60	0,9937
61	0,9930
62	0,9922
63	0,9917
64	0,9906
65	0,9898

Note: Survival probabilities for each age group are reported in column (2).

Table A.2: Descriptive Statistics of the sub-sample for the Mincer earnings equations

	Primary Education	VMBO	MBO/HAVO/VWO	HBO	WO	Doctorate	Total
Hourly Wage	13.09	12.08	17.31	26.16	31.97	42.58	20.59
	(7.54)	(7.91)	(9.24)	(13.09)	(21.83)	(24.07)	(14.33)
Annual Income	14619.59	13173.06	22447.91	37830.44	47489.42	67085.18	28045.54
	(13843.60)	(14213.94)	(17000.51)	(22856.00)	(36169.64)	(41437.08)	(24851.98)
Yearly Hours	947.63	890.24	1199.26	1403.98	1401.46	1547.53	1219.22
	(654.67)	(625.51)	(574.79)	(450.60)	(499.69)	(359.29)	(578.81)
Age	39.40	31.75	35.74	40.43	38.91	44.55	36.87
	(16.67)	(16.13)	(13.61)	(11.37)	(11.21)	(9.86)	(13.69)
Female	0.44	0.47	0.49	0.53	0.50	0.45	0.49
Observations	2,121,180	7,341,219	20,932,412	11,054,808	7,105,099	218,034	48,772,752

Note: The sample is split into groups based on the highest-achieved education level. For the variables Hourly Wage, Yearly Compensation, Yearly Hours, and Age the mean and standard deviation (in parentheses) are shown. For the binary variable Female the fraction of the population for whom the value is equal to one is shown. Yearly Compensation and Hourly Wages are measured in Euros. Age is measured in years.

Table A.3: Hourly wage returns per level of education, 2015-2022

Education Level	2015	2016	2017	2018	2019	2020	2021	2022
VMBO	11.58	11.50	11.74	11.83	12.68	9.91	9.33	10.79
	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
MBO/HAVO/VWO	48.62	50.25	51.44	51.39	52.47	49.35	49.26	49.88
	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
HBO	111.02	113.28	113.32	111.84	112.53	107.87	106.23	106.52
	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
WO	166.54	171.31	170.66	168.81	167.92	161.85	160.16	157.93
	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
Doctorate	236.65	238.16	239.95	231.55	229.99	220.23	223.00	218.01
	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.65	0.67	0.67	0.66	0.66	0.65	0.65	0.64
Observations	5,341,489	5,558,097	5,822,177	6,150,668	6,255,588	6,360,727	6,545,166	6,738,837

*Note: The reported returns are the exponentially transformed coefficients from the analyses performed on the sample of employees with a known education level; measured in percent. The set of controls includes Experience, Experience-squared, and a binary variable for sex. The robust standard errors of the original coefficients are given in parentheses. * 10%, ** 5%, *** 1%.*

Table A.4: Returns on hours worked per level of education, 2015-2022

Education Level	2015	2016	2017	2018	2019	2020	2021	2022
VMBO	16.35	18.31	20.47	20.47	20.35	12.63	10.49	12.39
	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
MBO/HAVO/VWO	64.76	66.63	69.51	67.63	67.30	53.52	52.83	51.40
	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
HBO	109.90	107.83	106.66	100.58	100.41	81.59	81.07	76.77
	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
WO	119.14	114.64	113.92	106.03	104.55	86.31	86.67	81.57
	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
Doctorate	138.80	136.41	136.51	122.48	124.64	96.54	97.52	87.43
	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.23	0.22	0.22	0.23	0.23	0.22	0.23	0.22
Observations	5,341,489	5,558,097	5,822,177	6,150,668	6,255,588	6,360,727	6,545,166	6,738,837

*Note: The reported returns are the exponentially transformed coefficients from the analyses performed on the sample of employees with a known education level; measured in percent. The set of controls includes Experience, Experience-squared, and a binary variable for sex. The robust standard errors of the original coefficients are given in parentheses. * 10%, ** 5%, *** 1%.*

