CHARLES UNIVERSITY
FACULTY OF SOCIAL SCIENCES
Institute of Economic Studies

Standing Tall Pays Off: A Meta-Analysis of Height Premium

Master's thesis

Author: Bc. Martina Juračková
Study program: Economics and Finance
Supervisor: doc. PhDr. Zuzana Havránková, Ph.D.
Year of defense: 2023

Declaration of Authorship

The author hereby declares that he or she compiled this thesis independently, using only the listed resources and literature, and the thesis has not been used to obtain any other academic title.

The author grants to Charles University permission to reproduce and to distribute copies of this thesis in whole or in part and agrees with the thesis being used for study and scientific purposes.

Prague, August 1, 2023

Martina Juračková

Abstract

As has been demonstrated by empirical research, height is an important physical feature impacting various aspects of the life of an individual. This thesis deals with the relationship between height and income, also referred to as height premium. With the help of modern meta-analytic methods, we aim to quantitatively summarize the empirical evidence on the impact of height on income. After introducing the topic of height premium, data collection and methodological framework, we test for publication bias. The analysis is conducted on 1084 height premium estimates collected from 67 studies. The results of publication bias testing indicate that height premium literature contains positive publication bias which persists even after we control for additional variables capturing study characteristics or, in other words, the heterogeneity of collected estimates. Based on Bayesian Model Averaging results, we conclude that geographical factors, the longitudinal nature of the dataset, restriction of the dataset with respect to gender, or adding a gender control variable into the regression are the most important factors explaining the variability of height premium effects.

JEL Classification	C11, E24, J31, J41
Keywords	height, wage, wage determinants, Mincer equation, meta-analysis, publication bias, Bayesian Model Averaging
Title	Standing Tall Pays Off: A Meta-Analysis of Height Premium

Abstrakt

Jak bylo prokázáno v akademické literatuře, výška je důležitý faktor ovlivňující různé aspekty života jednotlivce. Tato práce se zabývá vztahem mezi výškou a příjmy, jinak také nazývaným jako výškové premium. Naším cílem je za použití moderních metod meta-analýzy kvantitativně shrnout výsledky literatury zabývající se vlivem výšky na výši příjmů. Po představení výškového premia, procesu sběru dat a metodologických principů zkoumáme přítomnost publikační selektivity. Analýza je provedena na 1084 efektech výškového premia sesbíraných z 67 studií. Výsledky testů publikační analýzy naznačují, že se v literatuře zabývající se výškovým premiem nachází pozitivní publikační bias, který v literatuře zůstává, i když bereme v úvahu i proměnné zachycující vlastnosti studie neboli jinými slovy heterogeneitu sesbíraných efektů. Na základě výsledků Bayesovského průměrování modelů pak děláme závěr, že nejdůležitějšími faktory při zachycování variability efektů výškového premia je geografická lokace, longitudinální povaha datasetu, omezení dat na základě pohlaví, či zahnutí kontrolní proměnné pro pohlaví do regrese.

Klasifikace JEL	C11, E24, J31, J41
Klíčová slova	výška, mzda, determinanty mzdy, Mincerova rovnice, meta-analýza, publikační selektivita, Bayesovské průměrování modelů
Název práce	Stát zpříma se vyplácí: Meta-analýza výškového prémia

Acknowledgments

The author of this thesis would like to express great gratitude to doc. PhDr. Zuzana Havránková, Ph.D. for her guidance, willingness, helpful comments, and expertise provided during writing this thesis. Further, the author is grateful to prof. doc. PhDr. Tomáš Havránek, Ph.D. and PhDr. Jiří Schwarz, Ph.D. for their comments. Finally, the author would like to thank her family and friends for support and encouragement during her studies.

Typeset in LaTeXusing the IES Thesis Template.

Bibliographic Record

Juračková, Martina: *Standing Tall Pays Off: A Meta-Analysis of Height Premium.* Master's thesis. Charles University, Faculty of Social Sciences, Institute of Economic Studies, Prague. 2023, pages 110. Advisor: doc. PhDr. Zuzana Havránková, Ph.D.

Contents

List of Tables viii

List of Figures ix

Acronyms x

Thesis Proposal xi

1 Introduction 1

2 On height and its link to income 4

2.1 Height . . . 4

2.1.1 Why is height important . . . 4

2.1.2 What determines height . . . 6

2.1.3 How is height measured . . . 8

2.2 Link height-income . . . 9

2.2.1 Literature review . . . 9

2.2.2 Computation . . . 12

2.3 Issues and biases associated with height-income relationship measurement . . . 13

3 Data 17

3.1 Data collection . . . 17

3.2 Data adjustments . . . 19

3.3 Basic summary statistics . . . 21

4 Publication bias 26

4.1 Definition of publication bias . . . 26

4.2 Testing for publication bias . . . 27

4.2.1 Funnel plot . . . 27

4.2.2 Linear tests . . . 30
4.2.3 The problem of endogeneity . . . 32
4.2.4 Non linear tests . . . 36

5 Heterogeneity **39**
5.1 Explanatory variables . . . 39
5.2 Estimation method . . . 44
5.2.1 BMA explanation . . . 47
5.2.2 BMA implementation . . . 50
5.2.3 FMA . . . 51
5.2.4 Frequentist check . . . 51
5.3 Results . . . 52
5.4 Best practice estimate . . . 57

6 Conclusion **61**

Bibliography **82**

A Overview of height premium studies in specific countries **I**

B Other physical features in Mincer equation - elaborated **III**

C Additional characteristics of the causal and noncausal subsample of height premium estimates **VIII**

D Diagnostics for BMA **XIV**

List of Tables

3.1 List of studies included in the meta-analysis 20
3.2 Descriptive statistics of the effects collected from the primary studies 25

4.1 Linear tests for publication bias 33
4.2 Tests for publication bias with relaxed exogeneity assumption . 35
4.3 Nonlinear tests for publication bias 38

5.1 Descriptive statistics of variables used in the heterogeneity analysis 45
5.2 Explaining heterogeneity in height premium estimates - BMA and Frequentist check specification 58
5.3 Explaining heterogeneity in height premium estimates - FMA specification 59
5.4 Subjectively predicted best practice estimate 60

A.1 Studies of height premium in particular countries II

D.1 Summary statistics of the BMA model applied on the full sample XV

List of Figures

2.1 Stature determinants based on Steckel (1995) 8

3.1 Histogram of the collected height premium effects for the full sample 22
3.2 Variation of height premium effects within and across countries 23

4.1 Funnel plot of the collected effects of height premium for the full sample 29

5.1 Model inclusion of the BMA estimation 53

C.1 Variation of height premium estimates within and across studies - subsample of causal effects VIII
C.2 Histogram of the collected height premium effects for the subsample of causal effects IX
C.3 Funnel plot of the collected effects of height premium for the causal subsample X
C.4 Variation of height premium estimates within and across studies - subsample of noncausal effects XI
C.5 Histogram of the collected height premium effects for the subsample of noncausal effects XII
C.6 Funnel plot of the collected effects of height premium for the noncausal subsample XIII

D.1 Correlation matrix of heterogeneity variables XIV
D.2 Posterior model size and convergence for the BMA model applied on the full sample XV

Acronyms

BMA Bayesian Model Averaging

FAT Funnel Assymetry Test

FE Fixed Effects

FMA Frequentist Model Averaging

IV Instrumental Variable

MLE Maximum Likelihood Estimation

OLS Ordinary Least Squares

PET Precision Effect Test

PIP Posterior Inclusion Probability

PMP Posterior Model Probability

RE Random Effects

Master's Thesis Proposal

Author	Bc. Martina Juračková
Supervisor	doc. PhDr. Zuzana Havránková, Ph.D.
Proposed topic	Standing Tall Pays Off: A Meta-Analysis of Height Premium

Motivation Mincer wage equation, describing how years of schooling and potential experience impact wage, was firstly published in 1974. The benefits of schooling are indisputable regardless of years of schooling or levels of educational attainment as explained in the review by Balcar (2012). Goldsmith Veum (2002) conclude that worker's experience is rewarded with a comparable return. Since Mincer equation introduction, the researchers have been extending it with various features. Apart from education and experience, thousands of papers estimate e.g. the effects of particular skills, social and economic background, beauty, health or social capital of individuals on their earnings. These are among the most often contemplated wage level determinants. But what if we extend Mincer wage equation with physical characteristics of an individual — specifically his height? Generally, the academic literature concludes that the relationship between wage and height is positive and significant. In economics, this effect is called height premium. However, Hübler (2015) points out that not only height but also additional physical factors such as gender, age, weight, early-life nutrition or physical capacity might play a role as well.

In my thesis, I would like to focus exclusively on height premium — its true effect and what is it driven by. There are numerous papers investigating the relationship between height and wages theoretically and empirically but to my knowledge, there is no meta-analysis dealing with this topic. Given the ubiquitous nature of wage height premium results in the academic literature, I believe there is also room for publication bias detection as the results raise some suspicions.

Hypotheses

Hypothesis #1: The literature on wage height premium is affected by publication bias.

Hypothesis #2: After accounting for publication bias, the height premium is lower than commonly thought.

Hypothesis #3: Height is not the only physical factor affecting wage (gender, age, weight etc. matter as well).

Methodology At the beginning of any meta-analysis, there are studies search and data collection. Based on my keywords, I will create different combinations of search queries in Google Scholar. Relevant studies will be gathered (i.e. studies that report estimates of the effect I am interested in and also include standard errors). As there is no previous meta-analysis, whose dataset I could build upon and extend with new studies or cross-check my search with, I expect this part to take a lot of time and diligent work. Moreover, I will do snowballing to make sure all the relevant studies published within the last 3 years were indeed included. PRISMA diagram will be created.

The next step will be the data collection itself. I will collect the estimates of the effects (i.e. estimates of height premium), their standard errors and also variables that account for differences in individual studies such as age range of individuals, time range of the data, country, sample size, number of citations of the study etc. This will be followed by data inspection and data cleaning i.e. checking for outliers or suspicious values, trying to identify the source of these possible mistakes and correcting them.

Publication bias occurs due to the fact that sometimes studies with unintuitive or statistically insignificant results are not published (they are "filed away in a drawer"). For example in our case we can imagine it as the negative height premium estimates of several studies that are not published because such results are not in line with the author's expectations. But as these "bad" estimates are not reported, the simple mean of the published literature is consequently biased upwards. Publication bias will be corrected via both linear (FAT-PET, fixed effects, between effects and weighting) and non-linear techniques (selection model by Andrews & Kasy (2019) or endogenous-kink model by Bom & Rachinger (2019) - they assume that publication bias is not a linear function of standard error). Heterogeneity will be examined with the help of Bayesian and frequentist model averaging. Finally, best-practice estimate will be constructed.

Expected Contribution As remarked by Balcar (2012), the volume of papers discussing the relation between various characteristics of individuals and their wage levels is enormous. A considerable fraction of those papers deals with height premium. Nevertheless, no meta-analysis on this particular topic has been formed yet. Therefore, this thesis represents a valuable contribution as it quantitatively estimates the height premium effects, investigates and corrects publication bias and thus, reveals the true effect. On top of that, best-practice estimate is provided.

Outline

1. Introduction
2. Height premium: How is the effect estimated? Why is it important? What does previous academic literature say about it?
3. Data collection: How did I collect the data? What were the selection criteria? Basic summary statistics
4. Publication bias What is publication bias? Why could it be present? Testing for publication bias
5. Why the estimates vary: Coding for heterogeneity and identifying its sources, best-practice estimate
6. Conclusion

Core bibliography

Andrews, I., Kasy, M., 2019. Identification of and Correction for Publication Bias. American Economic Review 109(8): pp. 2766-2794

Balcar, J., 2012. Supply Side Wage Determinants: Overview of Empirical Literature. Review of economic perspectives, 12(4), 207-222.

Bom, P. R. D., Rachinger, H., 2019. â€śA Kinked Meta-Regression Model for Publication Bias Correction.â€ť Research Synthesis Methods 10(4): pp. 497-514.

Goldsmith, A. H., Veum, J. R., 2002. Wages and the Composition of Experience. Southern economic journal, 69(2), 429.

Hübler, O., 2015. Height and Wages. In J. Komlos, I. R. Kelly (Eds.), Oxford Handbooks Online. Oxford University Press

Chapter 1

Introduction

More than a decade ago while discussing the optimal taxation framework, Mankiw & Weinzierl (2010) in good spirit suggested that we should tax people based on their height. It is of course meant primarily as a joke and exaggeration to lighten the topic of income taxation a bit. At first glance, this seems like a rather odd and quirky idea. But is it?

Height can be assigned to the group of genetically predetermined features of a person. Similarly to beauty or gender, it is a physical characteristic of an individual that impacts his wage. Yet unlike education, experience, or specific skills and abilities that need to be firstly gained to be then consequently compensated for in salary, height is reflected in wage straightaway without any effort required on the individual's part. Height is not something you need to work hard for to get it. It is encoded in your DNA and you have no control over it. On top of that, it is a physical feature that is extremely difficult to alter. Utilizing this line of thought, two workers with identical attained education and set of skills but different heights should be awarded distinctly - with the wage premium accountable solely to their height difference.

Interestingly, gender-based, beauty-based, or weight-based wage discrimination is discussed plentifully (Hamermesh & Biddle 1993; Mitra 2003; d'Hombres & Brunello 2005; Ñopo 2009; Cipriani & Zago 2011; Schallenkamp *et al.* 2012; Averett *et al.* 2012; Borland & Leigh 2014; Lee 2015; Clément *et al.* 2020; Meléndez *et al.* 2021 are just a few examples). Yet, no public debate is led on height-based wage discrimination, despite the fact the empirical literature suggests the relation between wage and height is significant and positive. This brings us back to the proposition of Mankiw & Weinzierl - should we tax people based on their height? And conversely, should a person be paid extra for something that is not a result of hard work but rather a feature inherited from their parents? Though the idea is amusing, Mankiw & Weinzierl (2010) point out that it stems from the standard approach to optimal taxation policy.

As this thesis is focused on the meta-analysis of height premium in wages, the optimal taxation framework will not be discussed. Therefore, answers to the above-mentioned questions will not be provided here. However, one of the goals of this paper is to bring height discrimination in wages into public focus and possibly initiate a political debate on this topic. Also, to the author's best knowledge, a meta-analysis focused on height-premium does not exist yet. Therefore, the contribution of this paper is manifold.

The objective of this thesis is to perform a quantitative meta-analysis of height premium effects that capture the relationship between height and wages. As mentioned in Havránek *et al.* (2020), meta-analysis can be described as *the systematic review and quantitative synthesis of empirical economic evidence on a given hypothesis, phenomenon, or effect*. Via implementing the meta-analysis regressions, we will test for the presence of publication bias in the height premium literature. On top of that, with the help of current model averaging techniques, we will address the heterogeneity of the primary study's design and determine factors that explain the variability among the height premium estimates.

To be able to carry out the analysis, we collected 1084 effects of the impact of height on wages. Our dataset contains both causal and noncausal height premium associations. The results of the applied tests and models suggest that there is a positive publication bias in the literature dealing with wage returns to height, suggesting an under-representation of negative effects. After accounting for the bias, the mean beyond bias effect is negligible for the noncausal effects, however, remains pronounced for the causal associations of height premium. Additionally, we observe that the positive publication bias remains present even if we control for additional variables specifying different study characteristics. Including tens of variables into the regression yields uncertainty which can be handled by model averaging techniques. In our case we apply Bayesian Model averaging (BMA), Frequentist model averaging (FMA) and Frequency check. The heterogeneity analysis reveals that using dataset from Africa or America (excluding the USA) is associated with systematic reporting of positive effects, while the opposite is true if the researchers use longitudinal data, wage (dependent variable) of monthly frequency, restrict the dataset based on gender, or explicitly include gender control into their model.

The thesis is structured as follows: Chapter 2 discusses in detail height determinants and the role of height in the life of individuals. Moreover, it provides an overview of height premium literature and also lists some issues researchers encounter when estimating the height-wage relationship. In Chapter 3 we elaborate on the data collection procedure and data adjustments and provide basic summary statistics of the dataset used. Chapter 4 deals with publication bias. In Chapter 5, the results of the heterogeneity analysis are presented. Finally, Chapter 6 concludes.

Chapter 2

On height and its link to income

The purpose of this particular chapter is to introduce the reader to the topic of height premium. Firstly, in Section 2.1 we will be dealing with height only - why is it important, what qualities do we expect individuals to have based solely on their height and we will conclude with how is stature determined and measured. Secondly, Section 2.2 will provide a general overview of height premium (i.e. link of height to income) in the form of a literature review but we will also add a few words on how the economists determine it. Thirdly, there are several challenges associated with height premium and its measurement, as will be described in Section 2.3.

2.1 Height

2.1.1 Why is height important

According to Deaton (2007), on average people of greater stature earn more, score better in case of cognitive tests and have longer life expectancy. Also, height is strongly linked to labour market success (Case & Paxson 2008b) and social class status (Steckel 1995). There are several theories that try to address the issue of why short people may receive a different treatment compared to taller people, namely the leading theory in social psychology, the concept of evolutionary selection, height acting as a signal of good health and genetic robustness to illness and deprivation, theory emphasizing self-esteem, source of height premium originating in birth or general employer discrimination, occupation-specific discrimination, customer discrimination and others (Harper 2000; Persico *et al.* 2004). In the following paragraphs, some of those theories will be introduced in more detail.

Firstly, we will begin with evolutionary theory, which is in fact quite intuitive - in nature, animals are forced to make fight-or-flight decisions and following the rule ”the bigger, the more dangerous”, height simply commands respect and emits power. Although humans would perhaps like to believe otherwise, they have similar notions embedded in their subconscious minds. The proof of which is provided by Persico *et al.* (2004) who on an example of 13 US presidential elections demonstrate that firstly, in 10 cases the taller candidate has won and secondly, the height of presidents is generally higher than the estimated average height of the population. Therefore, height has a vast impact on how individuals regard themselves as well as how they are regarded by others both of which then transmit through several channels into their career success. Self-esteem and social esteem most definitely impact a person’s job performance and are key to professional success. Judge & Cable (2004) attribute such an observation to human perceptual bias: *People expect a positive relationship between an entity’s size and its value or status. Thus, studies have shown that people perceive more valuable things as larger than less valuable things; for example, coins are perceived as larger than cardboard disks of identical diameter, and jars filled with candy are judged to be heavier than jars of equal weight filled with sand. This perceptual bias also extends to judgments about people’s height and the extent to which they are esteemed by others.* The abovementioned thus insinuates that in some regards, the life of taller people may be easier as is indeed supported by empirical evidence. We have already established that height is a desirable asset (Roberts & Herman 2022). People of greater stature are perceived as more attractive (Harrison & Saeed 1977; Freedman 1979), more persuasive (Young & French 1996) and more likely to be prescribed as leaders (Higham & Carment 1992). On top of that, the literature research may also serve as support material for employers when hiring new employees since customers view salespeople that are taller than average as more persuasive, competent, and impressive (Kurtz 1969) while e.g. shorter police officers were a priori presumed to cause more disciplinary problems, display poorer morale, or receive more complaints (Lester & Sheehan 1980). On the other hand, such conduct might lead to employer discrimination resulting in employers preferring employees with particular physical characteristics.

Next, we will focus on leadership theory. Underdal (1994) defines leadership as *an asymmetrical relationship of influence in which one actor guides or directs the behaviour of others toward a certain goal over a certain period of time.* Subsequently, the leadership theory captures the personality traits or qualities that predetermine an individual to become a leader. As already remarked, taller people are naturally elected as leaders (Stogdill 1948) because their height signals capability, earns respect and high esteem, and serves as an indicator of talent.

The last line of thought that will be described introduces height as a pointer of good health and robustness to illness. As mentioned by Perkins *et al.* (2016), linkages between adult height and health exist both within and across generations. Moreover, the authors suggest that height should be commonly measured as it is *a useful marker of variation in cumulative net nutrition, biological deprivation, and standard of living between and within populations.* Furthermore, Deaton (2007) adds that taller people generally also live longer which might be due to economic and health reasons since either insufficient nutrition or disease in early childhood (more likely to be experienced by children from lower social class) consequently impact growth and overall health (Gunnell *et al.* 2001). On that note, Case & Paxson (2008a) argue that a healthy life environment (measured by height) transcribes from adulthood to older age. Taller individuals above the age of 50 experience significantly fewer difficulties with regular daily tasks, have better physical and mental health, and overall better cognitive functions. On the other hand, as the studies focused on the connection between height and longevity date back almost 100 years, the results have been mixed and e.g. Samaras *et al.* (2003) contradicts Deaton (2007) with the claim that shorter people have on average longer lifespans.

2.1.2 What determines height

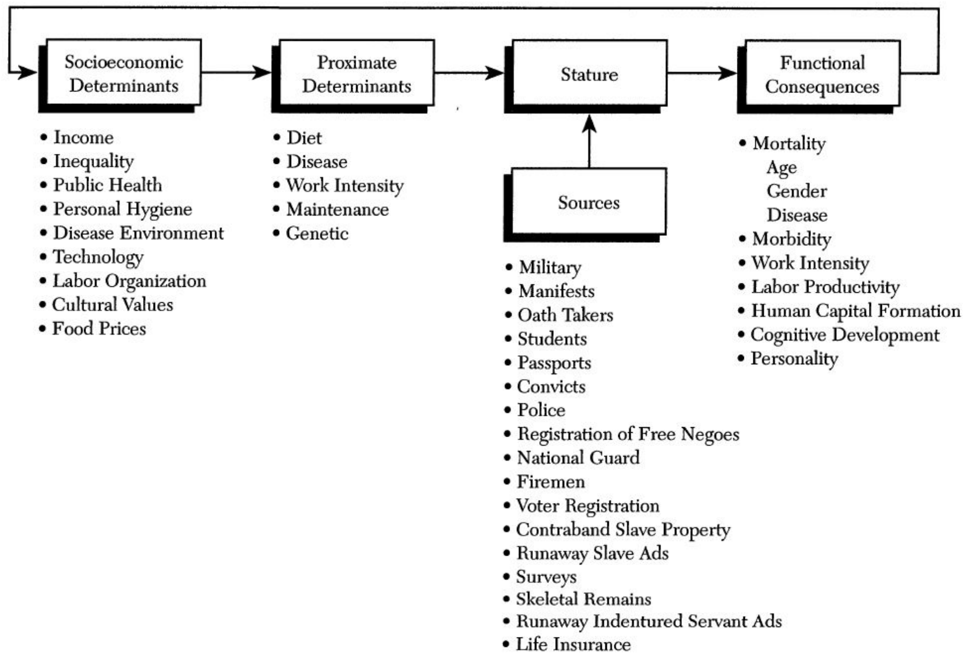
As has been already denoted, the height of a person is mostly a matter of genetics and inheritance but hormonal and psychological differences play a role as well. McEvoy & Visscher (2009) explain that so far 50 genes and regions of the genome have been identified as important determinants of height. Nevertheless, hundreds or even thousands of genes yet remain to be identified. Still, to a certain extent, height can be affected by external factors - Tanner & Tanner (1990) list the following: diet and nutrition, disease exposure, social class, psychological stress, climate, or urbanization. From these, net nutrition is probably the key one (this is also acknowledged by Thomas & Strauss (1997) who utilize both protein and calorie intakes as measures of health as well as wage determinants in urban Brazil). Protein, calcium, vitamin D, and zinc are considered as vital nutrients directly affecting individuals' growth. In case children's diet is lacking a sufficient amount of calories or key nutrients, their growth rate decelerates. As long as the deficient diet is not of a long-term nature, the human body can rebound and catch up. Nonetheless, with the impact of the prolonged growth period into the early 20s (Tanner *et al.* 1978; Zehetmayer 2013).

The ratio of genetic versus environmental determinants is often reported as 80% and 20%, respectively (Zehetmayer 2013). Case & Paxson (2008b) confirm the aforementioned proportion is particularly true for Western countries and while only 20% of variation in body stature can be attributed to environmental aspects, curiously enough those are the factors that are actually responsible for most of the differences in average heights across populations. Moreover, Lång & Nystedt (2018) imply that from the environmental aspects, prenatal conditions do not play a role either. Instead, approximately 90% and 80% of height premium for males and females, respectively, originate in environmental conditions in childhood and adolescence.

