CHARLES UNIVERSITY
FACULTY OF SOCIAL SCIENCES

Institute of Economic Studies

How much does intelligence predict lifetime income? A Meta-Analysis

Bachelor's thesis

Author: Van Anh Nguyenová

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

Van Anh Nguyenova

Abstract

Despite growing interest and extensive empirical research in economic returns to ability, a consensus regarding the true impact of intelligence on financial outcomes remains elusive. While psychology literature has made efforts to unify divergent findings, economics is yet to produce a comprehensive meta-analysis addressing this issue. Addressing this gap, our thesis utilizes cutting-edge meta-analytic techniques to analyze a unique dataset of 765 estimates drawn from 38 studies, providing a clearer picture of intelligence's impact on income. We uncover a notable positive publication bias, which, after correction, yields a diminished yet statistically significant effect. Specifically, our results indicate that a standard deviation increase in cognitive ability results in a less than 10% increase in financial outcomes. Leveraging over 30 variables in our Bayesian and frequentist averaging models, we identify key determinants of this effect, including the data collection year, outcome specifications, methodological choices, country-specific factors, and the number of estimates reported per study. Additionally, when adjusting for factors such as gender, residential location, work experience, and family attributes, we observe substantial variations in effect size.

JEL Classification	J24, J31, D31, C11
Keywords	intelligence and income, returns to ability, meta-analysis, publication bias, Bayesian model averaging
Title	How much does intelligence predict lifetime income? A Meta-Analysis

Abstrakt

Navzdory rostoucímu zájmu a rozsáhlému empirickému výzkumu ekonomických výnosů schopností, dosáhnutí konsenzu o skutečném dopadu inteligence na finanční výsledky zůstává neuchopitelné. Zatímco psychologická literatura se snaží sjednotit rozdílné závěry, ekonomie ještě nepřinesla ucelenou metaanalýzu řešící tuto problematiku. Naše práce se snaží tuto mezeru vyplnit využitím nejmodernějších meta-analytických technik pro analýzu jedinečného datasetu obsahující 765 odhadů pocházejících ze 38 studií, čímž poskytuje jasnější obraz dopadu inteligence na příjem. Odhalujeme výraznou pozitivní publikační selektivitu, která po korekci vede k oslabenému, ale statisticky významnému efektu. Naše výsledky ukazují, že zvýšení kognitivních schopností o jednu standardní odchylku vede k něco méně než 10% nárůstu ve finančních výsledcích. Využitím více než 30 proměnných v našich bayesovských a frekventistických průměrovacích modelech identifikujeme klíčové determinanty tohoto efektu, včetně roku sběru dat, specifikací výsledné proměnné, metodologie, původ dat a počtu odhadů obdržené ze studie. Dále, po korekci faktorů, jako je pohlaví, místo bydliště, pracovní zkušenosti a rodinné atributy, pozorujeme významné variace ve velikosti efektu.

Klasifikace JEL J24, J31, D31, C11
Klíèová slova inteligence a příjem, výnosy schopností, metaanalýza, publikační selektivita, Bayesovské průměrování modelů
Název práce Nakolik vypovídá inteligence o celoživotních příjmech? Meta-analýza

Acknowledgments

The author wishes to express profound gratitude to doc. PhDr. Zuzana Havránková, Ph.D. for her invaluable guidance through the challenging nature of meta-analyses, her in-depth expertise and empathetic approach, and her detailed and insightful comments that were provided whenever needed. Similarly, the author is eternally grateful to Bc. Petr Čala, who generously helped with the coding aspect of the thesis, provided invaluable and prompt advice along the way. Finally, I would like to extend heartfelt appreciation to my family and friends for their unwavering support and encouragement throughout my studies.

Typeset in LaTeXusing the IES Thesis Template.

Bibliographic Record

Nguyenová, Van Anh: *How much does intelligence predict lifetime income? A Meta-Analysis.* Bachelor's thesis. Charles University, Faculty of Social Sciences, Institute of Economic Studies, Prague. 2023, pages 81. Advisor: doc. PhDr. Zuzana Havránková, Ph.D.