Table A.5: Returns on annual income per level of education, 2015-2022

Education Level	2015	2016	2017	2018	2019	2020	2021	2022
VMBO	29.82	31.92	34.62	34.72	35.62	23.79	20.80	24.52
	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
MBO/HAVO/VWO	144.86	150.37	156.70	153.78	155.08	129.28	128.11	126.91
	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
HBO	342.94	343.25	340.86	324.90	325.93	277.46	273.43	265.07
	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
WO	484.10	482.33	479.01	453.84	448.04	387.86	385.64	368.32
	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
Doctorate	703.90	699.42	704.00	637.62	641.28	529.38	538.00	496.05
	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.46	0.46	0.47	0.47	0.48	0.46	0.47	0.47
Observations	5,341,489	5,558,097	5,822,177	6,150,668	6,255,588	6,360,727	6,545,166	6,738,837

*Note: The reported returns are the exponentially transformed coefficients from the analyses performed on the sample of employees with a known education level; measured in percent. The set of controls includes Experience, Experience-squared, and a binary variable for sex. The robust standard errors of the original coefficients are given in parentheses. * 10%, ** 5%, *** 1%.*

Table A.6: Wage returns to an extra year of schooling, 2015-2022

Year	Women	Men	All
2015	11.11	11.20	11.18
	(0.00)***	(0.00)***	(0.00)***
	[0.63]	[0.64]	[0.63]
2016	11.37	11.30	11.34
	(0.00)***	(0.00)***	(0.00)***
	[0.65]	[0.66]	[0.65]
2017	11.32	11.27	11.30
	(0.00)***	(0.00)***	(0.00)***
	[0.66]	[0.66]	[0.65]
2018	11.24	11.08	11.15
	(0.00)***	(0.00)***	(0.00)***
	[0.65]	[0.65]	[0.64]
2019	11.10	11.00	11.03
	(0.00)***	(0.00)***	(0.00)***
	[0.65]	[0.65]	[0.65]
2020	11.16	10.86	10.99
	(0.00)***	(0.00)***	(0.00)***
	[0.64]	[0.64]	[0.63]
2021	11.17	10.87	11.01
	(0.00)***	(0.00)***	(0.00)***
	[0.63]	[0.64]	[0.63]
2022	10.98	10.66	10.80
	(0.00)***	(0.00)***	(0.00)***
	[0.63]	[0.63]	[0.62]
Controls	Yes	Yes	Yes
Observations	>2.5M	>2.5M	>5M

*Note: The reported returns are the exponentially transformed coefficients from the analyses performed on the sample of employees with a known education level. The set of controls includes Experience, Experience-squared, and a binary variable for sex. The robust standard errors of the original coefficients are given in parentheses; R-squared are given in square brackets. * 10%, ** 5%, *** 1%.*

Table A.7: Hours worked returns to an extra year of schooling, 2015-2022

Year	Women	Men	All
2015	9.93	6.93	8.30
	(0.00)***	(0.00)***	(0.00)***
	[0.19]	[0.24]	[0.20]
2016	9.41	6.52	7.80
	(0.00)***	(0.00)***	(0.00)***
	[0.17]	[0.23]	[0.19]
2017	9.12	6.24	7.51
	(0.00)***	(0.00)***	(0.00)***
	[0.18]	[0.24]	[0.20]
2018	8.61	5.80	7.04
	(0.00)***	(0.00)***	(0.00)***
	[0.18]	[0.24]	[0.20]
2019	8.33	5.84	6.92
	(0.00)***	(0.00)***	(0.00)***
	[0.18]	[0.24]	[0.20]
2020	7.74	5.08	6.24
	(0.00)***	(0.00)***	(0.00)***
	[0.17]	[0.23]	[0.20]
2021	7.83	5.13	6.33
	(0.00)***	(0.00)***	(0.00)***
	[0.18]	[0.24]	[0.20]
2022	7.17	4.78	5.84
	(0.00)***	(0.00)***	(0.00)***
	[0.18]	[0.24]	[0.20]
Controls	Yes	Yes	Yes
Observations	>2.5M	>2.5M	>5M

*Note: The reported returns are the exponentially transformed coefficients from the analyses performed on the sample of employees with a known education level. The set of controls includes Experience, Experience-squared, and a binary variable for sex. The robust standard errors of the original coefficients are given in parentheses; R-squared are given in square brackets. * 10%, ** 5%, *** 1%.*