Studies that exploit datasets dating back more than a century reveal that the growth patterns used to be quite different in the past. These days we do not suffer from lack of food, we are not forced to carry out extremely physically demanding work and our health conditions are much improved. All of this leads to the fact that objectively we are better off than our ancestors e.g. two centuries ago and consequently, adolescents nowadays experience growth spurts at earlier age. Boys and girls in Europe go through peak growth at ages 18 and 16 on average, respectively, whereas in the 19th century, such growth pace was typically undergone 8-10 years later (i.e. at age 26).

Height is for the most part predetermined in DNA and thus quite difficult to alter, even though as mentioned above, some external factors (especially nutrition) can affect it to a certain extent. Those attributes are nicely distinguished in Steckel (1995) - see Figure 2.1 - where stature is perceived primarily as a function of proximate determinants and access to resources. According to Schultz (2002), a child's nutrition and health conditions are especially important in the following 3 life stages - fetal development during pregnancy, the first four years of childhood, and adolescent physical growth spurt. In terms of numbers, Case & Paxson (2010) add that after a period of swift growth at ages 0 to 3, children grow at a constant pace of approximately 6 cm per year until adolescence is reached when a growth spurt of 10 cm per year is typical. The duration and timing of the adolescent growth spurt then determine the height in adulthood.

Figure 2.1: Stature determinants based on Steckel (1995)



However unfair it might seem height, as a preordained characteristic, impacts the lives of individuals in many aspects. Be it their health conditions (stunted people are at greater risk of chronic diseases or premature death as discovered by Fogel (1994)), leadership abilities (Lindqvist 2012), happiness and well-being (Carrieri & De Paola 2012; Sohn 2016; Lee & Zhao 2017; Denny 2017), social status (e.g. Stulp *et al.* (2015) document not only a positive relationship between height and social status but also a positive link of height to interpersonal dominance) or e.g. risk-taking preferences (Rieger 2015).

2.1.3 How is height measured

Height used by researchers in their analyses can be of two types - either self-measured (in which case there are certain concerns - see Section 2.3) or reported by physicians. Depending on the country of origin, the units of measurement might differ (will be addressed in Section 2.2.2). Typically, height is incorporated into the models in its continuous form, however, sometimes scientists like to transform it into height deciles, quantiles, other height intervals or height dummies usually in the form of short/medium/tall or below median height/above median height. On top of that, the authors most of the time expect height to be homogenous and exogenous but such an approach also poses a few questions (see Section 2.3).

2.2 Link height-income

What affects an individual's wage is arguably one of the most (if not the most) examined topics in economics. In the academic literature, there is extensive research acknowledging education (a form of human capital) as a positive significant contributor to one's labour market return. In fact, the variety of datasets, methodologies and diversity of estimated results in the area of returns to education even led the authors to conduct several meta-analyses on this topic (Groot & Van Den Brink 2000; Van der Sluis *et al.* 2005; Liu *et al.* 2013; Churchill & Mishra 2018; Ma & Iwasaki 2021). Nevertheless, as pointed out by Thomas & Strauss (1997), human capital has more layers than just education. To be more precise, these authors pinpoint also the health condition of an individual reflected in height, body mass index (BMI), calorie intake, and protein intake. From the mentioned components of human capital, this paper will be focused on height only. The relationship between height and income is also recognized as height premium and will be referred to as such throughout the whole paper.

The importance of stature and how it impacts the lives of individuals have been to a large extent already described above. Section 2.1 mentions several channels through which height operates and affects wages. Here we will only add that Harper (2000) suggests the link between stature (and not only stature but also attractiveness or body mass) and income could originate in employer discrimination, occupation-specific discrimination, or productivity differences arising from customer discrimination. Moreover, e.g. Persico *et al.* (2004) identifies that the height premium is not necessarily present due to height as such but because of some characteristics correlated with height and therefore, one should bear in mind the potential problem of causality and correlation.

2.2.1 Literature review

Height premium

Generally, the empirical literature confirms the notion that taller men and women earn more (Loh 1993; Thomas & Strauss 1997; Harper 2000; Schultz 2002; Deaton 2007; Case *et al.* 2009; Böckerman *et al.* 2010; Kedir *et al.* 2013; Anderson 2018; Gu & Ji 2019). Moreover, this sentiment can be identified among many countries e.g. USA, Russia, Sweden, UK, Germany, China, India, Indonesia, Ethiopia, Australia, Colombia, Brazil, and others (see Appendix A). And as can be observed, the height premium is not limited only to developing countries, it can be identified in developed countries as well in which case greater stature is often associated with higher social and economic class (Steckel 1995).

While commonly the relationship between height and wage is assumed to be linear, Hübler (2009) identifies it is rather curvilinear (i.e. inverted U-shaped); in the private sector men who are taller than the average (but not the tallest) receive on average the highest wages, while in the public sector, the recipients of such advantages are women shorter than average. Heineck (2008) also finds a U-shaped relationship between height and wage in the case of females employed in administrative and secretarial occupations.

Most of the time, the researchers estimate the height premium to be within the 1% and 10% range for a 10 cm increase in an individual's height (Hübler 2016). However, Gu & Ji (2019) point out that differences in wages are attributable to variation in characteristics of individuals and that *after accounting for the social network, human capital and other endowment characteristics* the estimated impacts of height on wages are smaller and even statistically insignificant. Similarly, when individuals' occupation is incorporated into the regression, then the height premium in wages dissolves (Heineck 2008). Nevertheless, the premium is not a matter of just recent history, it could be observed already more than 100 years ago in the Victorian and Edwardian periods as revealed by Anderson (2018).

Bargain & Zeidan (2017) offer that height premium is translated into wages through the human capital channel - taller workers are likely to achieve higher levels of education and consequently, workers that are more educated are drawn to better-paid jobs. Case & Paxson (2008b) observe that the link between height and earnings can be found through cognitive abilities - taller people have better cognitive abilities and this fact is rewarded. This is supported by Anderson (2018). Apart from cognitive skills, some authors account the height premium also to social skills (Persico *et al.* 2004) or employer-based discrimination and occupational sorting (Cinnirella & Winter 2009). But Lundborg *et al.* (2009) remark that when they controlled for the physical capacity of an individual, they managed to lower the height premium by 80%. Additionally, controlling for physical capacity, cognitive skills and non-cognitive skills meant the whole height premium was covered.

The majority of the researchers do not distinguish between height premiums for different occupations but Harper (2000) finds that in occupations that require direct personal contact with customers are taller workers in particular likely to receive a wage bonus. Moreover, Mitra (2001) suggests that while men remain unaffected, taller women in blue-collar jobs earn more. Böckerman *et al.* (2010) form 4 work strain categories (sedentary work, light manufacturing work, heavy manufacturing work, physically very demanding work) and although there are not discovered any significant differences between height premiums of the respective groups, they also add that the tallest males are typically representatives of sedentary occupations, whereas the shortest men perform physically demanding jobs.

Employing the OLS estimation methods, Böckerman *et al.* (2017b) confirm that taller workers have higher wages. But at the same time, they point out that the OLS estimate is possibly biased. When they use the genetic score of an individual (based on 180 single nucleotide polymorphisms connected with height) as an instrumental variable under Mendelian randomization, then the estimated impact of height on wage is lower and more importantly insignificant. Clearly, the estimation method matters. Wang & Shen (2017) as well come to the conclusion that compared to OLS, IV is a more valid estimation method, while Gao & Smyth (2010) account for factors impacting height during childhood via instrumental variables and find that under TSLS regression, the returns to height are significantly larger compared to OLS estimates.

Intriguingly, height can be looked upon also from a different perspective - as an income inequality indicator. Choi (2020) argues that height inequality can be used as an alternative to more traditional income inequality measures (e.g. the Gini coefficient) with the finding that height inequality of individuals born in Korea between 1890 and 1919 is related to income inequality.

As noted by Price (2013), the positive link between stature and wage can also have a biological foundation. That is because *end-of-the-workday fatigue, or lack of energy, varies inversely with stature.* Therefore compared to shorter workers, the body of taller workers produces more energy to accomplish the required work tasks, resulting in less exhausted and more productive employees.

An interesting observation has been made by Sohn (2015b) - *a 1 cm reduction in husband's height relative to his wife's height costs him approximately 3% of his earnings.* Also, Wang (2015) reveals that in both the US and the UK immigrants have higher wage returns to height than domestic workers.

Other physical features

As has been already mentioned, height is not the only physical feature the classical Mincer equation has been augmented with over the years. For the most part, researchers focus on 4 other features - gender, ethnicity, beauty, and weight. For a detailed review, see Appendix B.

Overall, a person's height is clearly an important feature, as apart from wage, it is also associated e.g. with their level of happiness (Carrieri & De Paola 2012; Sohn 2016), well-being (Lee & Zhao 2017; Denny 2017), life satisfaction (Salahodjaev & Ibragimova 2020), career success and social esteem (Judge & Cable 2004), health outcomes of elderly (Huang *et al.* 2013), leadership abilities (Lindqvist 2012; Blaker *et al.* 2013), criminal activity (Bodenhorn *et al.* 2012), cognitive skills (Heineck 2009), perceived physical strength (Undurraga *et al.* 2012) or risk-taking preferences (Rieger 2015).

2.2.2 Computation

Commonly in the majority of the studies, the impact of height on wage is estimated via utilizing the classical Mincer equation augmented for height. As a dependent variable, we can most often observe the wage, income or wealth of an individual (usually in log form), while among the independent variables height is included, meaning the equation as a whole is in a so-called semi-elasticity form. This can be formally written as

$$log\, wage_i = \beta X_i + \epsilon_i \tag{2.1}$$

where $log\, wage_i$ is logarithm of hourly wage rate of individual i, X_i is vector of explanatory variables, β is the coefficient vector and finally ϵ_i is the error term.

As the literature on height premium used in the meta-analysis comes from various countries with different metric systems, height can be reported e.g. in centimetres, metres, or feet and inches. Therefore, transformations will be required for the height premium effects to be comparable among all the studies.

2.3 Issues and biases associated with height-income relationship measurement

As it happens in social sciences and especially in economics, empirical research is not without its challenges. When studying height and subsequently its link to income, the authors face numerous difficulties, varying from (at least at first sight) simple height measurement to methodological or modelling problems. Below are listed a few examples of obstacles the researchers come across in the context of height premium the most often.

Measurement error

Due to time efficiency reasons, the authors mostly decide to utilize already conducted and available longitudinal cohort studies. As those are of good quality and provide a great deal of information, it is not usually necessary to perform interview surveys on their own. On the other hand, such an approach makes the researchers dependent on the information reported and specifically in the case of height, there is a threat of self-reported bias. That is because, in the majority of cohort studies, the height is self-measured and self-reported by the respondents. The authors are aware of this shortcoming (e.g. Persico *et al.* (2004), Judge & Cable (2004), Case & Paxson (2008b), Cortez (2014) recognize this flaw), nevertheless, there is not much that can be done, save perhaps as Schultz (2002) or Gao & Smyth (2010) mention, using instrumental variables. To at least examine whether the bias might occur in the data, Persico *et al.* (2004) compare the distribution of height from the data with the height distribution of adults from a national survey conducted under the supervision of general practitioners, assuming the national survey should serve as a reliable reference group of population height. However, Boström & Diderichsen (1997) argue that in the case of self-reporting, weight suffers from potential bias much more strongly than height.

On that note, for the most part, if height measurement is taken from an official medical examination administered either by an anthropometrist or physician, then it is less prone to measurement error and overall the results display lower attenuation bias as confirmed by Case & Paxson (2008b).

Sample selection bias

Harper (2000) for example uses data from a British NCDS longitudinal survey of individuals born in the week 3-9 March 1958. He uses only a sample of employees, excluding self-employed, unemployed or not economically active. Such conduct might pose a question of whether there exists a bias stemming from employment selection. However, Harper (2000) showcases that a similar procedure was undertaken by Connolly *et al.* (1992) and Harper & Haq (1997) who detect no evidence of estimation bias associated with systematic data loss or drop-outs. Moreover, Harper (2000) presents an explicit test for sample selection bias which reveals that the potential selection bias is in this case almost negligible.

Confounding effects of gender or race

When estimating wage determinants Kuhn & Weinberger (2005) present an interesting way of dealing with the confounding effects of race and gender discrimination. Though they primarily focus on the impact of leadership in high school on wages in adulthood, height is also included in the model as one of the controls. They combine 3 datasets - Project TALENT, National Longitudinal Study of the Class of 1972 and High School and Beyond cohort of 1982 sophomores - from which they choose a subsample of *white men only* and thus clear the analysis of the effects of race or gender which are likely confounding. The results show that individuals who acted as e.g. team captain or club president in high school earn significantly more about 10 years later and intriguingly the size of the marginal effect does not vary with the level of attained education. An identical approach was taken by Persico *et al.* (2004): *To avoid confounding the effects of race, gender, and height discrimination, we focus our attention primarily on white men. In Britain this implies excluding the small number of native-born nonwhites ... in the United States, we focus on the 2 063 white, non-Hispanic men from whom there exist both adequate height data and other information.* The authors identify a wage premium for taller workers. At the same time, they also hint at a link between teen height and adulthood wage. Taller teenagers are more likely to participate in extracurricular activities such as athletics, youth groups, student government, hobby clubs or any school clubs in general where they acquire a special set of skills or human capital that distinguishes them from their peers and as Persico *et al.* (2004) remark, establishes the root of the wage differences in adulthood.

Endogeneity of height

When estimating the effect of height on labour market outcomes, be it productivity, worker's earnings, or household income, the researchers generally assume that adult height is homogeneous, measured without error, and exogenous with respect to an individual's wage. However, Schultz (2002) challenges this view by using various sets of instrumental variables reflecting regional conditions, prices, parent education, ethnicity, race, or even their interactions. The results not only indicate that height is indeed endogenous with respect to wage but also suggest that the OLS estimation method could suffer from bias as the OLS estimates are manyfold smaller than the instrumental variable estimates. The bias may stem from the fact that the majority of the observed variation in height is attributable to genetic factors that are unlike health-related human capital factors linked to productivity rather weakly.

If the scientists encounter endogenous height, they typically take the route of instrumental variables approach with e.g. food prices, genetic score, sibling's height, parent education, regional conditions, average number of health institutions, or ethnicity used as instruments. Sometimes they apply fixed effects to mitigate the time-invariant factors. However, Rietveld *et al.* (2015) point out certain limitations to this approach as height is time-invariant to a large extent as well (Frieze *et al.* 1990). One could argue that the height of an individual evolves as the stature of an infant or teenager is not identical to the stature of an adult. Which is a reasonable claim. Nevertheless, the academic literature studying height-income relationships utilizes almost exclusively adulthood height - thus, the time-invariance assumption.

Overall, following the notion that height in adulthood is essentially fixed at least until the individual's fifties and contemporary lifestyle has little impact on it (Rashad 2008), generally in the majority of cases the authors decide not to treat adult height as endogenous.

Causality

As regards causality, there are two main issues related to it - its direction and causality vs. correlation. The direction of causality may be of some concern since while better health captured by greater height leads to enhanced productivity, the resulting higher income could be spent on health improvement which is reminded in Thomas & Strauss (1997). Thus, simultaneity bias might contaminate the estimated effects. What is more, one should be also cautious when inspecting the income benefits of height as correlation and causation ought not to be used interchangeably. As is argued by Persico *et al.* (2004), the height premium of taller workers may have roots in particular characteristics correlated with height and not necessarily in height per se.

Chapter 3

Data

The aim of this paper is to conduct a meta-analysis of the effect of height on wages, namely to uncover possible publication bias and estimate the true effect of height premium. Modern meta-analysis methods will be applied to a dataset containing a total of 1084 estimates collected from 67 studies (127 causal effects and 957 noncausal effects). This chapter serves as an insight into the data collection procedure, describes data adjustments realized on the data gathered and finally, provides basic summary statistics.

3.1 Data collection

To the author's best knowledge, no meta-analysis dealing with the topic of height premium has been published yet, save perhaps Judge & Cable (2004). Nevertheless, their meta-analysis is focused on the link between height and social esteem, leadership emergence and job performance; in the second half of the paper they also mention 4 studies that highlight the relationship between physical height and income but this reference serves more as a suggestion for future research to draw attention to this, from the point of meta-analysis, rather unexplored research area.

Therefore, our data collection starts from the ground up with a studies search on Google Scholar. This is done via inputting specific keywords as search queries into the Google Scholar engine - we decided to use *height premium and wage*. The Google Scholar search returns thousands of studies. Following the general practice, the abstract of the first 500 of them was inspected. Those that could potentially contain empirical estimates of interest were categorized and saved for further analysis. The search started in September 2022 and was terminated in December 2022. In order to be included in our dataset, the study needed to meet the following criteria:

1. The study examines the impact of physical height on an individual's income, wage, earnings, productivity, or wealth.
2. The study reports the estimated coefficients together with the measures of precision (typically standard errors, t-statistics, or p-values).
3. Both height and income are examined at the level of an *individual*, so that the effects are comparable (meaning e.g. studies estimating the link between the average height of household members and *household* income are discarded).
4. The study uses quantitative methods.
5. The study is written (at least partially) in English so that information essential for the analysis can be extracted from it.

Following the abovementioned procedure, we identified 124 papers as suitable based on their topic and abstract. However, the list was narrowed down to the final number of 67 studies, as several of them were found lacking in some way or another. Typically for the following 3 reasons. Firstly, instead of using height in its continuous form, a significant portion of studies transforms it into height intervals or height dummies (see e.g. Loh 1993; Harper 2000; Kuhn & Weinberger 2005; Dinda *et al.* 2006; Cinnirella & Winter 2009; Loureiro *et al.* 2010). However, the height dummy coefficients are hardly comparable with the estimates of the impact of continuous height. Hence, papers of such kind needed to be excluded. Secondly, elaborating on the topic of the effects' comparability even more, we decided to include only studies applying the semi-elasticity form of the Mincer equation (meaning the measure of income serving as a dependent variable is in log form, while height remains continuous without any transformation). Thirdly, a few studies have the relationship reversed (see e.g. Steckel 1995; Komlos & Lauderdale 2007; Ranasinghe *et al.* 2011) - they examine the effect of *income on height*, whereas we target at the effect of *height on income*. Overall, it should be noted that even though these papers were not included in the dataset used for the meta-analysis in its empirical sense, they were a valuable source of information nonetheless and several are cited throughout this thesis.

As already mentioned, the literature on wage determinants and height premium, in particular, is very rich. Owing to this fact, we could afford to impose an additional restriction regarding the estimate's uncertainty measures without considerably affecting the sample size of our dataset. It was decided to collect only estimates accompanied by an explicit measure of precision, i.e. firmly stated standard error, t-statistic or p-value, and discard estimates reported only with signs corresponding to certain significance levels. It is possible to approximate the t-statistics of the graphically denoted significance levels but such a conduct introduces an error to our data and since we did not suffer from a lack of estimates, we decided to forego this procedure entirely.

To conclude, we select 67 studies from which the estimates of the effects of height on income in the form of causal effects, noncausal effects, or both together with their measures of precision are collected. We do not discriminate based on publication prestige - along with research articles published in peer-reviewed journals, we also incorporate several working papers, discussion papers, and conference articles. However, we do account for this characteristic as well as for others (e.g. estimation methods, type of data, publication characteristics, design of the analysis, or controls for individual's characteristics) during the data collection procedure in order to address the study heterogeneity. The topic of heterogeneity will be detailed in Chapter 5 and the specific characteristics of the study we control for in particular are discussed in Section 5.1. The list of studies incorporated into the meta-analysis is given in Table 3.1.

3.2 Data adjustments

Since the academic literature in the area of height premium is generous and we were able to impose several restriction criteria in order to ensure comparability of the individual effects, while at the same time maintaining the robustness of the dataset, we were not required to perform many adjustments.

First of all, as countries around the world utilize various metric systems, height can be expressed in e.g. in centimetres, metres, feet, or inches. This implies that a simple linear transformation is in order for the height premium effects to correspond. Accordingly, estimates and their respective standard errors originating from models with height measurement units different than centimetres were re-scaled in order to be converted into the selected common base metric - centimetres - and therefore directly comparable.

Table 3.1: List of studies included in the meta-analysis

Anderson (2018)	Hill (2004)	Peng *et al.* (2020)
Asadullah & Xiao (2020)	Hitsch *et al.* (2010)	Persico *et al.* (2004)
Baker & Cornelson (2019)	Hübler (2009)	Reddy (2014)
Bargain & Zeidan (2017)	Chen & Pastore (2021)	**Ribero (2000)**
Behrman & Rosenzweig (2001)*	Ibragimova & Salahodjaev (2020)	Rietveld *et al.* (2014)
Bleakley *et al.* (2014)	Johansson *et al.* (2009)	Rietveld *et al.* (2015)
Böckerman & Vainiomäki (2013)*	Johnston (2010)	Rooth (2011)
Böckerman *et al.* (2010)	Kanazawa & Still (2018)	Sargent & Blanchflower (1994)
Böckerman *et al.* (2017a)	**Kedir (2008)***	Schick & Steckel (2015)
Böckerman *et al.* (2017b)*	**Kedir (2009)**	**Schultz (2002)***
Bonilla *et al.* (2019)	Khasnobis & Dinda (2017)	Sohn (2015a)
Bossavie *et al.* (2017)*	Kim & Han (2017)	Sohn (2015b)
Case & Paxson (2008b)	Kortt & Leigh (2010)	Tao (2014)
Case & Paxson (2010)	Kropfhäußer (2016)	Vogl (2014)
Case *et al.* (2009)	Lång & Nystedt (2018)	Wang (2015)
Chen *et al.* (2019)*	Lee (2014)	**Wang & Shen (2017)**
Elu & Price (2013)*	Lee & Zhao (2017)	**Wang *et al.* (2020)***
Eschker *et al.* (2004)	Lindqvist (2012)	Yamamura *et al.* (2015)
Gao & Smyth (2010)*	Lundborg *et al.* (2009)	Yang *et al.* (2018)
Groothuis & Hill (2013)	Lundborg *et al.* (2014)	**Yimer & Fantaw (2011)**
Heineck (2005)*	Mitra (2001)	Zheng (2022)
Heineck (2008)	Oreffice & Quintana-Domeque (2016)	
Hersch (2008)	Park & Lee (2010)	

Note: The table presents a list of studies from which the estimates of height premium were collected. The studies that are in bold contain only causal effects, studies in bold with asterisks reveal both causal and noncausal effects. The rest of the studies report only noncausal effects.

Next, instead of standard errors, several studies disclose t-statistics (see e.g. Foster & Rosenzweig 1994; Behrman & Rosenzweig 2001; Schultz 2002; Johansson *et al.* 2009; Yamamura *et al.* 2015; Asadullah & Xiao 2020) or signs for significance levels in terms of asterisks (see e.g. Rashad 2008; Böckerman *et al.* 2017b; Tovar-García 2021) as the measure of the estimate's uncertainty. We, therefore, transform t-statistics into standard errors. As regards the visual depiction of the coefficient's significance, the approximation of the t-values from the number of asterisks or other graphical notation is difficult, especially at the 1% significance level. Hence, we opt to defer from this practice in order to avoid introducing errors to our dataset.

Additionally, in the case of Bleakley *et al.* (2014) and Ibragimova & Salahodjaev (2020) we replace the reported standard errors equal to 0 with 0.00001. As we collected only semi-elasticity estimates, the interpretation of the effect is clear and after the adjustments mentioned above, we encounter no more issues with econometric specification or measurement units and no other modification (for example transformation into partial correlation coefficients) is necessary. Lastly, since the dataset contains several outliers (see Figure 3.1 or 4.1), we decided to winsorize[1] the estimates and their standard errors at the 1% level.