Contents

List of Tables

List of Figures

Acronyms

SES Socioeconomic Status

SE Standard Error

TIP Test in Print

IQ Intelligence Quotient

MA Mental Age

CA Chronological Age

SD Standard Deviation

CI Confidence Interval

FAT Funnel Asymmetry Test

OLS Ordinary Least Squares

RE Random-effects

BE Between-effects

FE Fixed-effects

WAAP Weighted Average of Adequately Powered

IV Instrumental Variable

BMA Bayesian Model Averaging

VIF Variance Inflation Factors

PIP Posterior Inclusion Probability

FMA Frequentist Model Averaging

Chapter 1

Introduction

As we navigate the rapid currents of our technologically advanced world of the 21st century, the mechanics of wealth and income accumulation are influenced by a complex web of determinants. Among these, intelligence emerges as a particularly compelling factor that has piqued the interest of economists and social scientists alike. The broad consensus on the positive correlation between intelligence and economic prosperity is well-documented in research spanning several decades (Gottfredson 1997; Schmidt & Hunter 2004; Strenze 2007). However, the intricacies of this relationship require further exploration. In this thesis, we delve into the magnitude of intelligence's impact on financial prosperity. The scholarly community presents divided views on this matter: some researchers posit that the intelligence-income effect is substantially stronger than the general findings in psychological studies (Schmidt & Hunter 2004), whereas others contend that the predictive capacity of intelligence for financial outcomes is not significant enough to warrant attention (Bowles & Gintis 2002).

Given the divergent estimates presented by individual studies on the correlation between intelligence and income, there have been several efforts to merge these findings via meta-analyses. The earliest of these in the realm of psychology was provided by Ng *et al.* (2005), who examined a range of predictors of career success, including cognitive ability. Strenze (2007) then conducted a more comprehensive meta-analysis focusing exclusively on longitudinal studies, thereby assessing the causal influence of intelligence on socioeconomic success. His results resembled those of Ng *et al.* (2005), showcasing a modest correlation and highlighting intelligence as a significant but not the sole determinant of success, with factors like parental socioeconomic status and academic performance also playing crucial roles. These studies, however, primarily explored zero-order correlations and did not delve into the marginal effect sizes offered by wage-equation literature (Mincer 1974). Bowles *et al.* (2001) in his review, introduced this aspect into the economics literature, revealing an effect size similar to ours – a standard deviation difference in cognitive performance was linked with less than a 10% wage increase. Another piece of evidence is highlighted by Ozawa *et al.* (2022), who examined both economic and educational returns of cognitive ability in low- and middle-income nations. Their findings indicated a slightly smaller effect, but given the limited studies this conclusion is drawn from, it might be influenced by publication bias.

To our knowledge, a comprehensive meta-analysis specifically addressing the economic returns of cognitive abilities is yet to be conducted. Prior meta-analytical studies are limited by their sample size and have not accounted for publication bias, which could significantly skew interpretation. Acknowledging this gap, our thesis aspires to synthesize findings from numerous studies spanning decades of research into an exhaustive systematic review, relying exclusively on economic literature and yielding 765 estimates from 38 unique studies. Our contributions to the literature are manifold: we estimate the true effect size, adjusting for the prevalent positive publication bias found in the existing literature. Taking an unbiased average of all statistically significant effect estimates, we arrive at a value of 0.052, which implies a 5.2% rise in income corresponding to each standard deviation increase in intelligence score.

Additionally, we delve into the heterogeneity in the literature, thereby identifying a variety of determinants that influence the effect magnitude, including control parameters, such as family attributes, country-specific factors, temporal factors, and methodological considerations. Setting our study apart, it is the first of its kind to explore marginal effects, address publication bias, and study heterogeneity employing the latest methods in economics. These encompass the distinguished selection model by Andrews & Kasy (2019), as well as Bayesian model averaging, which effectively manages model uncertainty. Consequently, our research enriches the field of economics by offering refined insights into the intelligence-income relationship. These findings could, in turn, serve to guide economic and social policies aimed at fostering equitable and sustainable wealth distribution.

Our thesis is organized as follows: Chapter 2 lays the theoretical groundwork for our topic, discussing the existing literature in depth and elucidating the interpretation of the key terms—intelligence and financial returns. In Chapter 3, we detail the data collection process and the transformations necessary to standardize reported estimates to a common metric and present a descriptive statistical analysis of our sample. Chapter 4 focuses on assessing potential publication bias within our data, making necessary corrections, and interpreting the resulting findings. Chapter 5 provides a comprehensive overview of the variables used in the study, and addresses variations in study designs through the application of model averaging techniques. Following this, we share our findings and engage in a discussion of the factors contributing to heterogeneity amongst the estimates at hand. Lastly, Chapter 6 encapsulates our findings and acknowledges the inherent limitations of this meta-analysis.