Table A.8: Annual income returns to an extra year of schooling, 2015-2022

Year	Women	Men	All
2015	22.14	18.91	20.41
	(0.00)***	(0.00)***	(0.00)***
	[0.41]	[0.47]	[0.43]
2016	21.85	18.56	20.03
	(0.00)***	(0.00)***	(0.00)***
	[0.41]	[0.48]	[0.43]
2017	21.47	18.21	19.66
	(0.00)***	(0.00)***	(0.00)***
	[0.42]	[0.49]	[0.44]
2018	20.81	17.52	18.98
	(0.00)***	(0.00)***	(0.00)***
	[0.42]	[0.49]	[0.44]
2019	20.36	17.48	18.71
	(0.00)***	(0.00)***	(0.00)***
	[0.42]	[0.50]	[0.45]
2020	19.77	16.48	17.92
	(0.00)***	(0.00)***	(0.00)***
	[0.41]	[0.48]	[0.43]
2021	19.87	16.55	18.04
	(0.00)***	(0.00)***	(0.00)***
	[0.42]	[0.49]	[0.44]
2022	18.93	15.95	17.27
	(0.00)***	(0.00)***	(0.00)***
	[0.42]	[0.49]	[0.44]
Controls	Yes	Yes	Yes
Observations	>2.5M	>2.5M	>5M

*Note: The reported returns are the exponentially transformed coefficients from the analyses performed on the sample of employees with a known education level. The set of controls includes Experience, Experience-squared, and a binary variable for sex. The robust standard errors of the original coefficients are given in parentheses; R-squared are given in square brackets. * 10%, ** 5%, *** 1%.*

Table A.9: Mincer Earnings Estimates for Prediction

Education Level	log(FTE Comp)
VMBO	0.12
	(0.00)***
MBO/HAVO/VWO	0.45
	(0.00)***
HBO	0.82
	(0.00)***
WO	1.07
	(0.00)***
Doctorate	1.28
	(0.00)***
Controls	Yes
R-squared	0.67
Years	2019
Observations	6,283,509

*Note: The analysis is performed on the sample of employees with a known education level. The set of controls includes Experience, Experience-squared, binary variables for sex, region of residency, and migration background. The robust standard errors are given in parentheses. * 10%, ** 5%, *** 1%.*

Table A.10: Estimates of FTE Compensation prediction for Unknown Education category

	log(FTE Compensation)
Experience	0,03
	(0,00)***
Experience-squared	-0,00
	(0,00)***
Woman	-0,22
	(0,00)***
Randstad	0,12
	(0,00)***
Native	0,07
	(0,00)***
Intercept	10,20
	(0,00)***
R-squared	0,13
Observations	1,864,745

*Note: The robust standard errors are given in parentheses. * 10%, ** 5%, *** 1%.*

Table A.11: Average income profiles depending on the method of HC calculation per age group

Age	'Realized'	'FTE'	'Potential'	
			No productivity gap	50% productivity gap
16				
17	1983.33	9590.10	2167.41	1142.36
18	3122.64	11632.78	2406.30	1485.96
19	4581.52	13913.34	4538.19	3430.74
20	6251.36	16483.26	7142.08	5716.83
21	8480.91	19745.75	10161.28	8466.95
22	11315.50	22450.78	13943.73	12030.93