3.3 Basic summary statistics

The purpose of this subsection is to introduce the basic summary statistics of the collected estimates before we proceed further to the publication bias and heterogeneity analysis. As stated earlier, the utilization of the selection criteria left us with 67 studies suitable for data extraction. From those, a total of 1084 estimates of the relationship between height and income were retrieved (for the detailed list of papers see Table 3.1). Within our sample of estimates, we distinguish 127 instances of height premium causal effects and 957 cases of noncausal links of height to income. The study search was realized from September 2022 to December 2022. The last study was added in December 2022, then the literature search was terminated. The search queries were inputted in English, but papers written in other languages were not discarded as long as they included valuable information in English that could be extracted

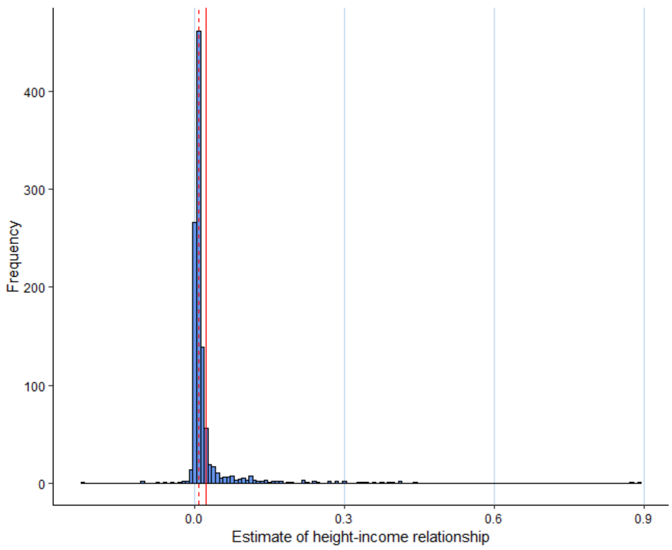
[1]Winsorization belongs to one of the common methods of dealing with outliers. It is based on replacing the outlier values with less extreme ones. In the case of the 1% level winsorization, the data points below the 1st percentile and above the 99th percentile are altered to the value of the 1st percentile and 99th percentile, respectively. Another method is e.g. trimming which completely drops the outlier from the dataset.

(e.g. Park & Lee 2010). The oldest study we use was published by Sargent & Blanchflower (1994) almost three decades ago, and the most recent one by Zheng (2022).

Proceeding from the individual studies to the estimates themselves, Figure 3.1, Figure 3.2 and Table 3.2 should equip the reader with an elementary knowledge of the structure of the dataset we gathered.

Figure 3.1 shows the distribution of the effects describing the relation of height and wages published by the researchers. The histogram is unimodal and right-skewed, meaning the outlier values are positive rather than negative. The majority of the estimates are concentrated around zero. The solid line indicates the mean of the sample. The median is signified by the dashed line.

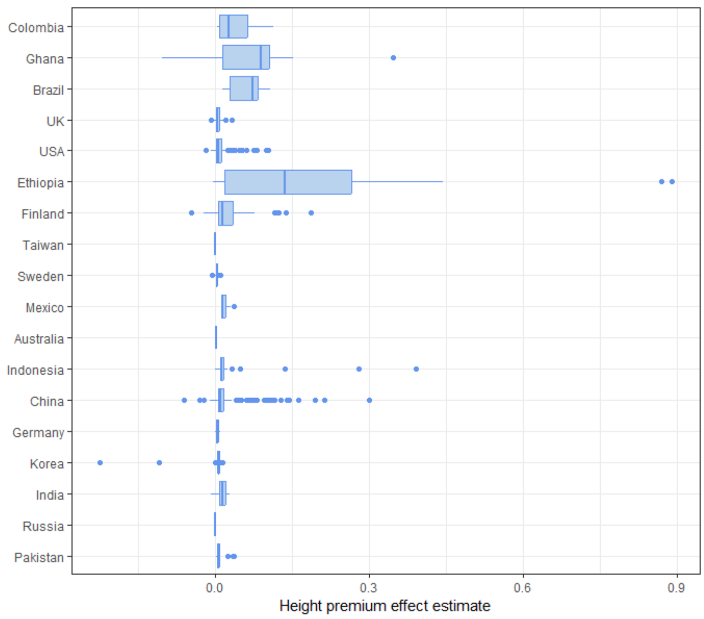
Figure 3.1: Histogram of the collected height premium effects for the full sample



Note: The figure displays the distribution of the height premium estimates reported by the primary studies. The estimated effects are on the horizontal axis and their frequency is on the vertical axis. The solid line represents the mean, the dashed line represents the median. Unwinsorized data are used.

The estimates of height premium aggregated by countries can be found in Figure 3.2. The boxes represent the interquartile range (i.e. the spread of the data between the 25th and 75th percentile, sometimes also called the middle half of the data) and the solid line inside the box stands for the median. The lower and upper whiskers illustrate the lowest and highest 25% of the data, respectively. The outliers are portrayed as dots outside the whiskers. Therefore, the boxplot depicted in Figure 3.2 demonstrates that the estimates collected from academic literature exhibit only a mild variability both within and across countries, especially in the case of the developed ones. However, the opposite is true for Brazil, Colombia, Ethiopia, or Ghana (i.e. developing countries[2]). For the boxplots of the subsample of causal and noncausal effects capturing the impacts of height on income, see Figure C.1 and Figure C.4 in Appendix C.

Figure 3.2: Variation of height premium effects within and across countries



Note: The figure shows a boxplot of the collected estimates on the effect of height on income both within and across various countries. Unwinsorized data are used.

Finally, Table 3.2 provides a statistical overview of the full sample. When inspecting the histograms and boxplots created for causal and noncausal effects alone (see Appendix C), we learned that the two subsamples exhibit noticeable differences. The histogram of causal effects appears to be more heavy-tailed. Also judging from the comparison of Figure C.1 and Figure C.4, the causal effects display more variability both across and within studies in contrast to the noncausal effects subsample. On the grounds of this finding, we would like to conduct separate meta-analyses for the respective subsamples and therefore, we complement Table 3.2 with descriptive statistics on causal and noncausal shares of the sample as well. We discerned causation and correlation already

[2]Brazil, Colombia, Ethiopia, and Ghana can be found on the IMF list of developing countries.

during the primary studies data collection procedure. In the further stages of the analysis, the distinction is ensured by explicitly encoding a dummy variable *Endogeneity* equal to 1 if the primary study accounts for the endogeneity of height and thus, reports the causal effects, or equal to 0 if the estimate is classified as noncausal effect (see Table 5.1).

The subsample of causal effects of height premium is characterized by a median of 0.078 and a simple mean equal to 0.115, while the noncausal effects of height premium can be described with a median and simple mean equal to 0.007 and 0.009, respectively. However, in the setting of meta-analysis, a simple (unweighted) mean should be regarded with reservations. That is because simple mean does not account for the number of effects a particular study reports. On the contrary, this averaging technique actually assigns more weight to studies with a higher number of estimates published. Ergo, weighted mean (weighted by the inverse of the number of estimates per study) should serve as a more informative and reliable tool. The denoted disproportions are easily visible when comparing the weighted and unweighted means and are especially pronounced for the subsample of causal effects[3].

[3]For the noncausal effects subsample, we have more than 10 studies that report only 1 or 2 estimates, while e.g. Wang (2015) discloses 65 height premium effects. In the case of the causal effects, over 25% of studies in the subsample published 2 estimates of causal effect at maximum but e.g. Elu & Price (2013) reveal 33 causal effects.

Table 3.2: Descriptive statistics of the effects collected from the primary studies

	causal	noncausal	full sample
unweighted mean	0.115	0.009	0.021
weighted mean	0.086	0.007	0.02
median	0.078	0.007	0.007
standard deviation	0.149	0.021	0.063
MIN	-0.102	-0.224	-0.224
MAX	0.89	0.391	0.89
studies	15	64	67
observations	127	957	1084

Note: The table provides basic summary statistics of the effects capturing the impact of height on income broken into subsamples of causal and noncausal effects. Summary statistics for the overall sample are presented as well. Weighted mean uses the inverse of the number of estimates per study as weights. Note that several studies report both causal and noncausal estimates of the height premium. Unwinsorized data are used.

Chapter 4

Publication bias

4.1 Definition of publication bias

Publication bias occurs due to the fact that sometimes studies with unintuitive or statistically insignificant results are not published (they are filed away in a drawer). Card & Krueger (1995) argue that reporting statistically significant or advantageous results appears more persuasive and trustworthy and increases the probability of publishing. Moreover, they also imply that in some cases, the researchers might alter the analysis so that it provides them with the anticipated results. Applied to our case we can imagine it as the negative height premium estimates of several studies that are not published because such results are not in line with the author's expectations. The publication selection phenomenon is not limited solely to economics. It can be observed in all scientific fields. As regards its quantitative effect in general, Ioannidis *et al.* (2017) come to the conclusion that in the economics literature average estimate is overestimated at least by a factor of two.

When we study the literature on wage height premium, we observe that a great number of papers report a statistically significant positive relationship between wage and height. Negative or insignificant effects of height premium are possible but because those are not in line with the results of height premium typically reported by the researchers (i.e. positive and significant), the authors might rarely interpret and publish them. Instead, they are inclined to report findings that are favourable with regard to the probability of publishing. Unfortunately, studies having high contribution potential but not that pleasing results then remain concealed (Ioannidis *et al.* 2017). Such discrimination leads to upward-biased mean estimates and thus publication bias.

However, publication bias does not necessarily equal intentional cheating. Negative estimates in our case could originate in e.g. small sample, bad model specification or measurement error. Thus, abandoning the unintuitive estimated effects benefits the study as focusing on wrong estimates will not improve the study's quality. On the level of individual studies, undertaking such action is understandable and for the sake of their quality and informativeness, one could say even desirable. Nevertheless, suppose this strategy is applied cumulatively by the majority of authors. In that case, the results inferred from the literature as a whole (in meta-analysis achieved via aggregating the published estimates into the mean effect) can be misleading.

To conclude, publication bias is natural to some extent and as argued by Havranek *et al.* (2022), those who review and interpret the literature are supposed to correct for the bias. As will be demonstrated later in this chapter, height premium literature is strongly affected by selective reporting. Applied publication bias tests suggest that researchers refrain from reporting negative and insignificant estimates too often.

4.2 Testing for publication bias

Publication bias tests in general are based on the assumption of no correlation between reported estimates and their standard errors (Havranek *et al.* 2022). The presence of publication bias can be inspected either visually via a funnel plot or more formally by applying linear and non-linear publication bias tests. To support the robustness of our results, we conduct several tests from both linear and non-linear categories.

4.2.1 Funnel plot

Firstly, we will begin the inspection of the presence of publication bias in the height premium literature by employing a visual tool called funnel plot. Funnel plot is a scatter plot of estimates obtained from the primary studies on the horizontal axis and their precision (defined as 1/standard error) on the vertical axis (Sterne *et al.* 2005).

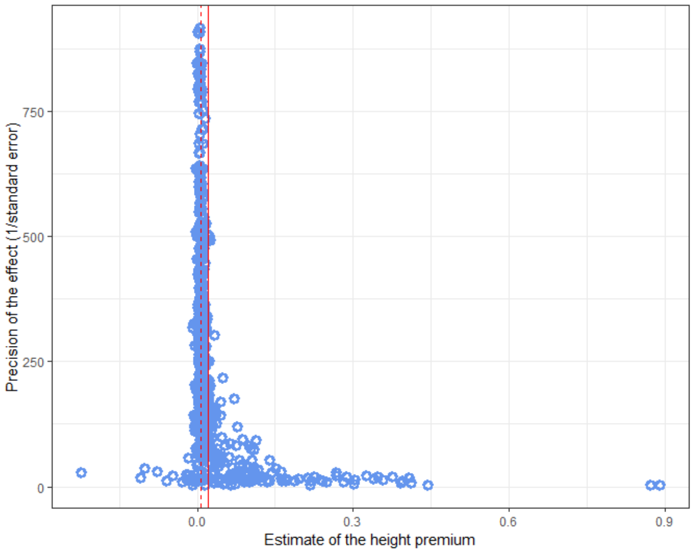
If the publication bias is not present, the funnel plot should take the shape of a symmetrical inverted funnel. Otherwise, its asymmetry suggests that the authors might favour either positive or negative estimates and publication bias possibly needs to be dealt with. It should be noted that the top of the funnel plot hints at the mean height premium corrected for bias. Following the assumption that all the studies we incorporated in the analysis evaluate the same underlying height premium, the most precise estimates are likely to overlap with the underlying mean as remarked by Havranek *et al.* (2022).

Egger *et al.* (1997) explain that the aforementioned notion stems from the hypothesis that the most precise estimates should be located in the top part of the graph coinciding with the true mean whereas as the effects become less precise, the graph widens to the bottom. Studies with large sample sizes are characterized by estimates with small standard errors (and thus high precision), while low precision estimates and large standard errors are typical for studies with small sample sizes. Finally, Stanley (2005) adds that wide and hollow funnel plot reveals that statistically significant results are reported systematically more often than insignificant ones.

The funnel plot of the collected height premium estimates is depicted in Figure 4.1. The estimates and their precision (1/standard error) can be found on the horizontal and vertical axis, respectively. The usage of the inverse of the standard errors (i.e. precision of the corresponding estimates) is a standard practice in meta-analyses (see e.g. Havránek 2015; Bajzik *et al.* 2020; Gechert *et al.* 2022; Havranek *et al.* 2022; Ehrenbergerova *et al.* 2023). The figure presents the estimated effects of both causal and noncausal effects - i.e. the full sample - and their overall mean (solid line) and median (dashed line). We observe that the shape of the plot roughly resembles symmetry, though the prevalence of large positive values and the lack of negative estimates of high magnitude points at its right-skewness and thus, under-reporting of large negative effects in the academic height premium literature. The most precise estimates are close to 0 and overall the graph is more or less centered around this value. Furthermore, judging from the great portion of values at the bottom of the diagram, the authors of the primary studies also abundantly report imprecise estimates.

It should be noted that outliers (in terms of precision) are excluded from the figure for ease of visualisation but otherwise included in all statistical tests. Several studies report standard errors almost equal to 0 and therefore, their precision nears infinity - in the graph, those would be depicted at the top at the level of the solid line representing the mean. Additionally, when creating the funnel plot, we utilized the original dataset we gathered (i.e. unwinsorized). However, Figure 4.1 revealed that the collected coefficients capturing the effect of height on wages contain both positive and negative outliers. Therefore, we perform winsorization at the 1% level to give the extreme values less weight and from now on employ only winsorized data.

Figure 4.1: Funnel plot of the collected effects of height premium for the full sample



Note: The figure depicts a funnel plot of the full sample (i.e. both causal and noncausal) estimates of the relationship between height and wage. The estimated effects are on the horizontal axis and their precision (1/standard error) is on the vertical axis. The solid line represents the mean, the dashed line represents the median. Unwinsorized data are used. However, for the empirical tests, winsorized data will be used. Outliers are excluded for ease of visualisation.

Even though the lack of negative estimates might suggest positive publication bias, we should bear in mind a remark by Stanley (2005) - asymmetry in funnel plot could be also to a certain extent attributable to the various methods and approaches the researchers adopt (i.e. the heterogeneity of studies - will be described in more detail in Chapter 5). On top of that, the funnel plot technique can hardly be considered a substitute for rigorous empirical tests since it heavily depends on the subjective interpretation of the researcher. Although we cannot draw any definite conclusions based on this visualisation tool, it can serve as a first dive into the problem of publication bias in the height premium literature akin to the initial exploratory data analysis in an empirical setting if you will - the purpose of both is to provide us with an initial outlook into our data. To conclude, the existence of publication bias will be further empirically examined by both linear and non-linear publication bias tests - see Section 4.2.2 and Section 4.2.4.

4.2.2 Linear tests

Publication bias tests in general are based on the assumption of no correlation between reported estimates and their standard errors (Havranek *et al.* 2022). Linear publication bias tests rely on the assumption of publication selection being a linear function of the standard error. This can be subsequently verified by regressing the estimates on their standard errors. If the slope coefficient is statistically significantly different from zero, then we can conclude that the reported estimates and standard errors are correlated and consequently, publication bias is present (e.g. Stanley 2005; Havranek & Irsova 2011; Havranek *et al.* 2018a). Note that the intercept in the aforementioned regression is also the true mean effect corrected for publication bias. This can be formally described by the following equation:

$$effect_{ij} = \beta_0 + \beta_1 \cdot SE(effect_{ij}) + \epsilon_{ij} \tag{4.1}$$

where $effect_{ij}$ is the *i-th* estimated effect of the relationship between height and income obtained from the *j-th* primary study (either causal or noncausal), $SE(effect_{ij})$ is the corresponding standard error, β_0 denotes the mean beyond bias effect (i.e. true mean publication bias corrected effect), β_1 and its significance level reveals the existence, direction and size of publication bias and finally ϵ_{ij} is the error term.

Egger *et al.* (1997) remark that since this particular regression is a quantitative alternative to the funnel plot, it is also referred to as a funnel asymmetry test (FAT). As a rule of thumb, FAT testing should be conducted on a dataset containing estimates from at least 10 studies and on top of that, the studies should be of different sizes. Otherwise, the meta-analysis regressions will perform poorly. Fortunately, that is of no concern for us as our dataset consists of effects collected from almost 70 studies, and the respective subsamples of causal and noncausal effects on which we perform the publication bias testing comprise of 15 and 64 studies, respectively[1].

When estimating the discussed meta-regression, we utilize various specifications - OLS, fixed-effects model (FE) that enables to filter out idiosyncratic study-level effects[2], and between-effects model (BE) of between study variance.

[1] Note that several studies report both causal and noncausal estimates of height premium effects - see Table 3.1.

[2] Nevertheless, in our case the FE estimator probably will not be much informative, as in terms of the number of estimated height premium effects per study, our dataset is relatively unbalanced (i.e. contains studies with both small and large number of reported height

Moreover, Equation 4.1 might suffer from heteroskedasticity. This issue is acknowledged by applying a weighting scheme, namely, we assign more weight to more precise estimates. This is achieved via weighting Equation 4.1 with the inverse of the standard errors (1/SE = precision). The meta-regression model then takes the following form:

$$effect_{ij} \cdot \frac{1}{SE(effect_{ij})} = \beta_0 \cdot \frac{1}{SE(effect_{ij})} + \beta_1 + \epsilon_{ij} \cdot \frac{1}{SE(effect_{ij})} \tag{4.2}$$

which can be rewritten as

$$t_{ij} = \beta_0 \cdot \frac{1}{SE(effect_{ij})} + \beta_1 + \nu_{ij} \tag{4.3}$$

Note that t_{ij} is t-value of the *i-th* estimate from the *j-th* study and the coefficients β_0 and β_1 are now reversed - β_0 tests for publication bias, while β_1 captures the mean effect. The Equation 4.3 is base for the precision asymmetry (PET) testing (Stanley 2005). During our analysis, we apply both FAT and PET tests simultaneously. The procedure exploits the fact that as the sample size nears infinity, the standard error reaches zero. Thus, β_0 should be close to the real effect and according to Stanley (2008), we can then test whether the true (publication bias corrected) effect is statistically significantly different from zero with $H_0 : \beta_0 = 0$. In addition, the significance and magnitude of β_1 provide us with information on the presence of publication bias (see the interpretation of the estimated $\hat{\beta_1}$ coefficient based on Doucouliagos & Stanley 2013 below).

Apart from weighting the equation by precision, we additionally use the inverse of the number of estimates per study as weights, so that each study has equal impact (Gechert *et al.* 2022). Lastly, we cannot be certain that standard errors within a specific paper are uncorrelated. Therefore, we use clustering of the standard errors at the study level that explicitly assumes the estimates of the same study to be correlated with each other, but on the other hand independent across different studies. When possible, we also enclose 95% wild bootstrap confidence intervals to account for the fact that clusters might be unbalanced (see e.g. Gechert *et al.* 2019).

The results of the discussed FAT-PET meta-regression models are presented in Table 4.1. The table is divided into two parts - Panel A for causal effects

premium effects) which will subsequently impact the FE estimation results (Havranek *et al.* 2018b).

and Panel B for noncausal effects. Both panels report the results of the unweighted (OLS, FE, BE) and weighted models (weighted by the inverse of the number of estimated collected per study, weighted by precision). As regards the interpretation of the estimated coefficients, Doucouliagos & Stanley (2013) provide us with a guidance:

- if $|\hat{\beta}_1|$ <1 or insignificant, then publication bias is *little to modest*
- if $1 \leq |\hat{\beta}_1| \leq 2$, then publication bias is *substantial*
- if $|\hat{\beta}_1| > 2$, then publication bias is *severe*

Based on the estimated effects from Table 4.1, we can conclude the following. For the subsample of causal effects, we identify substantial publication bias[3]. The estimated true mean effects are lower than the computed weighted mean of the causal subsample (0.086), providing confirmation of the hypothesis that the height premium literature suffers from positive publication bias. In the case of the noncausal subsample, publication bias can be denoted as modest which is further confirmed by the fact that the publication bias corrected mean effects are quite close to the computed weighted mean of 0.007 (see Table 3.2 for the reference of the weighted means). However, Stanley (2005) brings attention to the fact that the FAT-PET results might suffer from the endogeneity of standard errors. For this reason, we also estimate instrumental variable regression with a variety of forms and transformations of study sample sizes used as instruments (see Section 4.2.3).

4.2.3 The problem of endogeneity

The tests conducted in Section 4.2.2 consistently indicate that in the literature focused on height premium, publication bias is likely present. Yet, the results could be biased. Equation 4.1 assumes that in relation to the examined effect the standard errors are exogenous, meaning not correlated with the error term. However, Stanley (2005) suggests that the design of the analysis, sampling error, or other unobserved factors may be correlated with both the estimated height premium effects and their standard errors and thus, produce biased results. To account for the potential endogeneity between the collected

[3]The coefficient of FE model might suggest that the publication bias could be described even as severe. However, we would like to express certain reservations about this result, as our causal subsample contains both studies with low and high number or height premium effects which possibly biases the FE estimation results.

Table 4.1: Linear tests for publication bias

	Standard error (*Publication bias*)	Constant (*Mean beyond bias*)
Panel A: Causal effects		
OLS	0.874***	0.058***
	(0.314)	(0.015)
	[-0.057, 1.928]	[0.002, 0.138]
BE	1.567***	0.028***
	(0.219)	(0.010)
FE	2.242***	0.005***
	(0.115)	(0.001)
weighted study	1.027***	0.028***
	(0.366)	(0.010)
	[0.726, 2.098]	[0.005, 0.053]
weighted precision	1.529***	0.019***
	(0.232)	(0.006)
	[0.627, 2.857]	[-0.004, 0.037]
Panel B: Noncausal effects		
OLS	0.254***	0.008***
	(0.070)	(0.000)
	[-0.012, 0.678]	[0.006, 0.009]
BE	1.311***	0.004***
	(0.074)	(0.000)
FE	-0.397***	0.006***
	(0.033)	(0.000)
weighted study	0.076	0.007***
	(0.072)	(0.000)
	[-0.096, 0.675]	[0.005, 0.008]
weighted precision	0.620***	0.006***
	(0.262)	(0.001)
	[-0.21, 1.631]	[0.001, 0.009]

Note: The table displays the results of linear publication bias tests for causal and noncausal effects. OLS = ordinary least squared, BE = study-level between effects, FE = study-level fixed effects, weighted study = weighted least squares with the inverse of the number of height premium effects per study used as weights, weighted precision = weighted least squares with precision (1/SE) used as weights. Subsample of causal effects contains 15 studies (127 estimates of height premium) and subsample of noncausal effects contains 64 studies (957 estimates of height premium). The signs *, **, *** denote statistical significance at the 10%, 5% and 1%, respectively. Standard errors clustered at the study level are in parentheses. Wild bootstrap 95% confidence intervals in square brackets (we used Rademacher with 9999 iterations).

estimates and their standard errors, we apply the instrumental variable and p-value* approach. Apart from addressing the relaxed exogeneity assumption, the models should also serve as an additional publication bias robustness check.

In the instrumental variable model, we gradually use 4 instruments - the inverse of the number of observations ($1/nobs$); the inverse of the square root of the number of observations ($1/\sqrt{nobs}$); the inverse of the number of observations squared ($1/nobs^2$); and finally logarithm of the number of observations ($log(nobs)$). Both the exogeneity and relevance conditions (key conditions for the instrument to be considered valid) should be fulfilled. The performed diagnostic tests confirm that we have strong instruments. Each instrument is correlated with standard errors (the instrumented variable) and at the same time, it is not probable (though admittedly not completely impossible) that the estimation method and the sample size of the study are correlated.

In their paper van Aert & Van Assen (2021) extend and improve p-uniform and p-value methodologies and present a new method called *p-uniform** which is based on the assumption of uniform distribution of p-values around the true effect. This new method is able to address the three main weaknesses of the previously used p-uniform. That is first and foremost, the usage of statistically significant estimates only (i.e. omission of the insignificant effects) which leads to the p-uniform method being inefficient as the estimates are often accompanied by large variances. Secondly, in the presence of between study variance the effect sizes tend to be overestimated. Thirdly, the presence and possible estimation of between study variance is not tested. Accordingly, van Aert & Van Assen (2021) demonstrate that incorporating statistically insignificant effects into the estimation settles the deficiencies of p-uniform and denotes *p-uniform** as a preferred method.

The results of the specifications mentioned above are shown in Table 4.2. With the exception of the noncausal effect instrumental variable specification using the logarithm of the study's sample size as an instrument, the results suggest that the publication bias in height premium literature can be described as little to modest for both causal subsample and noncausal subsample. Mean beyond bias is however statistically significantly different from zero and for noncausal effects quite close to the computed weighted mean equal to 0.007. On the other hand, contradictory to the previous findings, the estimated true effects for the causal effects subsample are higher than the computed weighted mean (0.086).

Table 4.2: Tests for publication bias with relaxed exogeneity assumption

	Standard error (*Publication bias*)	Constant (*Mean beyond bias*)
Panel A: Causal effects		
IV - $1/nobs$	-0.195	0.110***
	(0.558)	(0.036)
IV - $log(nobs)$	0.566	0.065***
	(0.409)	(0.025)
IV - $1/\sqrt{nobs}$	0.031	0.096***
	(0.491)	(0.032)
IV - $1/nobs^2$	-0.156	0.107***
	(0.696)	(0.041)
p-uniform*		0.165***
Panel B: Noncausal effects		
IV - $1/nobs$	0.591	0.006***
	(0.441)	(0.002)
IV - $log(nobs)$	1.183***	0.002***
	(0.202)	(0.001)
IV - $1/\sqrt{nobs}$	0.907***	0.004***
	(0.273)	(0.002)
IV - $1/nobs^2$	0.342	0.007
	(0.901)	(0.005)
p-uniform*		0.007***

Note: The table shows the results of instrumental variable regressions (with respective instruments) and p-uniform* model by van Aert & Van Assen (2021). Subsample of causal effects contains 15 studies (117 estimates of height premium) and subsample of noncausal effects contains 63 studies (956 estimates of height premium). The signs *, **, *** denote statistical significance at the 10%, 5% and 1%, respectively. Standard errors clustered at the study level are in parentheses. Note that p-uniform* model in statistical software R provides only the mean beyond bias effect and does not test for publication bias.

4.2.4 Non linear tests

Unlike linear methods of testing for publication bias, non-linear models abandon the linearity assumption and instead work with the fact that estimates and standard errors are independent or uncorrelated. Stanley *et al.* (2010) explain that for highly precise estimates located at the top of the funnel plot, the linear publication bias tests are probably overstating the results as estimates of high precision are associated with low standard errors, Or in other words, the publication bias is overestimated, while the mean beyond bias effects reported from linear tests are lower than they should be. Therefore, in line with current meta-analysis research (see e.g. Bajzik *et al.* 2020; Havranek *et al.* 2020; Gechert *et al.* 2022; Matousek *et al.* 2022), to help us uncover the true mean cleaned of the publication bias, we also apply several non-linear tests. The following paragraphs will be dealing a description of the nonlinear publication bias methods employed. The numerical results of the respective tests are presented in Table 4.3.

Stanley *et al.* (2010) introduce a Top 10 method which is based on utilizing only the top 10% of the most precise estimates (hence the name *Top 10 method*). The rest of the estimates are discarded. Assuming the most precise estimates are associated with little publication bias, the true effect cleared of publication bias is then computed via averaging across the top 10% subsample.

Next, we compute weighted average of adequately powered (WAAP). Ioannidis *et al.* (2017) speculate that studies with small estimates and small sample sizes exhibit low statistical power and consequently, are more prone to various biases including publication bias. Thus, Ioannidis *et al.* (2017) propose to employ only estimates with statistical power over 80% (i.e. of adequate power) and compute the mean beyond bias effects as the weighted average of adequately powered (WAAP) with optimal weights set to $1/SE^2$.

Similarly as Top 10 method, Stem-based model is also established around the assumption that the most precise estimates should exhibit minimal publication bias. However, in contrast to Top 10 method, Stem-based by Furukawa (2019) uses a different approach when contemplating whether a particular study should be included in the sample used in the analysis. Instead of arbitrarily setting the threshold at 10% of the most precise estimates, it is determined via minimizing the mean squared error of the estimates. At the same time, the technique tries to find the optimal trade-off between efficiency on one side and precision on the other - the number of observations included should be high enough to warrant efficiency but simultaneously considerably low so that only a few imprecise estimates are included.

To publish, or not to publish, that is the question. As remarked by Hedges (1992), a key factor impacting the decision of the researcher regarding the disclosure of a particular finding is the effect's significance. Andrews & Kasy (2019) further support this claim by stating that significant coefficients have a ten times higher probability of being reported compared to their insignificant counterparts. Following this logic, Andrews & Kasy (2019) develop a selection model that uses standard errors and via the maximum likelihood estimation method determines the probability of certain estimation results being published. In the next step, the weights are re-calculated so that the effects with lower reporting probability have more weight and selective reporting is corrected.

The last test we use is Endogenous kink model by Bom & Rachinger (2019). Similarly as Stem-based model or Top 10 method, Endogenous kink also builds upon the assumption that the most precise estimates are the ones least affected by publication bias. Based on an endogenously determined threshold obtained by piece-wise linear meta-regression, the most precise estimates are extracted and then the mean beyond bias effect is computed via averaging. The threshold is actually the kink (or the intersection) of two linear functions - with high precision (likely unbiased) estimates and potentially biased effects. Bom & Rachinger (2019) also notice that Endogenous kink shares a similar feature with WAAP by Ioannidis *et al.* (2017) - both exploit precision as a selection criterion. Though opposed to Endogenous kink that does not exclude any estimates, WAAP corrects for publication bias only among estimates of adequate statistical power. Also, they differ in the cut-off values - Endogenous kink uses statistical significance, while WAAP relies on statistical power.

The results show that we are indeed correct to hypothesize there is a substantial publication bias in the height premium literature. The nonlinear tests that yield statistically significant results imply that in the case of causal effects, the true effects are lower than the computed weighted mean (0.086). In the case of subsample of noncausal effects, mean beyond bias effect is estimated to be at the level of or in the majority of cases below the computed weighted mean (0.007).

Table 4.3: Nonlinear tests for publication bias

	Constant (*Mean beyond bias*)
Panel A: Causal effects	
Top 10	0.017***
	(0.006)
WAAP	0.026***
	(0.006)
Stem-based	0.007
	(0.015)
Selection model	-0.001
	(0.004)
Endogenous kink	0.005
	(0.004)
Panel B: Noncausal effects	
Top 10	0.005***
	(0.000)
WAAP	0.005***
	(0.000)
Stem-based	0.007*
	(0.004)
Selection model	0.003***
	(0.000)
Endogenous kink	0.006***
	(0.000)

Note: The table displays the results of nonlinear publication bias tests for causal and noncausal effects. Top 10 = Top 10 method proposed by Stanley *et al.* (2010), WAAP = weighted average of adequately powered based on Ioannidis *et al.* (2017), Stem-based = Stem-based method suggested by Furukawa (2019), Selection model = selection model according to Andrews & Kasy (2019), Endogenous kink = endogenous kink model of Bom & Rachinger (2019). Subsample of causal effects contains 15 studies (127 estimates of height premium) and subsample of noncausal effects contains 64 studies (957 estimates of height premium). The signs *, **, *** denote statistical significance at the 10%, 5% and 1%, respectively. Standard errors clustered at the study level are in parentheses.

Chapter 5

Heterogeneity

As was already demonstrated, the reported coefficients of the impact of height on income are not unequivocally consistent (especially for the causal subsample - see figures in Appendix C). They differ in their magnitudes and signs. Although the previous chapter, focused on publication bias analysis, slightly introduced the issue of heterogeneity of studies, now we would like to dedicate a whole chapter to a thorough exploration of this topic. At this point, we have 3 main goals. First, we would like to test whether the publication bias remains present even after we control for various study characteristics. Second, among the factors we take into account, we intend to identify the ones that are responsible for the majority of differences in the height premium estimates. Third, based on the BMA results, we will construct a best-practice estimate that should account for both publication and attenuation bias.

The chapter is structured as follows: Firstly, Section 5.1 deals with the characteristics of the studies' context we collected based on which the heterogeneity testing will be performed. Next, we introduce the model averaging techniques applied (see Section 5.2). Finally, Section 5.3 presents the results of the heterogeneity analysis.

5.1 Explanatory variables

The studies our dataset is created from are indeed diverse. First and foremost it is the type of height premium effect the primary study reports - causal effect, noncausal effect (i.e. correlation) or both. Next, different types of data from distinct countries are utilized. There are 18 countries in our dataset, all of which base the analysis on either cross-sectional or longitudinal data. Moreover, while the majority of authors obtain the estimated effects by applying OLS or instrumental variable approach, there are several who utilize e.g. FE, RE or quantile regression. Also, the analysis is occasionally run on a somewhat restricted sample of individuals - e.g. females, males, self-employed, full-time workers, entrepreneurs, farmers, immigrants, or twins. Last but not least, even though the specification of the regression equation the researchers use is typically based on the Mincer earnings function, additional controls the authors decide to include are not always identical.

To capture the variability across studies, we create 35 variables describing their context. For the overview to be structured, the variables are divided into 6 groups: data characteristics, estimation method, design of the analysis, dependent variable specification, additional controls, and publication characteristics. The groups are detailed below. We are aware of the fact that the list of the variables created is not exhaustive. However, as there is no meta-analysis of height premium on which we could base and expand our own analysis, we decided to employ variables that should be able to capture the main differences, while maintaining a reasonable level of detail and reflecting the common choice of researchers.

Block 1 - Data characteristics

Across the primary studies, different types of data that are naturally characterized by various numbers of observations are used. In the height premium literature we examined, the researchers entertain only two types of data - cross-sectional, and longitudinal (or panel data which we treat as a subset of longitudinal data). To be more specific, longitudinal cohort studies are often employed and they are generally (though not always by default) a bit more extensive in terms of the number of subjects involved. To control for this aspect, we create variable *Sample size* that should address the possible imbalances by taking the logarithm of the study's sample size. Also, dummy variable *Longitudinal data* is equal to 1 for longitudinal (panel) data and equal to 0 for cross-sectional data.

Block 2 - Estimation method

The majority of the collected height premium effects were estimated with the help of the ordinary least squares (OLS) approach. The second most frequent estimation method is instrumental variable (IV) regression. Moreover, some authors employ also random effects (RE) and fixed effects (FE) panel methods (Behrman & Rosenzweig 2001; Hill 2004; Heineck 2005; Johnston 2010; Lee 2014; Kim & Han 2017), quantile regression (Kedir 2008; Sohn 2015b; Peng *et al.* 2020), or truncated regression and censored regression maximum likelihood methods (Eschker *et al.* 2004).

We decided to create the following dummy variables - *OLS*, *IV*, *Panel*, *QR*, and *MLE* to denote that the researchers used OLS, instrumental variable approach, panel methods (FE, RE), quantile regression, or maximum likelihood methods (truncated regression, censored regression), respectively. Owing to the fact that these techniques are based on various assumptions and the ways in which the differences of individuals are captured are diverse as well, we expect the estimation methods to impact the collected estimates considerably.

Block 3 - Design of the analysis

In contrast to publication bias testing, the heterogeneity analysis will be performed on the full sample of the collected effects. Therefore to encompass the variability of the methods estimating causal and noncausal associations and their underlying assumptions, we create binary variable *Endogeneity* and assign it with value 1 if the study controls for potential endogeneity of height.

As was already mentioned, the studies utilize data from 18 countries representing all continents, naturally with the exception of Antarctica. Therefore, following the geographical division, we decided to group the individual countries into 6 categories - *Australia*, *Africa*, *America*, *Asia*, *Europe*, and *USA*. The United States is put into a separate category because the remaining countries located in North and South America (namely Mexico, Colombia, and Brazil) are arguably less developed and height premium effects collected from developing countries exhibit more variability (see Figure 3.2). Moreover, we would like to examine whether the heterogeneity of height premium effects can be to a certain extent also driven by a gender of an individual. The gender differences in height premium effects are supported by empirical findings. For example Heineck (2008) suggests that the height premium coefficients are smaller for men, while Yang *et al.* (2018) claim the opposite and Case & Paxson (2008b) provide mixed results. For that reason, we establish dummy variables *Male only*, *Female only* indicating whether the primary study was conducted on a sample restricted only to males or females, respectively.

Block 4 - Dependent variable specification

It could be said that height premium literature discerns four basic forms of the dependent variable with respect to time - annual, monthly, weekly, and hourly. As a consequence, we create four binary variables *Dependent: annual*, *Dependent: monthly*, *Dependent: weekly*, *Dependent: hourly* equal to 1 if the dependent variable is of the respective specification and 0 otherwise. Apart from the qualifications above, wealth, wealth index, income score, expected salary, or career earnings are occasionally used. Those will be hidden under the variable *Dependent: other*.

Another possible distinction would be on the basis of the exact kind of the response variable - i.e. salary, wage, income, or earnings. But e.g. OECD does not distinguish between wages and earnings and defines both identically as *the total remuneration, in cash or in kind, payable to all persons counted on the payroll (including homeworkers), in return for work done during the accounting period*. Moreover, the authors of the primary studies do not pay much attention to this issue as well. Hence, throughout this thesis, we use the terms interchangeably and do not adopt them as criteria of any kind.

Block 5 - Additional controls

Although the general regression specification in the height premium literature follows the Mincer equation, the researchers augment it with controls they acknowledge as important. For example Case & Paxson (2008b) argue that incorporating cognitive ability is essential as height and cognitive ability are linked and that is consequently reflected in the labour market outcomes. The identical view is shared by Sohn (2015b). Moreover, Thomas & Strauss (1997) empirically demonstrate that controlling for employment sectors is important as well because, among the sectors, certain human capital endowments might be rewarded differently with the following case in point: *A labourer, for example, would presumably reap returns from strength and stature, whereas those characteristics are unlikely to be rewarded, in and of themselves, in a more sedentary occupation*. Therefore, dummy variables *Cognitive ability* and *Sector* are created. They are assigned a value of 1 if control variables for cognitive ability and employment sector are used, respectively.

The same logic applies to variables *Weight*, *Social class*, *Marital status* and *Parental* - they are equal to 1 if the regression specification accounts for the individual's weight, social class, marital status, or parental background characteristics (i.e. education, social class, earnings, height, political membership, or any other characteristic describing mother or father of an individual), accordingly. We also include dummies that control for (by empirical literature established) gender and ethnicity-based pay gap - *Gender* and *Ethnicity.* Last but not least, during the data collection we also paid attention to the fact whether the regression model contains information on the type of job that would specify the job in more detail (e.g. leading position, direct communication with customers etc.) or type of employment (self-employed, paid employee, unemployed, not economically active)

It is worth noting that we do not create dummy indicators for the presence of education or experience controls in the primary study regression. That is because conventionally, the researchers account for those on a regular basis. Thus, accounting for these particular control variables would not prove very helpful in capturing the variability of the regression models utilized.

Block 6 - Publication characteristics

Adding variables that to a certain extent gauge the quality of the primary study not reflected in data or methods is a common practice in meta-analyses (e.g. Bajzik *et al.* 2020; Havranek *et al.* 2022; Matousek *et al.* 2022). For that reason, we create the following variables *Publication year*, *Citations*, *Peer-reviewed*, and *Impact factor*.

The time span of studies we use ranges from 1994 to 2022. Therefore, the purpose of *Publication year* is to capture possible methodological changes and improvements that may have occurred in time. It is computed as the logarithm of the year the study was published minus 1994 (the year the oldest study in our dataset was published). With citations one ought to be careful, a high number of citations does not necessarily guarantee the top quality of the study. Arguably, sometimes it can be the case that a certain study is cited plentifully because it is showcased as a bad example. Also, the high number of citations can be associated with the long-time existence of the paper. Thus, *Citations* is defined as the logarithm of the total number of Google Scholar citations at the time of data collection divided by the number of years from the publishing year. *Peer-reviewed* is a dummy variable indicating whether the study was published in a peer-reviewed journal. Lastly, *Impact factor* stands for IDEAS/RePEc simple impact factor[1]. It should be noted that in

[1]We prefer RePEc recursive impact factor over JCR impact factor because the majority

several instances, the IDEAS/RePEc recursive discounted impact factor was not available (e.g. because the primary study was a discussion paper or it was published in a journal not included on the IDEAS/RePEc ranking list). In that case *Impact factor* is set to zero.

5.2 Estimation method

In the previous section, we established 35 variables reflecting the context in which the individual estimates were acquired (i.e. the heterogeneity of the respective studies). Now, we would like to determine whether there exists a relationship between the height premium effect (dependent variable) and the study characteristics (independent variables). In other words, whether the variables described above are able to capture the variability of the primary studies. Therefore, we will augment the model we already used for publication bias testing with additional control variables standing for the heterogeneity of studies. The equation will hence take the following form:

$$effect_{ij} = \beta_0 + \sum_k \beta_k X_{k,ij} + \delta \cdot SE(effect_{ij}) + \epsilon_{ij} \tag{5.1}$$

where $effect_{ij}$ is the *i-th* estimated effect of the relationship between height and income obtained from the *j-th* primary study (either causal or noncausal), X_{ij} contains additional controls described in Section 5.1, $SE(effect_{ij})$ is the corresponding standard error, β_0 denotes the mean beyond bias effect (i.e. true mean publication bias corrected effect), δ and its significance level reveals the existence, direction and size of publication bias and finally ϵ_{ij} is the error term.

The next step is thus to include the variables from Table 5.1 into our standard regression. Some of them are grounded in theoretical rationale, while others are included without any theoretical base and serve mainly as additional controls. Though we would like to believe that all of the abovementioned variables could be relevant, some will be inevitably redundant. But as we are not able to gauge their importance in advance, we encounter uncertainty. On the other hand, trying to find the optimal combination of K explanatory variables manually would lead to estimating 2^K regressions, which is both timewise and computationally challenging. Also, lowering the number of explanatory variables arbitrarily by force might appear as an acceptable approach but in that

of the studies in our dataset have been published in economic journals - i.e. type of journals RePEc commonly evaluates.

Table 5.1: Descriptive statistics of variables used in the heterogeneity analysis

Variable	Description	Mean	SD
Effect	causal (noncausal) effect of height premium	0.018	0.041
Standard error	standard error of the causal (noncausal) effect	0.010	0.022
Data characteristics			
Sample size	logarithm of the study's sample size	8.210	1.897
Longitudinal data	= 1 if primary study uses longitudinal (or panel) data; = 0 for cross-sectional data	0.614	0.487
Estimation methods			
OLS	= 1 if OLS approach is used for the estimation	0.803	0.398
IV	= 1 if instrumental variable approach is used for the estimation	0.105	0.307
Panel	= 1 if panel methods (RE, FE) are used for the estimation	0.042	0.201
MLE	= 1 if maximum likelihood methods (truncated or censored regression) are used for the estimation	0.023	0.150
QR (reference category)	= 1 if quantile regression is used for the estimation	0.026	0.159
Design of the analysis			
Endogeneity*	= 1 if estimation method accounts for potential endogeneity of height	0.105	0.307
Australia (reference category)	= 1 if country surveyed is part of Australia	0.028	0.164
Africa	= 1 if country surveyed is part of Africa	0.047	0.212
Asia	= 1 if country surveyed is part of Asia	0.247	0.432
Europe	= 1 if country surveyed is part of Europe	0.433	0.496
America	= 1 if country surveyed is part of North or South America (except the USA)	0.036	0.187
USA	= 1 if country surveyed is part of the USA	0.208	0.406
Male only	= 1 if the estimates are obtained on sample of males	0.603	0.489
Female only	= 1 if the estimates are obtained on sample of females	0.262	0.439
Dependent variable specification			
Dependent: annual	= 1 if dependent variable is specified as annual	0.322	0.467
Dependent: monthly	= 1 if dependent variable is specified as monthly	0.155	0.362
Dependent: weekly	= 1 if dependent variable is specified as weekly	0.018	0.134

Table 5.1: Descriptive statistics of variables used in the heterogeneity analysis (continued)

Variable	Description	Mean	SD
Dependent: hourly	= 1 if dependent variable is specified as hourly	0.0382	0.486
Dependent: other (reference category)	= 1 if dependent variable is specified in other way	0.122	0.327
Additional controls			
Cognitive ability	= 1 if primary study regression controls for cognitive ability	0.119	0.325
Sector	= 1 if primary study regression controls for occupation sector	0.022	0.147
Weight	= 1 if primary study regression controls for weight	0.063	0.243
Social class	= 1 if primary study regression controls for social class	0.009	0.093
Marital status	= 1 if primary study regression controls for marital status	0.201	0.401
Parental	= 1 if primary study regression controls for parental background	0.160	0.367
Gender	= 1 if primary study regression controls for gender	0.088	0.284
Ethnicity	= 1 if primary study regression controls for ethnicity	0.227	0.419
Job type	= 1 if primary study regression controls for type of job	0.174	0.379
Employment type	= 1 if primary study regression controls for type of employment	0.097	0.296
Publication characteristics			
Publication year	logarithm of the year the study was published minus 1994 (base year)	2.871	0.352
Citations	logarithm of the total number of citations (collected in January 2023) divided by the number of years from the publishing year	1.186	1.429
Peer-reviewed	= 1 if the study was published in peer-reviewed journal	0.896	0.305
Impact factor	RePEc recursive impact factor (collected in July 2023)	0.439	0.779

Note: The table presents the additional explanatory variables with their description and summary statistics. Mean = simple unweighted mean, SD = standard deviation. Dummy variables excluded due to dummy variable trap are denoted as reference category. Variables with asterisks are omitted from the analysis due to multicollinearity reasons.

way, we would neglect the variability of studies. As a solution to this dilemma, meta-analyses researchers recommend employing averaging techniques (see e.g. Havranek *et al.* 2018a; Bajzik *et al.* 2020; Gechert *et al.* 2022). In this thesis, we will apply the Bayesian model averaging technique (BMA) and as a robustness check, we will also estimate the Frequentist model averaging model (FMA) and Frequentist check. The models will be described in the paragraphs below.

5.2.1 BMA explanation

In principle, the Bayesian model averaging method (BMA) is established on averaging all the regression specifications resulting from various combinations of explanatory variables. Utilizing the methodology from Hasan *et al.* (2018), imagine the following model:

$$y = \alpha_s + X_s\beta_s + \epsilon \tag{5.2}$$

with X_s as a subset of K explanatory variables. Then there exists 2^K possible regression models denoted as $M_1, ..., M_s$, where $s \in [1, 2^K]$, that estimate 2^K possible combinations of the explanatory variables. The BMA then revolves around the subsequent key components:

Posterior Model Probability

Stemming from Bayes' rule, BMA defines Posterior model probability (PMP):

$$p(M_s \mid y, X) = \frac{p(y \mid M_s, X)p(M_s)}{p(y \mid X)} = \frac{p(y \mid M_s, X)p(M_s)}{\sum_{s=1}^{2^K} p(y \mid M_s, X)p(M_s)} \tag{5.3}$$

where $p(M_s)$ denotes prior model probability (i.e. the probability the researcher assigns to the model *prior* looking at the data), $p(y \mid X)$ is constant integrated (marginal) likelihood and $p(y \mid M_s, X)$ model's integrated likelihood corresponds to the probability of data given model M_s (Zeugner & Feldkircher 2015).