Continued on next page

Age	'Realized'	'FTE'	'Potential'	
			No productivity gap	50% productivity gap
23	13871.55	24363.93	17357.74	15279.87
24	16750.33	26219.97	21069.10	18916.84
25	19909.38	28203.09	24979.33	22807.56
26	22959.81	30106.73	29023.18	26786.97
27	25293.09	31762.89	32394.83	30037.35
28	27175.36	33351.90	35352.37	32837.34
29	28510.37	34666.23	37859.60	35129.63
30	29467.76	35847.73	39977.43	37009.30
31	30489.44	37222.90	43677.36	40450.13
32	31209.81	38089.45	45029.74	41559.60
33	31937.12	39068.74	46478.21	42773.48
34	32554.46	39899.99	47812.59	43856.29
35	33161.38	40681.21	49079.50	44880.35
36	33725.51	41412.61	50215.19	45813.90
37	34313.63	42047.19	51284.06	46665.63
38	35009.32	42901.41	52449.81	47675.61
39	35498.46	43466.24	53365.83	48416.03
40	35832.31	43770.14	53923.84	48846.99
41	36382.48	44280.68	54754.36	49517.52
42	36631.29	44684.49	55564.51	50124.50
43	37228.04	45028.09	55994.03	50511.06
44	37568.57	45167.26	56495.87	50831.57
45	37249.66	44674.89	56255.95	50465.42
46	37733.51	45672.59	57385.09	51528.84
47	37734.71	44858.11	56741.60	50799.86
48	37800.35	44895.17	56950.70	50922.94
49	37701.47	44662.52	56800.40	50731.46
50	37812.18	44723.00	56983.34	50853.17
51	37293.74	44167.43	56598.13	50382.77
52	37204.12	44017.61	56495.04	50256.33
53	36422.04	43428.31	56025.36	49726.84
54	36006.29	42941.66	55687.38	49314.52

Continued on next page

Age	'Realized'	'FTE'	'Potential'	
			No productivity gap	50% productivity gap
55	35440.16	42269.18	55244.35	48756.77
56	34875.98	42124.70	55199.84	48662.27
57	33834.13	40912.40	54443.60	47678.00
58	33420.46	40844.09	54625.84	47734.96
59	31995.59	39686.14	53949.80	46817.97
60	31048.63	38952.90	53706.27	46329.59
61	29180.72	37707.44	53195.73	45451.59
62	27170.05	37985.48	54435.08	46210.28
63	24408.18	34021.79	51737.92	42879.86
64	21211.19	31369.35	50695.09	41032.22
65	17249.19	27761.84	49352.91	38557.38

Note: Average incomes are reported in 2019 Euros. Values are calculated as weighted averages per age group based on average incomes of education groups. Values for the group of age 16 is omitted since it includes empty bins to ensure the non-disclosure of identifiable data when split by education level.

Table A.12: Shares of 'realized' individual HC relative to the 'potential' (no productivity gap)

	Education Levels				
Age	Low	Medium	High	Unknown	Overall
16	47.10	65.13		52.25	
17	47.15	65.11	72.40	53.10	49.77
18	47.04	65.00	72.43	54.00	54.66
19	46.82	64.87	72.39	54.92	58.36
20	46.57	64.71	72.39	55.77	60.78
21	46.35	64.51	72.39	56.59	62.66
22	46.20	64.26	72.21	57.44	64.13
23	46.07	63.96	72.03	58.28	65.11
24	46.08	63.60	71.84	59.00	65.64
25	46.16	63.23	71.60	59.62	65.85

Continued on next page

	Education Levels				
Age	Low	Medium	High	Unknown	Overall
26	46.23	62.83	71.31	60.07	65.83
27	46.28	62.43	70.99	60.37	65.60
28	46.36	62.04	70.64	60.60	65.39
29	46.41	61.66	70.29	60.73	65.09
30	46.42	61.31	69.95	60.84	64.81
31	46.40	60.99	69.63	60.92	64.57
32	46.37	60.81	69.39	61.01	64.41
33	46.33	60.61	69.15	61.15	64.17
34	46.27	60.40	68.93	61.26	63.93
35	46.24	60.22	68.71	61.32	63.70
36	46.18	60.03	68.50	61.31	63.51
37	46.13	59.87	68.30	61.27	63.35
38	46.08	59.73	68.08	61.20	63.25
39	46.08	59.59	67.88	61.07	63.04
40	46.08	59.46	67.66	60.93	62.84
41	46.07	59.33	67.43	60.78	62.67
42	46.07	59.19	67.17	60.62	62.44
43	46.08	59.04	66.92	60.48	62.22
44	46.06	58.86	66.60	60.29	61.93
45	46.02	58.69	66.24	60.07	61.55
46	45.99	58.49	65.87	59.83	61.19
47	45.93	58.27	65.52	59.55	60.82
48	45.83	58.00	64.99	59.26	60.38
49	45.68	57.69	64.40	58.93	59.86
50	45.50	57.35	63.70	58.57	59.32
51	45.24	56.97	62.92	58.12	58.63
52	44.97	56.45	62.08	57.61	57.92
53	44.53	55.89	61.14	56.96	57.06
54	44.11	55.25	60.18	56.19	56.17
55	43.66	54.49	59.11	55.27	55.22
56	43.13	53.57	57.88	54.23	54.13
57	42.43	52.53	56.50	53.06	52.92