Posterior model probability is crucial for weighting the models. Moreover, Havranek *et al.* (2015) add that PMP of separate models can serve as an indicator of the model's quality analogous to adjusted R squared.

Posterior Mean

As already mentioned, with the help of PMP, the actual weighted posterior means for the explanatory variables are computed as

$$E(\beta \mid y, X) = \sum_{s=1}^{2^K} E(\beta_s \mid M_s, y, X) p(M_s \mid y, X) \tag{5.4}$$

where $E(\beta \mid y, X)$ represents the weighted posterior mean of the variables' coefficients and $E(\beta_s \mid M_s, y, X)$ are the β_s coefficients estimated for model M_s. Also, the following equation describes how the posterior distribution of the coefficients is impacted by prior g, where $\hat{\beta}_s$ is OLS estimate (Hasan *et al.* 2018):

$$E(\beta_s \mid y, X, g, M_s) = \frac{g}{1+g} \hat{\beta}_s \tag{5.5}$$

Posterior Variance

Hasan *et al.* (2018) define weighted posterior variance in the following way:

$$Var(\beta \mid y, X) = \sum_{s=1}^{2^K} Var(\beta_s \mid M_s, y, X) p(M_s \mid y, X) + \sum_{s=1}^{2^K} (E(\beta_s \mid M_s, y, X) - E(\beta \mid y, X))^2 p(M_s \mid y, X) \tag{5.6}$$

where $E(\beta \mid y, X)$ is posterior mean already mentioned in Equation 5.4, $Var(\beta_s \mid M_s, y, X)$ is weighted average of variances across various regressions and $(E(\beta_s \mid M_s, y, X) - E(\beta \mid y, X))^2$ is weighted variance across different models.

Posterior Inclusion Probability

Last but not least, BMA also defines Posterior inclusion probability (PIP) as the summation of posterior model probabilities but the sum is limited only to models that contain variable k:

$$PIP = p(\beta_k \neq 0 \mid y, X) = \sum_{s=1}^{2^K} p(M_s \mid \beta_k \neq 0, y, X) \tag{5.7}$$

Based on PIP, we can determine whether a particular explanatory variable is a good predictor of the dependent variable or in other words, we could imagine it as a concept similar to significance, since PIP informs us of the probability the variable k will be included in the true model. PIP can take values from 0 to 1 and according to its magnitude, Jeffreys (1961) recommends the following distinction[2]:

- if PIP is between 0.5 and 0.75, then the effect is *weak*
- if PIP is between 0.75 and 0.95, then the effect is *positive*
- if PIP is between 0.95 and 0.99, then the effect is *strong*
- if PIP is between 0.99 and 1, then the effect is *decisive*

Priors

BMA estimation also requires a specification of two distribution priors - g and $p(M_s)$. These are priors over the parameter space and the model space, respectively, and can be found in Equation 5.3 and 5.5 above. The selection of priors should reflect all the available information the researchers have. However, in the majority of cases the prior knowledge is not vast. Therefore, typically uniform model prior (UMP) and unit information prior (UIP) are used. Uniform model prior characterizes the model space and assigns each model with equal prior probability. Unit information prior applied over the parameter space assumes that information in the prior and in the typical observation are almost identical. Via empirical testing, Eicher *et al.* (2011) confirm that the practice of preferring UMP and UIP priors over the others is for the best if the distribution is unknown to the researchers because, in the unfamiliar setting, these priors usually outperform the others.

Markov chain Monte Carlo

It should be noted that BMA is computationally very demanding. The estimation of 2^K regression models and solving the integrals of integrated likelihood $p(y \mid M_s, X)$ in PMP (Equation 5.3) can be very time-consuming. For the ease of calculation, BMA adopts Markov chain Monte Carlo approximation with Metropolis-Hastings algorithm specification. Markov chain Monte Carlo method allows only models with high PIP to be estimated. Those are selected with the help of Metropolis-Hastings algorithm - PMP of the competing model is compared with the PMP of the benchmark model. If the benchmark model is rejected in favour of the current model, the current model is labelled as a

[2]The effectiveness refers to the model's ability to explain the variance of the results, e.g. if PIP is equal to 1, then the all the effective models include the variable in question.

new benchmark model and compared to a new candidate model. Otherwise, the benchmark model remains and is compared with the competing model next in line. As at the beginning, a model with low PMP and low marginal likelihood could be selected, several of the initial iterations oftentimes need to be excluded (Zeugner & Feldkircher 2015).

5.2.2 BMA implementation

Even though we would be tempted to incorporate all variables from Table 5.1 into our regression, there exist several reasons that prevent us from doing so.

Firstly, the dummy variable trap. Including all of the dummy variables into our model will result in perfect multicollinearity (i.e. dummy variables will be perfectly correlated) which would influence the overall reliability and predictive power of our model. For that reason, it is essential to omit one variable (in Table 5.1 denoted as reference category) for each category that was created solely on the binary variable basis, meaning if the categorical variable can be assigned k different levels, the number of dummy variables we use in the regression is $k-1$.

Secondly, even after accounting for the dummy variable trap, we still need to inspect correlations between the respective variables. The correlation matrix in Figure D.1 (Appendix D) shows that the correlation between *Endogeneity* and *IV* is equal to 1. That is not surprising, as the instrumental variable approach controls for endogeneity. Also e.g. *OLS* and *IV*, or *Citations* and *Impact factor* are correlated. But based on Ratner (2009), the exhibited correlation is mild (ranges between 0.3 and 0.7 in absolute values). Therefore, after we drop variable *Endogeneity* for the sake of results' precision, multicollinearity should be dealt with. However, to account for the mild collinearity between various variables, instead of typically used uniform model prior, we will apply collinearity adjusted dilution model prior (George 2010) that does not maintain the zero correlation assumption.

Lastly, one additional adjustment was necessary before we inputted the data into the BMA model. As was already mentioned during the heterogeneity variables' description, several studies we used for data collection were not yet officially published or were published in a journal not included in the ranking of the IDEAS/RePEc recursive discounted impact factor. Overall, those could not be complemented with corresponding impact factor value. Hence, we opted for setting them to 0, otherwise, these data points would be disregarded by default.

For the BMA modelling, we utilize the *BMS* package by Zeugner & Feldkircher (2015). Because we know little of the parameter space, we follow the general practice and choose to use unit information prior (UIP). As regards the model space prior specification, due to mild collinearity between explanatory variables discussed above, we will apply the collinearity adjusted dilution model prior. We consider BMA with unit information prior and dilution model prior could as our baseline model and we also provide FMA (Frequentist model averaging) and OLS Frequentist check as a robustness check. Both FMA and Frequentist check will be described in the following sections.

5.2.3 FMA

Similarly to BMA, the frequentist model averaging technique (FMA) is also a powerful tool frequently employed in heterogeneity analyses (see e.g. Havranek *et al.* 2017; Gechert *et al.* 2022; Havranek *et al.* 2022). As recommended by Bayarri & Berger (2004), we will profit from the benefits of both of them and apply FMA along with the BMA model which in our case will serve as a robustness check for the BMA results. As opposed to BMA, in the case of FMA we are not expected to specify any priors but on the other hand, both FMA and BMA adopt a similar strategy when dealing with the model uncertainty - averaging across different models. Following Havranek *et al.* (2021)[3], we implement the Mallows model average estimator that aims to minimize the Mallows criterion and consequently select asymptotically optimal weights and improve the goodness-of-fit of the model (Hansen 2007). On top of that, as a result of the orthogonalization of the covariate space, the number of models that need to be estimated drops significantly from 2^K to only K models (Amini & Parmeter 2012).

5.2.4 Frequentist check

The other method we will use as a robustness check of our baseline BMA specification is Frequentist check. In this case, the standard errors are clustered at the study level. The mechanism of Frequentist check is based on simple OLS estimation. Nonetheless, it utilizes the results from BMA analysis as the model includes only variables for which the estimated posterior inclusion probability (PIP) is 0.5 at minimum (i.e. variables that exhibit at least weak importance).

[3]The empirical part of the FMA analysis is based on code provided by Havranek *et al.* (2021) in their online appendix.

5.3 Results

This section provides an overview of the results of the heterogeneity analysis performed on the full sample of height premium estimates. The final number of variables used is 32 including standard error. We excluded *Endogeneity* due to multicollinearity reasons and *QR*, *Australia*, *Dependent: other* were omitted to prevent the dummy variable trap (those are in Table 5.1 marked as reference category).

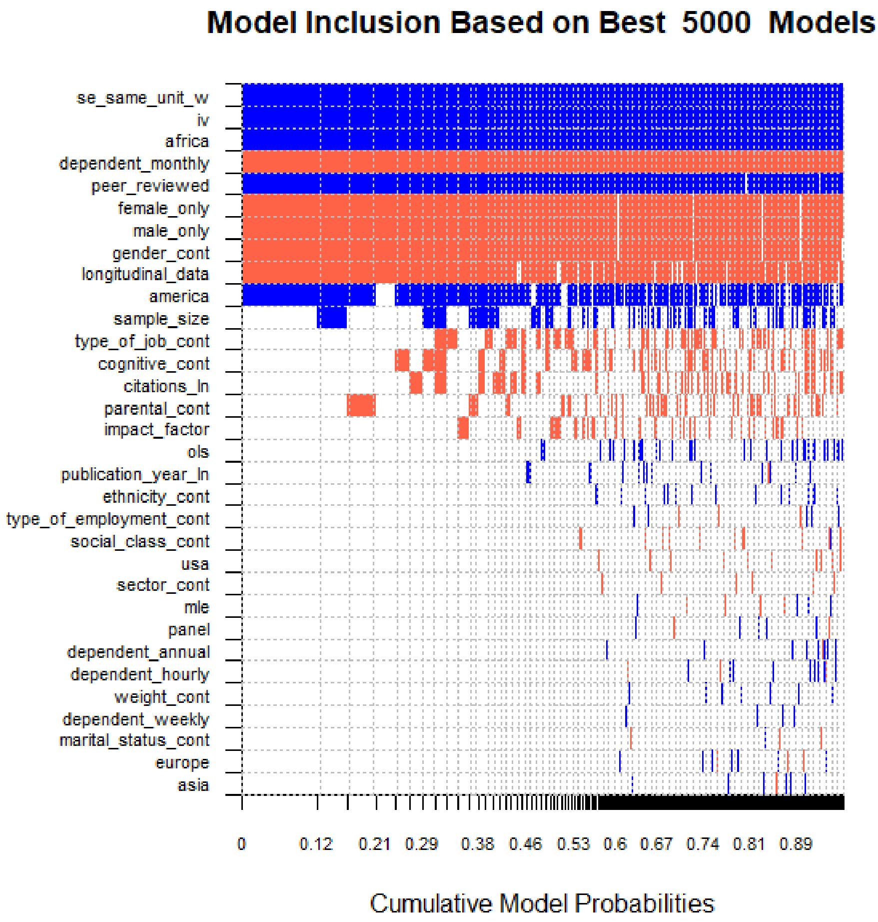
Figure 5.1 presents the visual results of the BMA model with unit information prior and dilution model prior. The explanatory variables can be found on the vertical axis, sorted according to posterior inclusion probability (PIP), meaning the most significant regressors with the highest PIP will be located at the top part of the plot, while those that are least likely to be included into the model are displayed at the bottom. The horizontal axis lists the posterior model probabilities. The columns refer to models and the width of the column represents the posterior model probability (PMP) of each model. The colours used in the figure carry the following type of information - red (will appear lighter in the greyscale picture) signifies the negative coefficient of a particular variable; blue (will appear darker in the greyscale picture) indicates a positive coefficient; white means the variable is not included in the model. By inspecting the graph, we identify 10 variables with PIP above 0.5 - *Standard error*, *Longitudinal data*, *IV*, *Africa*, *America*, *Male only*, *Female only*, *Dependent: monthly*, *Gender*, and *Peer-reviewed*.

The numerical outcomes of the BMA analysis are provided in Table 5.2. Also, Table 5.2 and Table 5.3 present Frequentist check and FMA specification results, respectively. On the grounds of the PIP interpretation based on Jeffreys (1961) (as was detailed in Section 5.2.2), the variables identified from the inclusion model plot as important can be assigned with the following labels - *America* has a positive effect, *Longitudinal data* and *Gender* exhibit a strong effect, while the findings suggest a decisive effect of *Standard error*, *IV*, *Africa*, *Male only*, *Female only*, *Dependent: monthly*, and *Peer-reviewed*. The estimation results are discussed in more detail below.

Block 1 - Publication bias and data characteristics

As regards the presence and sign of publication bias, we can see that BMA model assigns *Standard error* with positive posterior mean and PIP equal to 1, indicating that the positive publication bias persists even after we control for

Figure 5.1: Model inclusion of the BMA estimation



Note: The figure displays the BMA results in graphical form. The variables are listed on the vertical axis in descending order according to their posterior inclusion probabilities (PIP), posterior model probabilities are on the horizontal axis. Red (darker) = negative coefficient, blue (darker) = positive coefficient, white = variable is not included in the model.

additional variables of heterogeneity. Moreover, FMA and Frequentist check models estimate the relation between standard error and height premium coefficients to be positive and highly statistically significantly different from zero as well. Overall, this finding is in line with the conclusions of publication bias testing conducted in the previous chapter and we can surmise that the evidence of positive publication bias is convincing.

It is interesting to see that according to the findings of BMA model, the nature of the dataset, longitudinal (or panel) to be more specific, is associated with a systematic reporting of more negative estimates in the height premium literature, judging from the sign of posterior mean and the PIP value of *Longitudinal data*. This is again confirmed by both FMA and Frequentist check.

Presumably, *Sample size* has no impact on the heterogeneity of height premium estimates.

Block 2 - Estimation method

Among the estimation methods the researchers employ, *IV* has a decisive effect in explaining the differences in the collected height premium estimates. The sign of posterior mean of the variable *IV* together with the corresponding PIP value imply that the instrumental variable method systematically delivers positive effects when capturing the impact of height on income. Sometimes when employing the arguably more advanced methodologies, the researchers might encounter serious difficulties. In the case of the instrumental variable technique, it might prove difficult to find an instrument satisfying both the exogeneity and relevance assumptions. When instrumenting height in relation to income, the authors typically use food prices, genetic score, sibling's height, parent education, regional conditions, average number of health institutions, or ethnicity as instruments and the BMA analysis results indicate that such conduct is appropriate. This fact is also corroborated by McGovern *et al.* (2017) who demonstrate that studies adopting the instrumental variable strategy are characterized by higher returns to height in comparison with e.g. OLS.

Except for the instrumental variable approach, other estimation methods yield no significant differences and do not contribute to the height premium estimates heterogeneity.

Block 3 - Design of the analysis

BMA model denotes the effect of variables *Africa* and *America* as decisive and positive, respectively. This indicates that belonging to certain parts of the world has a significant impact on explaining the variety among the collected estimates. This is further confirmed by both FMA and Frequentist check.

In comparison with studies utilizing datasets from e.g. USA or Europe, the analyses conducted on countries in Africa or America (excluding the USA) provide more positive estimates of the height-income relationship. This implication is not shocking as the boxplot of the height premium effect across countries (Figure 3.2) has already hinted such an effect might be possible, especially among the developing countries of Africa and America. Moreover, academic literature partially signifies such a notion is not far from the truth (e.g. Ribero 2000; Kedir 2009; Sohn 2015a) and Hübler (2009) specifically brings attention to the fact that in developing countries, the wage returns to height tend to be higher.

Moreover, the outcomes of this block of heterogeneity characteristics also suggest that restricting a sample on which the primary estimation is performed on the basis of gender has a meaningful impact on the heterogeneity of the height premium estimates as also remarked by e.g. Case & Paxson (2008b) or Yang *et al.* (2018). This is further supported by FMA as well as Frequentist check.

Block 4 - Dependent variable specification

With respect to the type of the dependent variable, the BMA model concludes that the dependent variable in monthly form (*Dependent: monthly*) is a determinant that influences the height premium effects negatively. The remaining specifications of the dependent variable are not factors that would impact the differences in height premium estimates. That is however in contradiction with the results provided by FMA that identifies *Dependent: annual* as statistically significant at the 5% level and the remaining variables *Dependent: monthly*, *Dependent: weekly*, *Dependent: hourly* as statistically significant at the 10% level.

Block 5 - Additional controls

From all the additional control variables the authors of the primary studies augment the baseline Mincer equation with, the only one that is able to capture the heterogeneity among height premium effects is *Gender*. Though perhaps a bit unsatisfactory, the finding is not surprising because as has already been demonstrated in Block 3 - Design of the analysis, gender contributes to the variety of the estimates, be it by directly involving gender control in the model specification, or via gender-based sample restriction.

Admittedly, we expected more of the variables from the block of additional controls to be recognized as notable. For example Case & Paxson (2008b) highlight the role of cognitive ability and mark it as one of the channels through which the height premium effects can be explained. Therefore, Case & Paxson (2008b) consider the inclusion of cognitive ability control into the regression specification as essential. Also, the findings of Longhi & Brynin (2017) and Evans (2020) propose that accounting for ethnicity might be potentially important. They estimate that Bangladeshi ethnic minority living in the UK experiences on average a pay gap amounting to 20%. Moreover, Hübler (2009) advises not to leave out sector specification from the regression model because the height premiums vary across sectors. Last but not least, including parental controls, be it income, height, education level, or social class of mother and father, might help to address a substantial portion of the height premium effect as remarked by Vogl (2014).

Nevertheless, the results of the BMA estimation show that controlling for additional aspects other than gender does not contribute to explaining the differences among the height premium estimates.

Block 6 - Publication characteristics

In the category of publication characteristics, the heterogeneity analysis of the linkages between height and income generates the following results.

We find little empirical evidence that IDEAS/RePEc recursive impact factor, year of publishing, or number of Google Scholar citations represented by variables *Impact factor*, *Publication year*, and *Citations* would explain the variability among the height premium estimates or systematically impact the effects the authors of the primary studies report.

On the other hand, studies published in peer-reviewed journals systematically generate positive estimates of height premium and the effect is classified as decisive (as described by posterior mean of the variable *Peer-reviewed* and its PIP value). However, it could be argued that a revelation of this kind is not very much shocking because if the authors want their results to be published, they might be inclined to disclose effects that are plausible and in general agreement with the broad evidence supplied by the academic literature.

Those findings are in accordance with respect to all the specifications - BMA, FMA, and Frequentist check.

5.4 Best practice estimate

The results of the previous analyses demonstrate that there exists empirical evidence for the presence of publication bias and also that the additional control variables we utilized are able to explain the variability of the height premium estimates. At this point, we will aggregate the previous findings via the synthetic approach of best practice estimate, meaning we will try to estimate the height premium effect corrected for publication bias as well as for various study characteristics. The best practice estimation method utilizes linear regression with the variables from BMA as independent variables. By setting the values of explanatory variables to their minimum, maximum, or mean we specify our preferences.

In our baseline best practice model specification, we assign *Standard error* with its minimal value, as we aim to eliminate publication bias. Next, we would also like to remove possible endogeneity bias, so we set IV to its maximum. As regards the dummy variables that represent additional controls in the Mincer equation, we plug them all with 1. Lastly, we also consider the quality of the estimates. Therefore, we set the publication characteristics to their maximum values (though BMA modelling disregards some of them as unimportant). The only exception is *Citations* which we assign with its mean value. We have already expressed our reservations in connection to the number of citations as a measure of the study's quality. We believe that a high number of citations does not guarantee the superior quality of a particular study. For the rest of the variables we do not hold any preferences, thus we set them to their mean values.

Apart, from the baseline specification, we also construct best practice estimate for Africa, America, and gender-restricted settings as those are brought to our attention based on the BMA outcomes.

The results for the respective specifications and the approximations of 95% confidence intervals are given in Table 5.4. The best practice estimate stemming from the baseline specification is equal to 0.036 which is more or less at the level of mean beyond bias effects of causal subsample reported in Chapter 4.

Table 5.2: Explaining heterogeneity in height premium estimates - BMA and Frequentist check specification

	BMA			*Frequentist check*		
	Post. mean	*Post. SD*	*PIP*	*Coef.*	*SE*	*p-value*
Standard error	0.761	0.049	1.000	0.763	0.192	0.000
Data characteristics						
Sample size	0.000	0.001	0.340			
Longitudinal data	-0.007	0.003	0.927	-0.007	0.003	0.033
Estimation methods						
OLS	0.001	0.002	0.083			
IV	0.035	0.004	1.000	0.034	0.009	0.000
Panel	0.000	0.001	0.021			
MLE	0.000	0.002	0.022			
Design of the analysis						
Africa	0.055	0.005	1.000	0.055	0.020	0.006
Asia	0.000	0.000	0.017			
Europe	0.000	0.000	0.018			
America	0.014	0.007	0.873	0.015	0.005	0.004
USA	-0.000	0.000	0.023			
Male only	-0.020	0.004	0.990	-0.020	0.010	0.035
Female only	-0.021	0.004	0.990	-0.020	0.009	0.024
Dependent variable specification						
Dependent: annual	0.000	0.000	0.021			
Dependent: monthly	-0.020	0.003	1.000	-0.020	0.008	0.010
Dependent: weekly	0.000	0.001	0.018			
Dependent: hourly	0.000	0.000	0.019			
Additional controls						
Cognitive ability	-0.002	0.003	0.255			
Sector	-0.000	0.001	0.022			
Weight	0.000	0.001	0.019			
Social class	-0.000	0.002	0.025			
Marital status	-0.000	0.000	0.018			
Parental	-0.001	0.002	0.177			
Gender	-0.020	0.005	0.988	-0.019	0.010	0.054
Ethnicity	0.000	0.001	0.038			
Job type	-0.002	0.003	0.268			
Employment type	0.000	0.001	0.025			
Publication characteristics						
Publication year	0.000	0.001	0.044			
Citations	-0.000	0.001	0.200			
Peer-reviewed	0.014	0.004	0.998	0.013	0.006	0.032
Impact factor	-0.000	0.001	0.150			
Intercept	0.014	NA	1.000	0.019	0.010	0.048

Note: The table displays the results of the BMA and Frequentist check model specification. Post. mean = posterior mean, Post. SD = posterior standard deviation, PIP = posterior inclusion probability. BMA is set to use unit information prior and dilution prior. Frequentist check incorporates only variables with PIP above 0.5. The explanatory variables are described in Table 5.1.

Table 5.3: Explaining heterogeneity in height premium estimates - FMA specification

	FMA		
	Coef.	*SE*	*p-value*
Standard error	0.754	0.049	0.000
Data characteristics			
Sample size	0.002	0.001	0.005
Longitudinal data	-0.006	0.002	0.010
Estimation methods			
OLS	0.024	0.007	0.000
IV	0.056	0.007	0.000
Panel	0.019	0.008	0.016
MLE	0.022	0.009	0.020
Design of the analysis			
Africa	0.070	0.007	0.000
Asia	0.012	0.006	0.065
Europe	0.008	0.006	0.190
America	0.028	0.007	0.000
USA	0.013	0.006	0.053
Male only	-0.020	0.004	0.000
Female only	-0.023	0.004	0.000
Dependent variable specification			
Dependent: annual	0.007	0.004	0.046
Dependent: monthly	-0.009	0.005	0.053
Dependent: weekly	0.013	0.007	0.076
Dependent: hourly	0.011	0.004	0.007
Additional controls			
Cognitive ability	-0.011	0.003	0.000
Sector	-0.007	0.006	0.249
Weight	0.003	0.004	0.526
Social class	0.010	0.012	0.401
Marital status	-0.005	0.004	0.152
Parental	-0.003	0.003	0.233
Gender	-0.021	0.005	0.000
Ethnicity	0.005	0.002	0.037
Job type	-0.008	0.003	0.009
Employment type	0.007	0.004	0.088
Publication characteristics			
Publication year	0.005	0.004	0.216
Citations	-0.002	0.001	0.016
Peer-reviewed	0.014	0.004	0.001
Impact factor	-0.001	0.002	0.563
Intercept	-0.047	0.016	0.004

Note: The table displays the results of the FMA model specification. For optimal weights selection Mallows criterion is employed (Hansen 2007). Also, we use orthogonalization which reduces the number of models that need to be estimated from 2^K to K (Amini & Parmeter 2012). The explanatory variables are described in Table 5.1.

Table 5.4: Subjectively predicted best practice estimate

	predicted estimate	95% confidence interval	
All	0.036	0.006	0.066
Africa	0.103	0.058	0.147
America	0.063	0.034	0.092
Male only	0.028	-0.004	0.059
Female only	0.019	-0.015	0.054

Note: The table presents the subjectively determined best practice estimates of the height premium effects. The selection of specific variables used is explained in Section 5.4. OLS is used for the estimation of the 95% confidence interval approximation. Standard errors are clustered at the study level.

We observe approximately two times and three times higher height premium estimates for America and Africa, respectively. Moreover, studies that perform their analysis on a subsample of men have higher estimates of height premium compared to studies utilizing subsamples of women. Overall, the estimates differ both across countries and genders.

As a final remark, we would like to point out that the results of the best-practice method should be regarded with caution, as the best-practice method is a tool that depends on the choice of the author's ideal study. Also, the definition of best practice is highly subjective.

Chapter 6

Conclusion

To summarize, this thesis deals with the meta-analysis of the relationship between height and income, also referred to as height premium. We clarified that height is a crucial physical feature of individuals impacting not only the level of income but also other aspects of their lives. Taller people are considered to be more attractive (Harrison & Saeed 1977; Freedman 1979), more persuasive (Young & French 1996) and are naturally elected as leaders (Stogdill 1948). Height is also associated with health outcomes (Fogel 1994), overall social status, social esteem and self-esteem (Judge & Cable 2004; Stulp *et al.* 2015).

The objective of the performed meta-analysis is to provide a quantitative overview of the height premium literature while focusing on two main topics - publication bias and heterogeneity. To be more specific, we conduct visual and as well as more formally established empirical publication bias tests to draw conclusions about the presence of publication bias in this particular research area. The analysis is furthermore complemented with the identification of factors that help to explain the heterogeneity among the estimates of the height-income relationship. The meta-regression methodology is applied to a dataset consisting of 1084 effects of the collected returns to height estimates detailed in 67 studies.

The contribution of this thesis is threefold. First, to the author's best knowledge, no meta-analysis of the impact of height on an individual's income has been performed as of yet. Therefore this work proves to be an important missing piece supplied to the academic literature on wage determinants. Second, the principle of the synthesis of estimates meta-analysis is built upon allows us to collect height premium effects from tens of studies. The comprehensive nature of such a dataset enables us to examine whether the authors are inclined to report only effects of certain direction and subsequently estimate the true effect which is cleared of the publication bias. Also, by addressing the differences in methodologies, data, or study design in the heterogeneity part of our meta-analysis, we were able to identify factors explaining the variation of returns to height across individual studies. Third, there are also policy implications as our findings provide a basis for to this point not initiated public debate focused on height-based discrimination.

As explained by Havranek *et al.* (2022) publication bias testing is founded on the assumption of no correlation between the reported estimates (height premium effects in our case) and their standard errors. For the purpose of publication bias testing, we divide our dataset into two subsamples in accordance with the fact whether the primary study controls for the endogeneity of height. In other words, we deal with causal and noncausal impacts of height on income. The main findings are as follows.

Height premium literature exhibits a positive publication bias which is mainly modest or substantial. Hence, the empirical literature suffers from an under-representation of reported negative height-income relationship effects. Causal mean beyond bias estimates of height premium are consistently higher than their noncausal counterparts, notwithstanding the fact whether we employ linear publication bias tests, nonlinear publication bias tests or tests that control for endogeneity of standard errors. A similar pattern of persistence with respect to the effect's size applies to the noncausal subsample. The results constantly display true effects at the level of the simple mean or slightly below. Though, the noncausal effects are of small magnitude and close to zero, thus, almost negligible. On the other hand, the estimated true effects of causal associations are not inconsequential. The publication bias analysis results also imply that when compared to the simple mean, the causal mean beyond bias effects are generally lower. Although that is not always the case, especially when the endogeneity of standard errors is taken into account.

In the heterogeneity analysis, we are confronted with model uncertainty which can be overcome by the utilization of averaging techniques. In our setting, we employ Bayesian model averaging technique (BMA), Frequentist model averaging technique (FMA) and Frequentist check. First of all, the estimation results confirm that publication bias remains present across all three models even after we control for additional variables capturing the study heterogeneity. Therefore, we conclude that the existence of publication bias in height premium literature is robust. In addition, the heterogeneity analysis helps us with explaining the variability among the collected height premium estimates. The quantitative results of the BMA model suggest that the differences between the effects capturing the impact of height on income can be attributed to e.g. geographical location. Studies that conduct the analysis on data from Africa or America (excluding the USA) systematically report more positive estimates. Other factors that were identified to capture the heterogeneity of height premium estimates are the longitudinal nature of the dataset used, restricting the sample with respect to gender, monthly specification of the dependent variable, or inclusion of gender control into the regression model. All of those are associated with the systematic delivery of negative height premium estimates.

Apart from the contributions and results above, we would also like to denote a few limitations of this thesis. First, the FAT-PET regression models we used require the relationship between the estimate and standard error to be linear. Moreover, the estimates and their standard errors ought to be exogenous. The issue of linearity is solved via the adoption of non-linear publication bias tests (e.g. Stem-based method by Furukawa (2019), Weighted Average of Adequately Powered by Ioannidis *et al.* (2017), or Selection model of Andrews & Kasy (2019)). To account for the potential endogeneity between the collected height premium estimates and their standard errors, we apply the instrumental variable approach as well as p-value* approach.

Second, we are aware of the fact that we do not cover all the details associated with wage returns to height. However, the scope of this thesis is limited. Also, as this meta-analysis is a pioneering work on the grounds of height premium literature, we cannot profit from previously conducted research that could otherwise be used as a baseline reference material. Nevertheless, we believe that future research can build upon this work and focus in more detail on the heterogeneity part of the analysis. It could be interesting to see how the results change when for instance additional variables reflecting the age of an individual or specific parental controls (i.e. mother's level of education, father's height, mother's social class, father's income etc.) are incorporated. Another option is to be more restrictive in the selection criteria of papers considered for the analysis and e.g. focus exclusively on hourly or annual earnings specification of the dependent variable.

Bibliography

ABUEG, L. C., P. G. O. HUBILLA, F. M. LOZANO, P. I. L. VALDIVIESO, & M. M. S. VALENCIA (2020): "Penalties and some counterfactuals to beauty premium: evidence from a job search simulation experiment." *DLSU Business & Economics Review* **30(1)**: pp. 15–29.

VAN AERT, R. C. & M. VAN ASSEN (2021): "Correcting for publication bias in a meta-analysis with the p-uniform* method." *Manuscript submitted for publication Retrieved from: https://osfio/preprints/bitss/zqjr92018* .

AHN, R., T. H. KIM, & E. HAN (2019): "The moderation of obesity penalty on job market outcomes by employment efforts." *International journal of environmental research and public health* **16(16)**: p. 2974.

AMINI, S. M. & C. F. PARMETER (2012): "Comparison of model averaging techniques: Assessing growth determinants." *Journal of Applied Econometrics* **27(5)**: pp. 870–876.

ANDERSON, P. (2018): "â€tall and litheâ€™–the wage-height premium in the victorian and edwardian british railway industry." *Explorations in Economic History* **67**: pp. 152–162.

ANDREWS, I. & M. KASY (2019): "Identification of and correction for publication bias." *American Economic Review* **109(8)**: pp. 2766–94.

ANỲŽOVÁ, P. & P. MATĚJ (2018): "Beauty still matters: The role of attractiveness in labour market outcomes." *International Sociology* **33(3)**: pp. 269–291.

ASADULLAH, M. N. & S. XIAO (2020): "The changing pattern of wage returns to education in post-reform china." *Structural Change and Economic Dynamics* **53**: pp. 137–148.

Atal, J., H. Ñopo, & N. Winder (2009): "New century, old disparities: gender and ethnic wage gaps in latin america." .

Averett, S. & S. Korenman (1993): "The economic reality of the beauty myth."

Averett, S. L., L. M. Argys, & J. L. Kohn (2012): "Immigration, obesity and labor market outcomes in the uk." *IZA Journal of Migration* **1(1)**: pp. 1–19.

Bailey, J. (2013): "Who pays for obesity? evidence from health insurance benefit mandates." *Economics Letters* **121(2)**: pp. 287–289.

Bajzik, J., T. Havranek, Z. Irsova, & J. Schwarz (2020): "Estimating the armington elasticity: The importance of study design and publication bias." *Journal of International Economics* **127**: p. 103383.

Baker, M. & K. Cornelson (2019): "The tall and the short of the returns to height." *Technical report*, National Bureau of Economic Research.

Balcar, J. (2012): "Supply side wage determinants: Overview of empirical literature." *Review of Economic Perspectives* **12(4)**: pp. 207–222.

Bargain, O. & J. Zeidan (2017): "Stature, skills and adult life outcomes: Evidence from indonesia." *The Journal of Development Studies* **53(6)**: pp. 873–890.

Bayarri, M. J. & J. O. Berger (2004): "The interplay of bayesian and frequentist analysis." .

Behrman, J. & M. R. Rosenzweig (2001): "The returns to increasing body weight." *Available at SSRN 297919* .

Bishu, S. G. & M. G. Alkadry (2017): "A systematic review of the gender pay gap and factors that predict it." *Administration & Society* **49(1)**: pp. 65–104.

Blaker, N. M., I. Rompa, I. H. Dessing, A. F. Vriend, C. Herschberg, & M. Van Vugt (2013): "The height leadership advantage in men and women: Testing evolutionary psychology predictions about the perceptions of tall leaders." *Group Processes & Intergroup Relations* **16(1)**: pp. 17–27.

Bleakley, H., D. Costa, & A. Lleras-Muney (2014): "Health, education, and income in the united states, 1820–2000." In "Human Capital in History: The American Record," pp. 121–159. University of Chicago Press.

Böckerman, P., A. Bryson, J. Viinikainen, C. Hakulinen, L. Pulkki-Råback, O. Raitakari, & J. Pehkonen (2017a): "Biomarkers and long-term labour market outcomes: the case of creatine." *Journal of Economic Behavior & Organization* **142**: pp. 259–274.

Böckerman, P., E. Johansson, U. Kiiskinen, & M. Heliövaara (2010): "The relationship between physical work and the height premium: Finnish evidence." *Economics & Human Biology* **8(3)**: pp. 414–420.

Böckerman, P. & J. Vainiomäki (2013): "Stature and life-time labor market outcomes: Accounting for unobserved differences." *Labour Economics* **24**: pp. 86–96.

Böckerman, P., J. Viinikainen, J. Vainiomäki, M. Hintsanen, N. Pitkänen, T. Lehtimäki, J. Pehkonen, S. Rovio, & O. Raitakari (2017b): "Stature and long-term labor market outcomes: evidence using mendelian randomization." *Economics & Human Biology* **24**: pp. 18–29.

Bodenhorn, H., C. Moehling, & G. N. Price (2012): "Short criminals: Stature and crime in early america." *The Journal of Law and Economics* **55(2)**: pp. 393–419.

Bom, P. R. & H. Rachinger (2019): "A kinked meta-regression model for publication bias correction." *Research synthesis methods* **10(4)**: pp. 497–514.

Bonilla, R., F. Kiraly, & J. Wildman (2019): "Beauty premium and marriage premium in search equilibrium: theory and empirical test." *International Economic Review* **60(2)**: pp. 851–877.

Borland, J. & A. Leigh (2014): "Unpacking the beauty premium: What channels does it operate through, and has it changed over time?" *Economic Record* **90(288)**: pp. 17–32.

Bossavie, L., H. Alderman, J. Giles, & C. Mete (2017): "The effect of height on earnings: Is stature just a proxy for cognitive and non-cognitive skills?" *World Bank Policy Research Working Paper* **(8254)**.

Boström, G. & F. Diderichsen (1997): "Socioeconomic differentials in misclassification of height, weight and body mass index based on questionnaire data." *International journal of epidemiology* **26(4)**: pp. 860–866.

Breach, A. & Y. Li (2017): "Gender pay gap by ethnicity in britain." .

Brown, C. & P. W. Routon (2018): "On the distributional and evolutionary nature of the obesity wage penalty." *Economics & Human Biology* **28**: pp. 160–172.

Caliendo, M. & M. Gehrsitz (2016): "Obesity and the labor market: A fresh look at the weight penalty." *Economics & Human Biology* **23**: pp. 209–225.

Card, D. & A. B. Krueger (1995): "Time-series minimum-wage studies: a meta-analysis." *The American Economic Review* **85(2)**: pp. 238–243.

Carrieri, V. & M. De Paola (2012): "Height and subjective well-being in italy." *Economics & Human Biology* **10(3)**: pp. 289–298.

Case, A. & C. Paxson (2008a): "Height, health, and cognitive function at older ages." *American economic review* **98(2)**: pp. 463–467.

Case, A. & C. Paxson (2008b): "Stature and status: Height, ability, and labor market outcomes." *Journal of political Economy* **116(3)**: pp. 499–532.

Case, A. & C. Paxson (2010): "Causes and consequences of early-life health." *Demography* **47**: pp. S65–S85.

Case, A., C. Paxson, & M. Islam (2009): "Making sense of the labor market height premium: Evidence from the british household panel survey." *Economics letters* **102(3)**: pp. 174–176.

Chen, J. & F. Pastore (2021): "study hard and make progress every dayâ€ť: Updates on returns to education in china." *Technical report*, Global Labor Organization (GLO).

Chen, Q., C. Zhu, J. Wang, & G. Kong (2019): "Height and earnings: A mendelian randomization analysis from china." .

Choi, S.-j. (2020): "Height inequality and socioeconomic implications in korea: analysis of individuals born between 1890 and 1919." *Journal of biosocial science* **52(4)**: pp. 504–513.

Churchill, S. A. & V. Mishra (2018): "Returns to education in china: a meta-analysis." *Applied Economics* **50(54)**: pp. 5903–5919.

Cinnirella, F. & J. K. Winter (2009): "Size matters! body height and labor market discrimination: A cross-european analysis." .

Cipriani, G. P. & A. Zago (2011): "Productivity or discrimination? beauty and the exams." *Oxford Bulletin of Economics and Statistics* **73(3)**: pp. 428–447.

Clément, M., P. Levasseur, & S. Seetahul (2020): "Is excess weight penalised or rewarded in middle-income countriesâ€™ labour markets? comparative evidence from china, india and mexico." *Kyklos* **73(2)**: pp. 161–195.

Connolly, S., J. Micklewright, & S. Nickell (1992): "The occupational success of young men who left school at sixteen." *Oxford Economic Papers* **44(3)**: pp. 460–479.

Cortez, R. (2014): "Health and productivity in peru: An empirical analysis by gender and region1." *Invertir En Salud* p. 87.

Daly, A., X. Meng, A. Kawaguchi, & K. Mumford (2006): "The gender wage gap in four countries." *Economic Record* **82(257)**: pp. 165–176.

Deaton, A. (2007): "Height, health, and development." *Proceedings of the national academy of sciences* **104(33)**: pp. 13232–13237.

Dechter, E. K. (2015): "Physical appearance and earnings, hair color matters." *Labour Economics* **32**: pp. 15–26.

Denny, K. (2017): "Are the effects of height on well-being a tall tale?" *Journal of Happiness Studies* **18(5)**: pp. 1445–1458.

d'Hombres, B. & G. Brunello (2005): "Does obesity hurt your wages more in dublin than in madrid? evidence from echp." .

Dinda, S., P. Gangopadhyay, B. Chattopadhyay, H. Saiyed, M. Pal, & P. Bharati (2006): "Height, weight and earnings among coalminers in india." *Economics & Human Biology* **4(3)**: pp. 342–350.

Doorley, K. & E. Sierminska (2015): "Myth or fact? the beauty premium across the wage distribution in germany." *Economics Letters* **129**: pp. 29–34.

Doucouliagos, C. & T. D. Stanley (2013): "Are all economic facts greatly exaggerated? theory competition and selectivity." *Journal of Economic Surveys* **27(2)**: pp. 316–339.

Egger, M., G. D. Smith, M. Schneider, & C. Minder (1997): "Bias in meta-analysis detected by a simple, graphical test." *Bmj* **315(7109)**: pp. 629–634.

Ehrenbergerova, D., J. Bajzik, & T. Havranek (2023): "When does monetary policy sway house prices? a meta-analysis." *IMF Economic Review* **71(2)**: pp. 538–573.

Eicher, T. S., C. Papageorgiou, & A. E. Raftery (2011): "Default priors and predictive performance in bayesian model averaging, with application to growth determinants." *Journal of Applied Econometrics* **26(1)**: pp. 30–55.

Elu, J. U. & G. N. Price (2013): "Does ethnicity matter for access to childhood and adolescent health capital in china? evidence from the wage-height relationship in the 2006 china health and nutrition survey." *The Review of Black Political Economy* **40(3)**: pp. 315–339.

Eschker, E., S. J. Perez, & M. V. Siegler (2004): "The nba and the influx of international basketball players." *Applied Economics* **36(10)**: pp. 1009–1020.

Evans, T. (2020): "Ethnicity pay gaps in great britain 2019."

Fikkan, J. L. & E. D. Rothblum (2012): "Is fat a feminist issue? exploring the gendered nature of weight bias." *Sex Roles* **66(9)**: pp. 575–592.

Fletcher, J. M. (2009): "Beauty vs. brains: Early labor market outcomes of high school graduates." *Economics letters* **105(3)**: pp. 321–325.

Fogel, R. W. (1994): "Economic growth, population theory, and physiology: the bearing of long-term processes on the making of economic policy."

Foster, A. D. & M. R. Rosenzweig (1994): "A test for moral hazard in the labor market: Contractual arrangements, effort, and health." *The Review of Economics and Statistics* pp. 213–227.

Freedman, D. G. (1979): "Human sociobiology: A holistic approach." .

Frieze, I. H., J. E. Olson, & D. C. Good (1990): "Perceived and actual discrimination in the salaries of male and female managers 1." *Journal of Applied Social Psychology* **20(1)**: pp. 46–67.

Furukawa, C. (2019): "Publication bias under aggregation frictions: Theory, evidence, and a new correction method." *Evidence, and a New Correction Method (March 29, 2019)* .

Gao, W. & R. Smyth (2010): "Health human capital, height and wages in china." *The Journal of Development Studies* **46(3)**: pp. 466–484.

Gechert, S., T. Havránek, Z. Havránková, & D. Kolcunova (2019): "Death to the cobb-douglas production function." *Technical report*, IMK working paper.

Gechert, S., T. Havranek, Z. Irsova, & D. Kolcunova (2022): "Measuring capital-labor substitution: The importance of method choices and publication bias." *Review of Economic Dynamics* **45**: pp. 55–82.

George, E. I. (2010): "Dilution priors: Compensating for model space redundancy." In "Borrowing Strength: Theory Powering Applications–A Festschrift for Lawrence D. Brown," volume 6, pp. 158–166. Institute of Mathematical Statistics.

Gerard, F., L. Lagos, E. Severnini, & D. Card (2021): "Assortative matching or exclusionary hiring? the impact of employment and pay policies on racial wage differences in brazil." *American Economic Review* **111(10)**: pp. 3418–57.

Goldin, C. (2014): "A grand gender convergence: Its last chapter." *American Economic Review* **104(4)**: pp. 1091–1119.

Groot, W. & H. M. Van Den Brink (2000): "Overeducation in the labor market: a meta-analysis." *Economics of education review* **19(2)**: pp. 149–158.

Groothuis, P. A. & J. R. Hill (2013): "Pay discrimination, exit discrimination or both? another look at an old issue using nba data." *Journal of Sports Economics* **14(2)**: pp. 171–185.

Gu, T. & Y. Ji (2019): "Beauty premium in china's labor market: Is discrimination the main reason?" *China Economic Review* **57**: p. 101335.

GUNNELL, D., J. ROGERS, & P. DIEPPE (2001): "Height and health: predicting longevity from bone length in archaeological remains." *Journal of Epidemiology & Community Health* **55(7)**: pp. 505–507.

HAMERMESH, D. S. & J. BIDDLE (1993): "Beauty and the labor market."

HANSEN, B. E. (2007): "Least squares model averaging." *Econometrica* **75(4)**: pp. 1175–1189.

HARPER, B. (2000): "Beauty, stature and the labour market: A british cohort study." *Oxford Bulletin of Economics and Statistics* **62**: pp. 771–800.

HARPER, B. & M. HAQ (1997): "Occupational attainment of men in britain." *Oxford Economic Papers* **49(4)**: pp. 638–650.

HARRISON, A. A. & L. SAEED (1977): "Let's make a deal: An analysis of revelations and stipulations in lonely hearts advertisements." *Journal of Personality and Social Psychology* **35(4)**: p. 257.

HASAN, I., R. HORVATH, & J. MARES (2018): "What type of finance matters for growth? bayesian model averaging evidence." *The World Bank Economic Review* **32(2)**: pp. 383–409.

HAVRÁNEK, T. (2015): "Measuring intertemporal substitution: The importance of method choices and selective reporting." *Journal of the European Economic Association* **13(6)**: pp. 1180–1204.

HAVRANEK, T., R. HORVATH, Z. IRSOVA, & M. RUSNAK (2015): "Cross-country heterogeneity in intertemporal substitution." *Journal of International Economics* **96(1)**: pp. 100–118.

HAVRANEK, T. & Z. IRSOVA (2011): "Estimating vertical spillovers from fdi: Why results vary and what the true effect is." *Journal of International Economics* **85(2)**: pp. 234–244.

HAVRANEK, T., Z. IRSOVA, L. LASLOPOVA, & O. ZEYNALOVA (2020): "The elasticity of substitution between skilled and unskilled labor: A meta-analysis." .

HAVRANEK, T., Z. IRSOVA, L. LASLOPOVA, & O. ZEYNALOVA (2021): "Skilled and unskilled labor are less substitutable than commonly thought." .

Havranek, T., Z. Irsova, L. Laslopova, & O. Zeynalova (2022): "Publication and attenuation biases in measuring skill substitution." .

Havranek, T., Z. Irsova, & T. Vlach (2018a): "Measuring the income elasticity of water demand: the importance of publication and endogeneity biases." *Land Economics* **94(2)**: pp. 259–283.

Havranek, T., Z. Irsova, & O. Zeynalova (2018b): "Tuition fees and university enrolment: a meta-regression analysis." *Oxford Bulletin of Economics and Statistics* **80(6)**: pp. 1145–1184.

Havranek, T., M. Rusnak, & A. Sokolova (2017): "Habit formation in consumption: A meta-analysis." *European Economic Review* **95**: pp. 142–167.

Havránek, T., T. D. Stanley, H. Doucouliagos, P. Bom, J. Geyer-Klingeberg, I. Iwasaki, W. R. Reed, K. Rost, & R. C. van Aert (2020): "Reporting guidelines for meta-analysis in economics." *Journal of Economic Surveys* **34(3)**: pp. 469–475.

Hedges, L. V. (1992): "Modeling publication selection effects in meta-analysis." *Statistical Science* **7(2)**: pp. 246–255.

Heineck, G. (2005): "Up in the skies? the relationship between body height and earnings in germany." *Labour* **19(3)**: pp. 469–489.