Continued on next page

	Education Levels				
Age	Low	Medium	High	Unknown	Overall
58	41.70	51.35	54.98	51.72	51.62
59	40.88	49.99	53.16	50.12	50.03
60	40.09	48.43	51.10	48.35	48.32
61	38.85	46.57	48.65	46.20	46.20
62	37.50	44.37	46.01	43.76	43.82
63	35.79	41.80	42.76	41.85	41.52
64	34.16	38.87	39.00	38.83	38.48
65	31.96	35.29	35.01	35.26	34.95

Note: Columns (2) - (5) report the proportion of individual 'Realized' HC relative to 'Potential' (no productivity gap assumed) by education category measured in percent. Column (6) shows the weighted average of all education levels, representing the share for the whole age group. Bins with <10 observations were deleted to ensure the non-disclosure of identifiable data.

Table A.13: Shares of 'realized' individual HC relative to the 'potential' (50% productivity gap)

	Education Levels				
Age	Low	Medium	High	Unknown	Overall
16	57.20	72.85		63.47	
17	57.18	72.77	78.65	64.14	59.63
18	57.01	72.61	78.63	64.84	63.95
19	56.70	72.42	78.53	65.56	67.07
20	56.38	72.21	78.49	66.24	69.03
21	56.12	71.95	78.44	66.90	70.54
22	55.95	71.65	78.18	67.62	71.70
23	55.84	71.30	77.96	68.31	72.45
24	55.88	70.93	77.72	68.89	72.83
25	55.98	70.55	77.45	69.40	72.96
26	56.07	70.17	77.16	69.79	72.91
27	56.13	69.79	76.85	70.06	72.69
28	56.22	69.42	76.54	70.26	72.49

Continued on next page

	Education Levels				
Age	Low	Medium	High	Unknown	Overall
29	56.28	69.07	76.23	70.39	72.23
30	56.30	68.75	75.94	70.50	71.99
31	56.28	68.45	75.67	70.57	71.80
32	56.27	68.33	75.49	70.64	71.68
33	56.25	68.18	75.32	70.76	71.52
34	56.21	68.03	75.17	70.84	71.38
35	56.19	67.89	75.02	70.88	71.27
36	56.14	67.75	74.89	70.88	71.19
37	56.11	67.63	74.75	70.85	71.13
38	56.07	67.51	74.61	70.79	71.09
39	56.07	67.41	74.49	70.69	70.97
40	56.08	67.31	74.35	70.59	70.85
41	56.06	67.20	74.19	70.48	70.74
42	56.05	67.09	74.02	70.35	70.58
43	56.06	66.96	73.84	70.26	70.43
44	56.03	66.80	73.61	70.11	70.21
45	55.99	66.65	73.32	69.94	69.93
46	55.95	66.47	73.02	69.73	69.64
47	55.86	66.26	72.77	69.48	69.35
48	55.74	66.01	72.32	69.21	68.98
49	55.58	65.72	71.82	68.91	68.55
50	55.38	65.42	71.21	68.58	68.08
51	55.09	65.05	70.51	68.16	67.49
52	54.79	64.57	69.76	67.69	66.88
53	54.34	64.06	68.90	67.09	66.13
54	53.89	63.48	68.03	66.38	65.35
55	53.41	62.76	67.05	65.52	64.49
56	52.85	61.90	65.91	64.52	63.48
57	52.11	60.96	64.63	63.43	62.37
58	51.32	59.85	63.20	62.11	61.13
59	50.44	58.57	61.49	60.56	59.61
60	49.57	57.09	59.54	58.80	57.95

Continued on next page

	Education Levels				
Age	Low	Medium	High	Unknown	Overall
61	48.22	55.33	57.16	56.63	55.86
62	46.81	53.23	54.61	54.10	53.49
63	44.94	50.71	51.38	52.58	51.40
64	43.37	47.82	47.65	49.45	48.35
65	40.93	44.24	43.70	45.70	44.74

Note: Columns (2) - (5) report the proportion of individual 'Realized' HC relative to 'Potential' (50% productivity gap assumed) by education category measured in percent. Column (6) shows the weighted average of all education levels, representing the share for the whole age group. Bins with <10 observations were deleted to ensure the non-disclosure of identifiable data.

Table A.14: Share of 'realized' vs 'potential' HC stock by education group

Measure	Low	Medium	High	Unknown	Total
No productivity gap in 'potential'					
Share 'realized'	44.55	57.66	64.05	56.42	58.34
Extra HC, FTE	20.00	19.07	16.51	12.99	16.90
Extra HC, non-workers	35.45	23.26	19.44	30.59	24.76
50% productivity gap in 'potential'					
Share 'realized'	54.35	65.58	71.02	66.47	66.76
Extra HC, FTE	24.44	21.80	18.49	15.40	19.50
Extra HC, non-workers	21.21	12.62	10.48	18.12	13.74

Note: Share 'realized' reports the proportion of 'realized' HC stock relative to 'potential' measured in percent. Extra HC, FTE, and Extra HC, non-working, show the increase in the stock measure from FTE conversion of employees' earnings and earnings imputations for the non-workers, respectively, measured in percent, relative to the total value of the 'potential' measure.

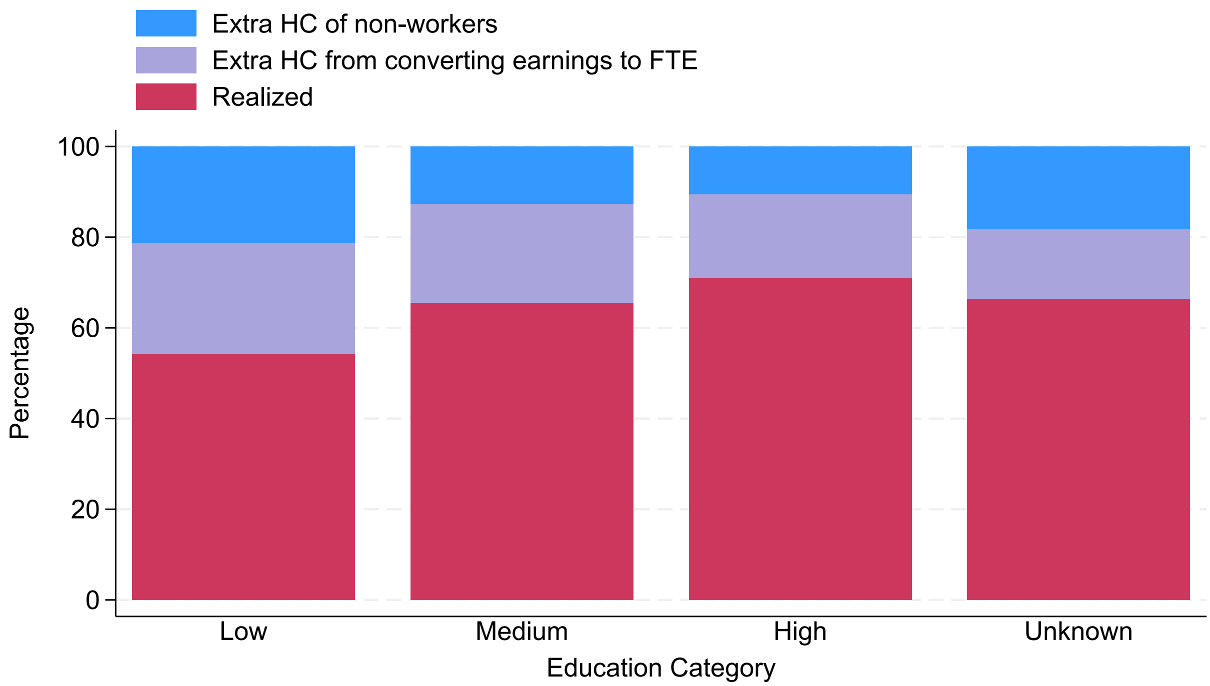


Figure A.1: Share of 'realized' vs 'potential' (50% productivity gap assumed) HC stock by education group