Heineck, G. (2008): "A note on the height–wage differential in the uk–cross-sectional evidence from the bhps." *Economics Letters* **98(3)**: pp. 288–293.

Heineck, G. (2009): "Too tall to be smart? the relationship between height and cognitive abilities." *Economics Letters* **105(1)**: pp. 78–80.

Hersch, J. (2008): "Profiling the new immigrant worker: The effects of skin color and height." *Journal of Labor Economics* **26(2)**: pp. 345–386.

Higham, P. A. & D. W. Carment (1992): "The rise and fall of politicians: The judged heights of broadbent, mulroney and turner before and after the 1988 canadian federal election." *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement* **24(3)**: p. 404.

Hill, J. R. (2004): "Pay discrimination in the nba revisited." *Quarterly Journal of Business and Economics* pp. 81–92.

Hitsch, G. J., A. Hortaçsu, & D. Ariely (2010): "What makes you click?â€"mate preferences in online dating." *Quantitative marketing and Economics* **8**: pp. 393–427.

Huang, J., H. M. Van den Brink, & W. Groot (2009): "A meta-analysis of the effect of education on social capital." *Economics of education review* **28(4)**: pp. 454–464.

Huang, W., X. Lei, G. Ridder, J. Strauss, & Y. Zhao (2013): "Health, height, height shrinkage, and ses at older ages: evidence from china." *American Economic Journal: Applied Economics* **5(2)**: pp. 86–121.

Hübler, O. (2009): "The nonlinear link between height and wages in germany, 1985–2004." *Economics & Human Biology* **7(2)**: pp. 191–199.

Hübler, O. (2016): "Height and wages." *Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199389292.013* **29**.

Ibragimova, N. & R. Salahodjaev (2020): "Does height matter for earnings? evidence from russia." *Journal of Biosocial Science* **52(6)**: pp. 885–894.

Ioannidis, J. P., T. D. Stanley, & H. Doucouliagos (2017): "The power of bias in economics research."

Iwasaki, I. & X. Ma (2020): "Gender wage gap in china: a large meta-analysis." *Journal for Labour Market Research* **54(1)**: pp. 1–19.

Jayachandran, S. (2015): "The roots of gender inequality in developing countries." *economics* **7(1)**: pp. 63–88.

Jeffreys, H. (1961): "Small corrections in the theory of surface waves." *Geophysical Journal International* **6(1)**: pp. 115–117.

Johansson, E., P. Böckerman, U. Kiiskinen, & M. Heliövaara (2009): "Obesity and labour market success in finland: The difference between having a high bmi and being fat." *Economics & Human Biology* **7(1)**: pp. 36–45.

Johnston, D. W. (2010): "Physical appearance and wages: Do blondes have more fun?" *Economics Letters* **108(1)**: pp. 10–12.

Judge, T. A. & D. M. Cable (2004): "The effect of physical height on workplace success and income: preliminary test of a theoretical model." *Journal of Applied Psychology* **89(3)**: p. 428.

Kanazawa, S. & M. C. Still (2018): "Is there really a beauty premium or an ugliness penalty on earnings?" *Journal of Business and Psychology* **33(2)**: pp. 249–262.

Kedir, A. M. (2008): "Health and wages: Panel evidence on men and women using iv quantile regression." .

Kedir, A. M. (2009): "Health and productivity: panel data evidence from ethiopia." *African Development Review* **21(1)**: pp. 59–72.

Kedir, A. M. *et al.* (2013): "Schooling, bmi, height and wages: Panel evidence on men and women." *Economic Issues* **18(2)**: pp. 1–18.

Khasnobis, P. & S. Dinda (2017): "Height differentiated wage premium in west bengal, india: An empirical study." .

Kim, T. H. & E. Han (2017): "Height premium for job performance." *Economics & Human Biology* **26**: pp. 13–20.

Komlos, J. & B. E. Lauderdale (2007): "Underperformance in affluence: the remarkable relative decline in us heights in the second half of the 20th century." *Social Science Quarterly* **88(2)**: pp. 283–305.

Kortt, M. & A. Leigh (2010): "Does size matter in australia?" *Economic Record* **86(272)**: pp. 71–83.

Kropfhäusser, F. (2016): "A fresh look at the labor market height premium in germany." *Economics Bulletin* **36(3)**: pp. 1376–1383.

Kuhn, P. & C. Weinberger (2005): "Leadership skills and wages." *Journal of Labor Economics* **23(3)**: pp. 395–436.

Kurtz, D. L. (1969): "Physical appearance and stature: Important variables in sales recruiting." *Personnel Journal* .

Lång, E. & P. Nystedt (2018): "Two by two, inch by inch: Height as an indicator of environmental conditions during childhood and its influence on earnings over the life cycle among twins." *Economics & Human Biology* **28**: pp. 53–66.

Lee, S. (2015): "Beauty pays but does investment in beauty?" *IZA World of labor* **(198)**.

Lee, W.-S. (2014): "Big and tall: Is there a height premium or obesity penalty in the labor market?" .

Lee, W.-S. & Z. Zhao (2017): "Height, weight and well-being for rural, urban and migrant workers in china." *Social indicators research* **132(1)**: pp. 117–136.

Lester, D. & D. Sheehan (1980): "Attitudes of supervisors toward short police officers." *Psychological reports* .

Li, M., M. d. C. Triana, S.-Y. Byun, & O. Chapa (2021): "Pay for beauty? a contingent perspective of ceo facial attractiveness on ceo compensation." *Human Resource Management* **60(6)**: pp. 843–862.

Lindqvist, E. (2012): "Height and leadership." *Review of Economics and statistics* **94(4)**: pp. 1191–1196.

Liu, E., S. Zhang *et al.* (2013): "A meta-analysis of the estimates of returns to schooling in china." *University of Houston Department of Economics Working Paper* **(201309855)**.

Loh, E. S. (1993): "The economic effects of physical appearance." *Social Science Quarterly* .

Longhi, S. (2020): "Does geographical location matter for ethnic wage gaps?" *Journal of Regional Science* **60(3)**: pp. 538–557.

Longhi, S. & M. Brynin (2017): "The ethnicity pay gap." *Equality and Human Rights Commission* .

Loureiro, P. R., A. Sachsida, & M. J. C. Mendonca (2010): "Links between physical appearance and wage discrimination: Further evidence." *International Review of Social Sciences and Humanities* **1(2)**: pp. 1–16.

Lundborg, P., P. Nystedt, & D.-O. Rooth (2009): "The height premium in earnings: the role of physical capacity and cognitive and non-cognitive skills." .

Lundborg, P., P. Nystedt, & D.-O. Rooth (2014): "Height and earnings: The role of cognitive and noncognitive skills." *Journal of Human Resources* **49(1)**: pp. 141–166.

Ma, X. & I. Iwasaki (2021): "Return to schooling in china: A large meta-analysis." *Education Economics* **29(4)**: pp. 379–410.

Mankiw, N. G. & M. Weinzierl (2010): "The optimal taxation of height: A case study of utilitarian income redistribution." *American Economic Journal: Economic Policy* **2(1)**: pp. 155–76.

Marianne, B. (2011): "New perspectives on gender." In "Handbook of labor economics," volume 4, pp. 1543–1590. Elsevier.

Matousek, J., T. Havranek, & Z. Irsova (2022): "Individual discount rates: a meta-analysis of experimental evidence." *Experimental Economics* **25(1)**: pp. 318–358.

McEvoy, B. P. & P. M. Visscher (2009): "Genetics of human height." *Economics & Human Biology* **7(3)**: pp. 294–306.

McGovern, M. E., A. Krishna, V. M. Aguayo, & S. Subramanian (2017): "A review of the evidence linking child stunting to economic outcomes." *International journal of epidemiology* **46(4)**: pp. 1171–1191.

Meléndez, L. V. *et al.* (2021): "Stylized facts about the gender wage gap: Evidences from a region of peru." *Turkish Journal of Computer and Mathematics Education (TURCOMAT)* **12(12)**: pp. 1340–1356.

Mitra, A. (2001): "Effects of physical attributes on the wages of males and females." *Applied Economics Letters* **8(11)**: pp. 731–735.

Mitra, A. (2003): "Access to supervisory jobs and the gender wage gap among professionals." *Journal of Economic Issues* **37(4)**: pp. 1023–1044.

Mobius, M. M. & T. S. Rosenblat (2006): "Why beauty matters." *American Economic Review* **96(1)**: pp. 222–235.

Moro, A., S. Tello-Trillo, & T. Tempesti (2019): "The impact of obesity on wages: The role of personal interactions and job selection." *Labour* **33(2)**: pp. 125–146.

Navarová, T. (2022): "International gender wage gap: A meta-analysis." .

Ñopo, H. (2009): "The gender wage gap in peru 1986-2000: Evidence from a matching comparisons approach." .

Octafia, T. P. & D. Setyonaluri (2022): "Beauty premium of working women in urban indonesia." *Makara Human Behavior Studies in Asia* **26(2)**: pp. 85–94.

Oostendorp, R. H. (2009): "Globalization and the gender wage gap." *The World Bank Economic Review* **23(1)**: pp. 141–161.

Oreffice, S. & C. Quintana-Domeque (2016): "Beauty, body size and wages: Evidence from a unique data set." *Economics & Human Biology* **22**: pp. 24–34.

Park, K. S. & I. Lee (2010): "Height premium in the korean labor market." *Journal of Labour Economics* **33(3)**: pp. 129–149.

Peng, L., X. Wang, & S. Ying (2020): "The heterogeneity of beauty premium in china: Evidence from cfps." *Economic Modelling* **90**: pp. 386–396.

Perkins, J. M., S. V. Subramanian, G. Davey Smith, & E. Özaltin (2016): "Adult height, nutrition, and population health." *Nutrition reviews* **74(3)**: pp. 149–165.

Persico, N., A. Postlewaite, & D. Silverman (2004): "The effect of adolescent experience on labor market outcomes: The case of height." *Journal of political Economy* **112(5)**: pp. 1019–1053.

Price, G. N. (2013): "The allometry of metabolism and stature: Worker fatigue and height in the tanzanian labor market." *Economics & Human Biology* **11(4)**: pp. 515–521.

Ranasinghe, P., M. N. A. Jayawardana, G. R. Constantine, M. R. Sheriff, D. R. Matthews, & P. Katulanda (2011): "Patterns and correlates of adult height in sri lanka." *Economics & human biology* **9(1)**: pp. 23–29.

Rashad, I. (2008): "Height, health, and income in the us, 1984–2005." *Economics & Human Biology* **6(1)**: pp. 108–126.

Ratner, B. (2009): “The correlation coefficient: Its values range between+ 1/- 1, or do they?” *Journal of targeting, measurement and analysis for marketing* **17(2)**: pp. 139–142.

Reddy, A. (2014): “Rural labour markets: Insights from indian villages.” *Asia-Pacific Development Journal* **21(1)**.

Reed, D. & J. Cheng (2003): *Racial and ethnic wage gaps in the California labor market*. Public Policy Institute of California San Francisco.

Ribero, R. (2000): “Earnings effects of household investment in health in colombia.” *Available at SSRN 217428* .

Rieger, M. (2015): “Risk aversion, time preference and health production: Theory and empirical evidence from cambodia.” *Economics & Human Biology* **17**: pp. 1–15.

Rietveld, C. A., J. Hessels, & P. van der Zwan (2014): “The stature of the self-employed and its premium.” .

Rietveld, C. A., J. Hessels, & P. van der Zwan (2015): “The stature of the self-employed and its relation with earnings and satisfaction.” *Economics & Human Biology* **17**: pp. 59–74.

Roberts, J. V. & C. P. Herman (2022): “The psychology of height: An empirical review.” *Physical appearance, stigma, and social behavior* pp. 113–140.

Rooth, D.-O. (2011): “Work out or out of workâ€”the labor market return to physical fitness and leisure sports activities.” *Labour Economics* **18(3)**: pp. 399–409.

Salahodjaev, R. & N. Ibragimova (2020): “Height and life satisfaction: evidence from russia.” *Applied Research in Quality of Life* **15(1)**: pp. 219–237.

Samaras, T. T., H. Elrick, & L. H. Storms (2003): “Is height related to longevity?” *Life sciences* **72(16)**: pp. 1781–1802.

Sargent, J. D. & D. G. Blanchflower (1994): “Obesity and stature in adolescence and earnings in young adulthood: analysis of a british birth cohort.” *Archives of pediatrics & adolescent medicine* **148(7)**: pp. 681–687.

Schallenkamp, K., R. DeBeaumont, & J. Houy (2012): "Weight-based discrimination in the workplace: Is legal protection necessary?" *Employee Responsibilities and Rights Journal* **24(4)**: pp. 251–259.

Schick, A. & R. H. Steckel (2015): "Height, human capital, and earnings: The contributions of cognitive and noncognitive ability." *Journal of Human Capital* **9(1)**: pp. 94–115.

Schultz, T. P. (2002): "Wage gains associated with height as a form of health human capital." *American Economic Review* **92(2)**: pp. 349–353.

Seguino, S. & C. Grown (2006): "Gender equity and globalization: Macroeconomic policy for developing countries." *Journal of International Development: The Journal of the Development Studies Association* **18(8)**: pp. 1081–1104.

Shin, E. J. (2022): "Representation and wage gaps in the planning profession: A focus on gender and race/ethnicity." *Journal of the American Planning Association* pp. 1–15.

Sierminska, E. (2015): "Does it pay to be beautiful?" *IZA World of Labor* .

Van der Sluis, J., M. Van Praag, & W. Vijverberg (2005): "Entrepreneurship selection and performance: A meta-analysis of the impact of education in developing economies." *The World Bank Economic Review* **19(2)**: pp. 225–261.

Sohn, K. (2015a): "The height premium in indonesia." *Economics & Human Biology* **16**: pp. 1–15.

Sohn, K. (2015b): "The value of male height in the marriage market." *Economics & Human Biology* **18**: pp. 110–124.

Sohn, K. (2016): "Height and happiness in a developing country." *Journal of Happiness Studies* **17(1)**: pp. 1–23.

Stanley, T. D. (2005): "Beyond publication bias." *Journal of economic surveys* **19(3)**: pp. 309–345.

Stanley, T. D. (2008): "Meta-regression methods for detecting and estimating empirical effects in the presence of publication selection." *Oxford Bulletin of Economics and statistics* **70(1)**: pp. 103–127.

Stanley, T. D., S. B. Jarrell, & H. Doucouliagos (2010): "Could it be better to discard 90% of the data? a statistical paradox." *The American Statistician* **64(1)**: pp. 70–77.

Steckel, R. H. (1995): "Stature and the standard of living." *Journal of economic literature* **33(4)**: pp. 1903–1940.

Sterne, J. A., B. J. Becker, & M. Egger (2005): "The funnel plot." *Publication bias in meta-analysis: Prevention, assessment and adjustments* pp. 73–98.

Stinebrickner, R., T. Stinebrickner, & P. Sullivan (2019): "Beauty, job tasks, and wages: A new conclusion about employer taste-based discrimination." *Review of Economics and Statistics* **101(4)**: pp. 602–615.

Stogdill, R. M. (1948): "Personal factors associated with leadership: A survey of the literature." *The Journal of psychology* **25(1)**: pp. 35–71.

Stulp, G., A. P. Buunk, S. Verhulst, & T. V. Pollet (2015): "Human height is positively related to interpersonal dominance in dyadic interactions." *PloS one* **10(2)**: p. e0117860.

Tanner, J. *et al.* (1978): "Foetus into man: physical growth and conception to maturity." *Foetus into man: physical growth and conception to maturity.* .

Tanner, J. M. & J. M. Tanner (1990): *Foetus into man: Physical growth from conception to maturity.* Harvard University Press.

Tao, H.-L. (2014): "Height, weight, and entry earnings of female graduates in taiwan." *Economics & Human Biology* **13**: pp. 85–98.

Thomas, D. & J. Strauss (1997): "Health and wages: Evidence on men and women in urban brazil." *Journal of econometrics* **77(1)**: pp. 159–185.

Tovar-García, E. D. (2021): "Participation in sports, physical exercise, and wage income: Evidence from russian longitudinal data." *German Journal of Exercise and Sport Research* **51(3)**: pp. 333–343.

Underdal, A. (1994): "Leadership theory." *International Multilateral Negotiation–Approaches to the Management of Complexity.-San Francisco* pp. 178–197.

Undurraga, E. A., L. Zebrowitz, D. T. Eisenberg, V. Reyes-García, T. B. S. Team, & R. A. Godoy (2012): "The perceived benefits of height: Strength, dominance, social concern, and knowledge among bolivian native amazonians." *PloS one* **7(5)**: p. e35391.

Vogl, T. S. (2014): "Height, skills, and labor market outcomes in mexico." *Journal of Development Economics* **107**: pp. 84–96.

Wang, J., Q. Chen, G. Chen, Y. Li, G. Kong, & C. Zhu (2020): "What is creating the height premium? new evidence from a mendelian randomization analysis in china." *Plos one* **15(4)**: p. e0230555.

Wang, L. & J. Shen (2017): "Examining the factors affecting personal income: An empirical study based on survey data in chinese cities." *Frontiers of Economics in China* **12(4)**.

Wang, S.-Y. (2015): "Statistical discrimination, productivity, and the height of immigrants." *ILR review* **68(3)**: pp. 529–557.

Weichselbaumer, D. & R. Winter-Ebmer (2005): "A meta-analysis of the international gender wage gap." *Journal of economic surveys* **19(3)**: pp. 479–511.

Yamamura, E., R. Smyth, & Y. Zhang (2015): "Decomposing the effect of height on income in china: The role of market and political channels." *Economics & Human Biology* **19**: pp. 62–74.

Yang, X., J. Gao, J.-H. Liu, & T. Zhou (2018): "Height conditions salary expectations: Evidence from large-scale data in china." *Physica A: Statistical Mechanics and its Applications* **501**: pp. 86–97.

Yimer, S. & O. Fantaw (2011): "The impacts of health and nutrition on wages in ethiopia." *African Journal of Business Management* **5(30)**: p. 12174.

Young, T. J. & L. A. French (1996): "Height and perceived competence of us presidents." *Perceptual and Motor Skills* **82(3)**: pp. 1002–1002.

Zehetmayer, M. (2013): "Health, market integration, and the urban height penalty in the us, 1847–1894." *Cliometrica* **7(2)**: pp. 161–187.

Zeugner, S. & M. Feldkircher (2015): "Bayesian model averaging employing fixed and flexible priors: The bms package for r." *Journal of Statistical Software* **68**: pp. 1–37.

Zheng, Y. (2022): "Does height affect wage? empirical evidence from china."
.

Appendix A

Overview of height premium studies in specific countries

Table A.1: Studies of height premium in particular countries

Country	*Author*
Australia	Kortt & Leigh (2010), Lee (2014), Lee & Zhao (2017)
Brazil	Schultz (2002)
China	Asadullah & Xiao (2020), Chen *et al.* (2019), Chen & Pastore (2021), Elu & Price (2013), Gao & Smyth (2010), Peng *et al.* (2020), Wang *et al.* (2020), Wang & Shen (2017), Yamamura *et al.* (2015), Yang *et al.* (2018), Zheng (2022)
Colombia	Ribero (2000)
Ethiopia	Kedir (2008), Kedir (2009), Yimer & Fantaw (2011)
Finland	Böckerman *et al.* (2010), Böckerman *et al.* (2017a), Böckerman *et al.* (2017b), Böckerman & Vainiomäki (2013), Johansson *et al.* (2009)
Germany	Heineck (2005), Hübler (2009), Kropfhäußer (2016), Oreffice & Quintana-Domeque (2016), Rietveld *et al.* (2014), Rietveld *et al.* (2015)
Ghana	Schultz (2002)
India	Bleakley *et al.* (2014), Khasnobis & Dinda (2017), , Reddy (2014)
Indonesia	Bargain & Zeidan (2017), Sohn (2015a), Sohn (2015b)
Korea	Kim & Han (2017), Park & Lee (2010)
Mexico	Vogl (2014)
Pakistan	Bossavie *et al.* (2017)
Russia	Ibragimova & Salahodjaev (2020)
Sweden	Lång & Nystedt (2018), Lindqvist (2012), Lundborg *et al.* (2009) Lundborg *et al.* (2014), Rooth (2011),
Taiwan	Tao (2014)
UK	Anderson (2018), Bonilla *et al.* (2019), Case *et al.* (2009), Case & Paxson (2008b), Case & Paxson (2010), Heineck (2008), Persico *et al.* (2004), Sargent & Blanchflower (1994), Schick & Steckel (2015), Wang (2015)
USA	Baker & Cornelson (2019), Behrman & Rosenzweig (2001), Bleakley *et al.* (2014), Case & Paxson (2008b), Case & Paxson (2010), Eschker *et al.* (2004), Groothuis & Hill (2013), Hersch (2008), Hill (2004), Hitsch *et al.* (2010), Johnston (2010), Kanazawa & Still (2018), Mitra (2001), Persico *et al.* (2004), Schultz (2002), Wang (2015)

Note: The table presents an overview of studies used in the analysis grouped by countries.

Appendix B

Other physical features in Mincer equation - elaborated

Gender and ethnicity premium

Gender and ethnicity-based discrimination are perhaps the most often discussed ones among the researchers as well as the wider public. According to EUROSTAT gender pay gap statistics in EU countries, in 2020 women's hourly wage was on average lower by 13 % when compared to men's hourly earnings. The highest gender pay gap was recorded in Latvia - 22.3 %, whereas in Luxembourg men and women were awarded almost identically with a pay gap of 0.7 %. Moreover, the statistics also show that the gender pay gap is lower for those who newly enter the labour market. As regards the sector or occupational specification, it is higher in the private sector and the financial and insurance areas.

In the last decades, the gender pay gap diminished considerably (Goldin 2014). However, the difference between salaries of men and women is still quite clearly observable (Mitra 2003; Huang *et al.* 2009; Ñopo 2009; Meléndez *et al.* 2021). As Balcar (2012) remarks, it ranges from smaller (3.9 % to 6 %) to greater values (over 20 %). Thus, the magnitude of the gap itself is quite wide and it may depend on many factors - choice of the sector (public or private - Bishu & Alkadry 2017; Iwasaki & Ma 2020), the relevance of the country's institution (Daly *et al.* 2006), certain regions (Iwasaki & Ma 2020), the country's openness to trade (Oostendorp 2009) or data restriction by targeting e.g. never-marrieds or specific occupation (Weichselbaumer & Winter-Ebmer 2005). This is also supported by Navarová (2022) who points out that when researchers pick a specific labour market group (e.g. full-time workers, fresh college graduates) instead of performing the analysis on the level of the whole market, then the estimated gender wage gap tends to be smaller.

Generally, the pay gap in developing countries has a habit of being higher than in developed countries. As the less advanced country develops and the educational or personal autonomy possibilities of its citizens improve, one could argue that gender inequality should consequently gradually decline. However, Jayachandran (2015) explains that in many poor countries (particularly in South Africa, the Middle East, and India), historically the society has developed peculiar cultural norms that heighten the prioritization of men. Seguino & Grown (2006) also add that even though in developing countries women's access to employment might have improved, they are mostly employed in short-term and low-paid (usually manufacturing) jobs which greatly contributes to gender inequality.

Nevertheless as reminded by Balcar (2012), the impacts of gender on wages should be regarded with apprehension and thoughtfulness. That is because sometimes their significance and estimated magnitude may suffer from bias stemming from the fact that in the analysis researchers may have not included variables correlated with gender - e.g. Marianne (2011) explains that certain psychological and socio-psychological factors (such as risk preferences, attitude towards competition or attitudes towards negotiation) are different for males and females and thus *may make some occupations more attractive to women and others more attractive to men.*

Based on the literature on the race and ethnicity pay gap, we can conclude that this physical feature impacts workers' wages considerably as well (Reed & Cheng 2003; Atal *et al.* 2009; Longhi 2020; Gerard *et al.* 2021; Shin 2022). Longhi & Brynin (2017) evaluate that among ethnic minorities living in the UK, Bangladeshi men and women experience the largest pay gap compared to White British employees - on average 20.2% as estimated by Evans (2020). But on the other hand, they also remark that this effect may be partially attributed to the fact that Bangladeshi workers are more probable to be performing low-paid jobs. Intriguingly, workers of Chinese ethnicity receive a wage premium compared to White British employees (Breach & Li 2017; Evans 2020). Again, similarly as with the gender pay gap, Balcar (2012) warns that the interpretation is not as straightforward as would seem - rigid labelling the coefficients from regression models as the magnitude of the racial discrimination effect would be a mistake as omitting skills or the role of immigrants in the sample can be accompanied by bias.

Beauty premium

As regards the beauty premium, the impact of looks on earnings has been proved by many (e.g. Hamermesh & Biddle 1993; Fletcher 2009; Borland & Leigh 2014; Anýžová & Matěj 2018; Peng *et al.* 2020; Li *et al.* 2021; Octafia & Setyonaluri 2022). Peng *et al.* (2020) estimate that good-looking workers receive on average beauty premium of 5.4%, while bad-looking employees earn on average 3.3% less. Similarly Harper (2000) concludes the penalty for plainness is 15% for men and 11% for women. Doorley & Sierminska (2015) state that beauty premium is lower for females (on average 2% to 4%) than for males (on average 5% to 7%) as supported by Hamermesh & Biddle (1993). However, Abueg *et al.* (2020) state that good-looking men earn less than good-looking women.

Most of the papers agree that beauty premiums can be identified uniformly among all types of occupations. This is contradicted by Stinebrickner *et al.* (2019) who found that attractive individuals are preferred only in positions requiring interpersonal interaction. Mobius & Rosenblat (2006) raise an interesting point - they suggest that because physically attractive workers are more confident, employers subconsciously perceive them as more competent and that consequently results in higher wages. On that note, choosing between general employer discrimination, occupation-specific effects, and productivity differences arising from customer discrimination, Harper (2000) provides robust evidence that the bulk of the differences in wages between attractive and unattractive individuals is attributable to employer discrimination. Moreover, he also adds that attractiveness plays an important role in the marriage market - attractive women are more likely to get married, while unattractive men may experience great difficulty in finding a suitable match.

As recalled by Dechter (2015) or Johnston (2010), hair colour matters as well. Dechter (2015) focuses on so-called "blonde myth" with the following results - typically, inexperienced blonde women earn less than workers of different hair colours but as the blonde female workers gain more experience over time, this trend reverses and eventually blonde women with more experience are paid better than non-blond workers. On the other hand, Johnston (2010) observes no changes in the blond premium over time, instead he figures that the blond premium of women is of similar magnitude as an additional year of education (approximately 7%).

Cipriani & Zago (2011) tried to discover whether the source of beauty-based wage differences is purely in discriminatory tastes of employers or due to the higher productivity of more handsome individuals with the conclusion of productivity-related discrimination. Sierminska (2015) recommends that policies attempting to mitigate appearance-based discrimination need to account for employer discrimination, customer discrimination, productivity, and occupational sorting - the channels through which beauty-based discrimination is realized.

Contradictory to the above-mentioned, Kanazawa & Still (2018) find a weak relationship between beauty and wages which disappears completely after controlling for health, intelligence, and personality factors. They argue that past studies did not account for those factors and thus incorrectly assigned their impact on wages under the term "beauty premium". They put an emphasis on the following: *more beautiful workers earn more, not because they are beautiful, but because they are healthier, more intelligent, and have a better personality.*

Weight premium

The impact of weight on individuals' wages is not crystal clear, the empirical findings are mixed. The majority of authors agree on wage penalties for obese or overweight women (e.g. Averett & Korenman 1993; Sargent & Blanchflower 1994; Harper 2000; Fikkan & Rothblum 2012; Caliendo & Gehrsitz 2016; Ahn *et al.* 2019; Moro *et al.* 2019) or obese workers (i.e. men or women) in general (Schallenkamp *et al.* 2012). Obese immigrants also receive a wage penalty, no matter their gender (Averett *et al.* 2012). Bailey (2013) argues that awarding a wage penalty to obese people is not fair, as they have higher health costs that are supposed to be paid by insurers or employers, but instead they are covered by the employees themselves in the form of lower wage.

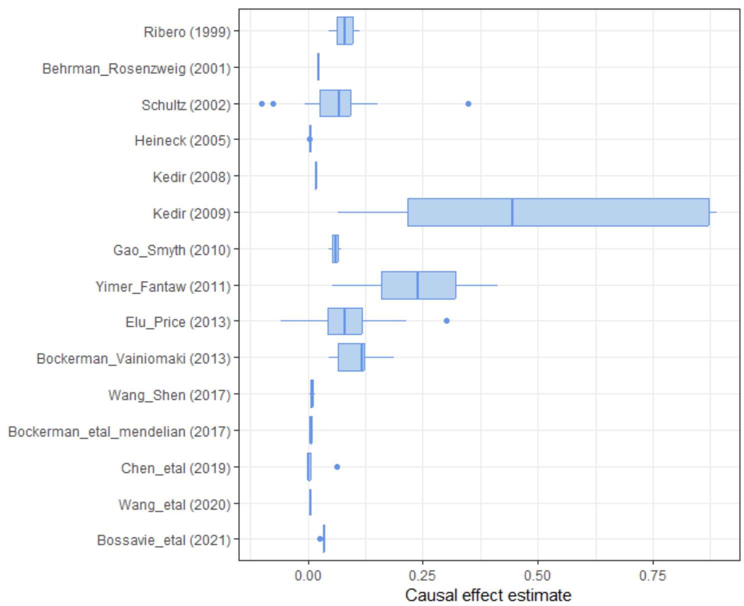
As regards the estimation techniques, Brown & Routon (2018) point out that OLS regression methods usually used by the researchers *tend to overstate obesity penalties for the lowest earners and understate obesity penalties for the highest earners* which is also supported by Caliendo & Gehrsitz (2016). Utilizing the applied fixed effects quantile regression models in their case provided the following results - underweight or overweight lowest earning individuals has no impact on their wage while obesity penalties are linked with higher quantiles of the wage distribution.

Interestingly enough, the literature on weight-based wage discrimination also provides evidence that the impact of individuals' weight on their salary can be positive, meaning overweight people receive wage premiums (are paid more). Such a relationship is in particular observed in India and Mexico, where the population chronically suffers from being underweight and overweight, respectively. But for example in China or Europe in general, obese workers deal with significant wage penalties (d'Hombres & Brunello 2005; Clément *et al.* 2020).

Appendix C

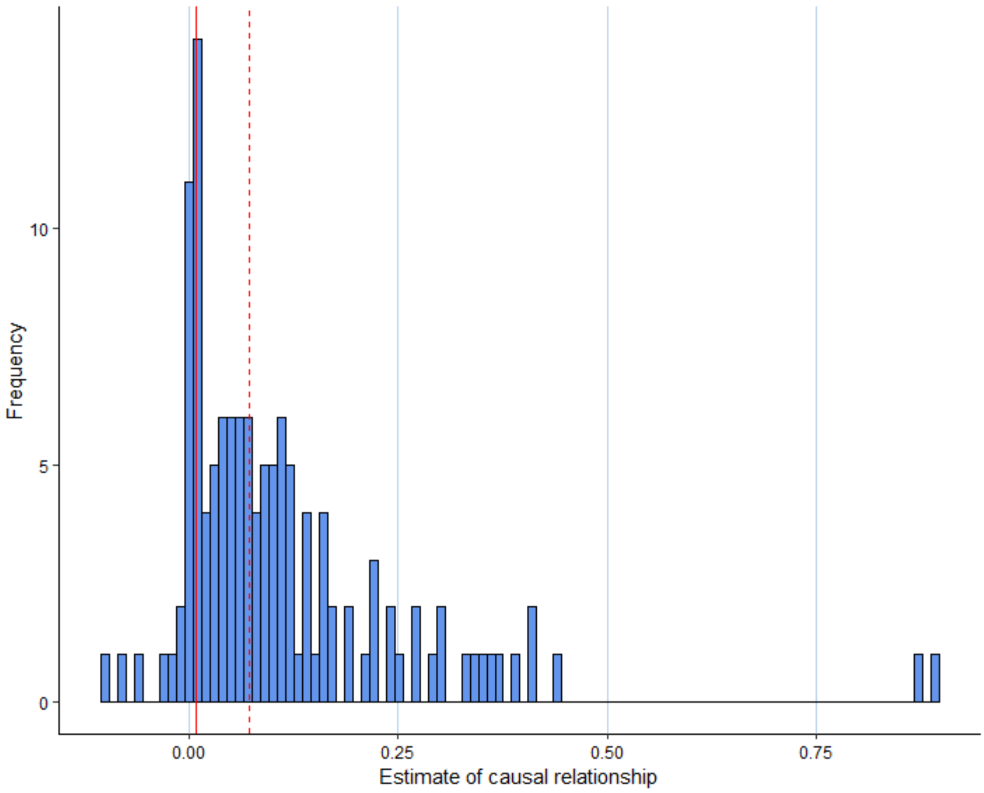
Additional characteristics of the causal and noncausal subsample of height premium estimates

Figure C.1: Variation of height premium estimates within and across studies - subsample of causal effects



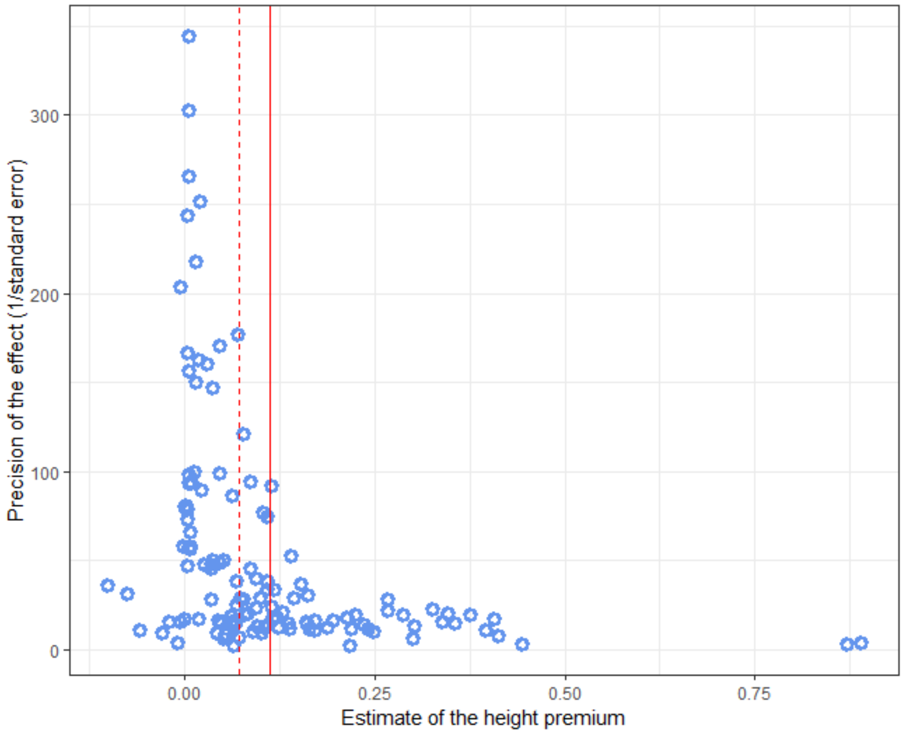
Note: The figure shows a boxplot of the collected estimates on the causal effect of height on income both within and across various studies. The boxes represent the interquartile range (i.e. the spread of the data between the 25th and 75th percentile sometimes also called the middle half of the data) and the solid line inside the box stands for median. The lower and upper whiskers illustrate the lowest and highest 25% of the data, respectively. The outliers are portrayed as dots outside the whiskers. Unwinsorized data are used.

Figure C.2: Histogram of the collected height premium effects for the subsample of causal effects



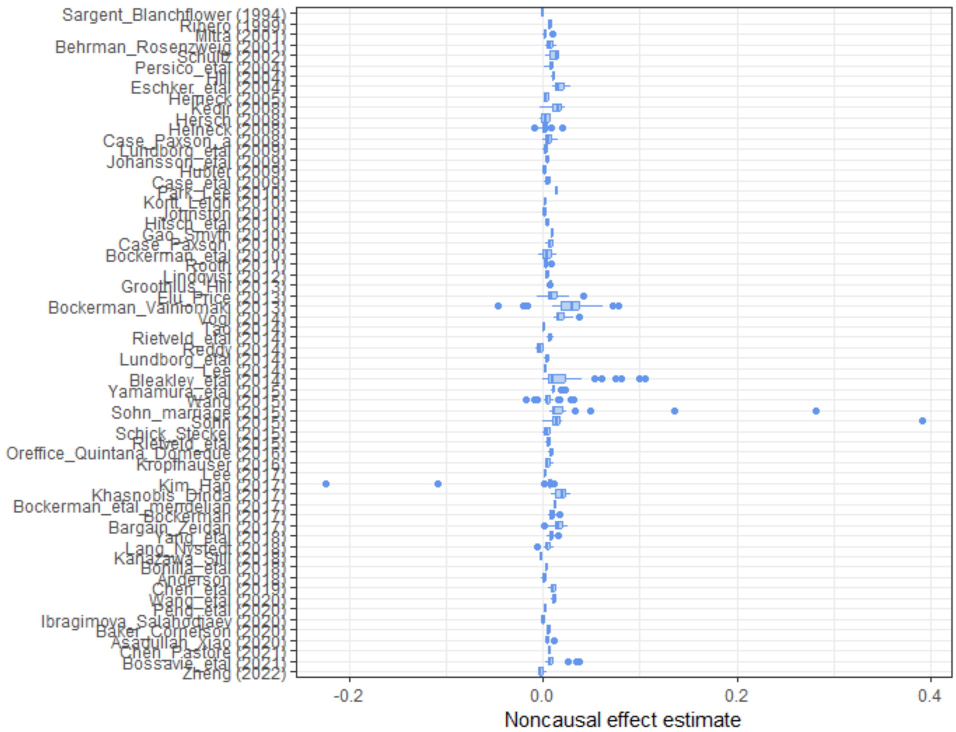
Note: The figure displays the distribution of the height premium causal estimates reported by the primary studies. The estimated effects are on the horizontal axis and their frequency is on the vertical axis. The solid line represents the mean, the dashed line represents the median. Unwinsorized data are used.

Figure C.3: Funnel plot of the collected effects of height premium for the causal subsample



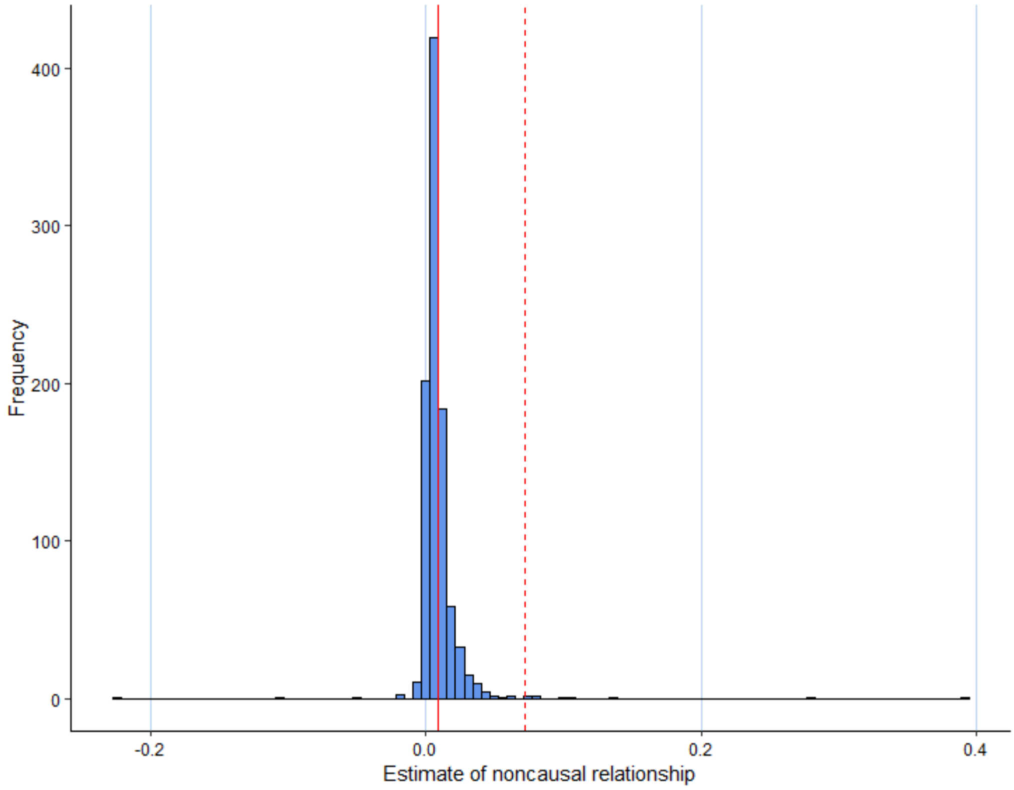
Note: The figure depicts a funnel plot of the causal estimates of the relationship between height and wage. The estimated effects are on the horizontal axis and their precision (1/standard error) is on the vertical axis. The solid line represents the mean, the dashed line represents the median. Unwinsorized data are used. However, for the empirical tests, winsorized data will be used. Outliers are excluded for ease of visualisation.

Figure C.4: Variation of height premium estimates within and across studies - subsample of noncausal effects



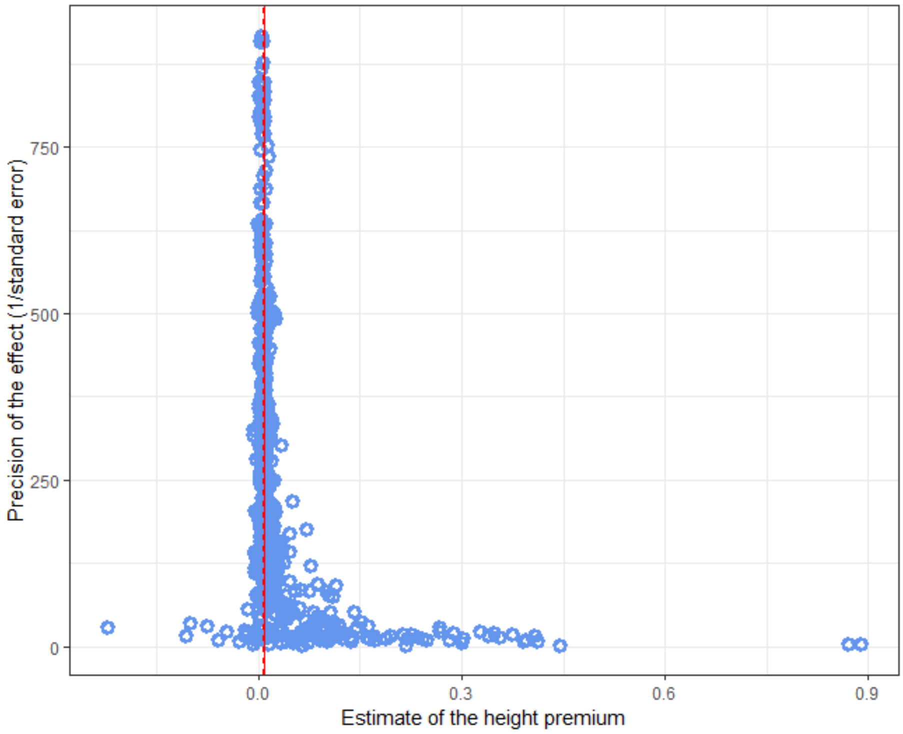
Note: The figure shows a boxplot of the collected estimates on the noncausal effect of height on income both within and across various studies. The boxes represent the interquartile range (i.e. the spread of the data between the 25th and 75th percentile sometimes also called the middle half of the data) and the solid line inside the box stands for median. The lower and upper whiskers illustrate the lowest and highest 25% of the data, respectively. The outliers are portrayed as dots outside the whiskers. Unwinsorized data are used.

Figure C.5: Histogram of the collected height premium effects for the subsample of noncausal effects



Note: The figure displays the distribution of the height premium noncausal estimates reported by the primary studies. The estimated effects are on the horizontal axis and their frequency is on the vertical axis. The solid line represents the mean, the dashed line represents the median. Unwinsorized data are used.

Figure C.6: Funnel plot of the collected effects of height premium for the noncausal subsample

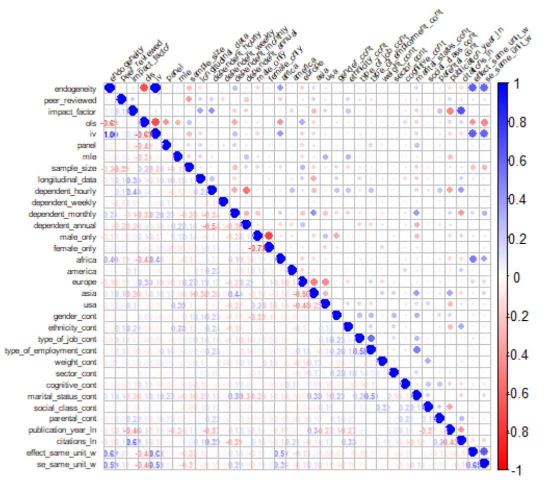


Note: The figure depicts a funnel plot of the noncausal estimates of the relationship between height and wage. The estimated effects are on the horizontal axis and their precision (1/standard error) is on the vertical axis. The solid line represents the mean, the dashed line represents the median. Unwinsorized data are used. However, for the empirical tests, winsorized data will be used. Outliers are excluded for ease of visualisation.

Appendix D

Diagnostics for BMA

Figure D.1: Correlation matrix of heterogeneity variables



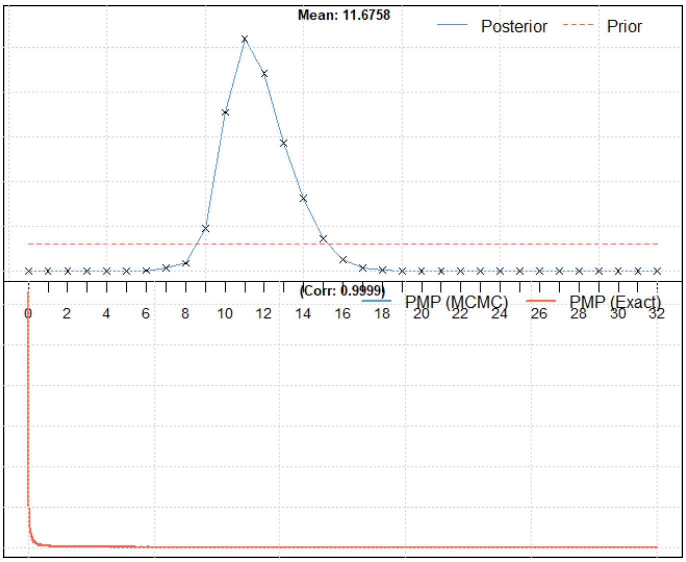
Note: The figure displays the correlation coefficients of variables used in the heterogeneity analysis. Variable *Endogeneity* is excluded due to high correlation. To account for the otherwise mild correlations, during the estimation we apply collinearity adjusted dilution model prior.

Table D.1: Summary statistics of the BMA model applied on the full sample

Mean no. regressors	*Draws*	*Burnins*
11.6758	3e+06	1e+06
Time	*No. models visited*	*Modelspace* 2^K
13.30924 mins	490796	4.3e+09
% visited	*% Topmodels*	*Corr PMP*
0.011	96	0.9999
No. Obs	*Model Prior*	*g-Prior*
1043	random / 16	UIP
Shrinkage-Stats		
Av=0.999		

Note: The table presents summary statistics of the BMA analysis. We use burn-ins = 1 million, draws = 3 million, unit information prior, and collinearity adjusted dilution model prior.

Figure D.2: Posterior model size and convergence for the BMA model applied on the full sample



Note: The figure shows posterior model size distribution (the top part) and posterior model probabilities (the bottom part) of the BMA analysis conducted in Chapter 5.