Chapter 2

Effect of Intelligence on Income

2.1 Motivation for the Effect

Studying the effect of intelligence on personal income can have a profound impact on the field of economics for several compelling reasons. Firstly, gaining insights into this effect can offer valuable perspectives on income disparities and socioeconomic inequality, empowering policymakers to formulate evidence-based strategies aimed at reducing income gaps and fostering equal opportunities. Secondly, the work of Schmidt & Hunter (2004) reveals a robust positive correlation between general intelligence and job performance, with coefficients ranging from 0.31 to 0.73. This finding implies that higher intelligence levels can potentially lead to increased income levels through enhanced task performance. By leveraging this knowledge, employers can establish tailored recruitment processes to attract and select highly competent individuals, enhancing overall organizational efficiency and productivity. Moreover, a wealth of research conducted by various scholars has consistently demonstrated a substantial correlation between educational attainment and cognitive ability (Strenze 2007), emphasizing the vital role of offering high-quality educational opportunities. Therefore, educational systems can be accordingly designed to promote intellectual development from a young age and improve individuals' prospects of achieving financial success.

Overall, intelligence, as a complex construct, is widely recognized in psychology research as one of the most accurate predictors of a person's long-term success, including the likelihood of achieving higher educational attainment and occupational prestige (Bajema 1968; Cheng & Furnham 2012; Sorjonen *et al.* 2012). These achievements, in turn, have the potential to lead to increased personal income in adulthood. Extensive research consistently demonstrates that when the g-factor (also known as general intelligence, general mental ability, or general intelligence factor) is taken into account, other determinants lose their predictive power in determining life outcomes. This phenomenon is often reflected in research titles with the phrase "not much more than g" (Ree & Earles 1991; Ree *et al.* 1994; Olea & Ree 1994; Ree & Earles 2013). Recent studies, including the work of Ganzach & Patel (2018), have reaffirmed the influential role of general intelligence in predicting wages, further supporting the notion that there is little beyond the impact of g. These findings underscore the robust and pervasive nature of general intelligence in forecasting a wide range of life outcomes.

However, it is noteworthy that general ability alone does not determine economic success. Numerous other factors, such as education, experience, personality traits, and family background, also contribute significantly to an individual's earning potential (Heckman *et al.* 2006). Before delving into the empirical research on the specific effect we are examining, we will provide an overview of the concept of intelligence and financial returns, summarize the findings related to its impact on labor outcomes, and outline our contributions to the existing literature.

2.2 Intelligence

Since the earliest attempts to understand the meaning of intelligence among Greek philosophers, the definition of intelligence has undergone modifications over the years as a result of centuries of study and discussion, leading to substantial controversy among scientists from various fields. According to the general intelligence theory proposed by Spearman (1927), intelligence is considered a single, unitary construct that underlies all intellectual abilities, commonly referred to as the g-factor. On the other hand, Howard Gardner's theory of multiple intelligences suggests that intelligence consists of several distinct abilities, including linguistic, logical-mathematical, spatial, musical, bodily-kinesthetic, interpersonal, and intrapersonal intelligence (Gardner 2012). Another approach to defining intelligence, supported by a group of 52 specialists in the study of intelligence and related fields, is described by Gottfredson (1997, pg. 13):

> *"Intelligence is a very general mental capability that, among other things, involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience. It is not merely book*

> *learning, a narrow academic skill, or test-taking smarts. Rather it reflects a broader and deeper capability for comprehending our surroundings - 'catching on,' 'making sense' of things, or 'figuring out' what to do."*

Despite ongoing debate about the definition of intelligence, researchers have made significant progress in developing standardized intelligence tests that can measure cognitive abilities across a range of disciplines (Deary 2012). Since the beginning of the 20th century, there has been a trend in the idea of assessing intelligence through tests and scores. Consequently, numerous intelligence tests have been designed to measure general cognitive functioning, with at least 200 tests listed in the Test in Print (TIP) series under the 'Intelligence and General Aptitude' category (Sternberg & Kaufman 2011). In particular, the Stanford-Binet and Wechsler intelligence tests are the most well-known and extensively studied (Camara *et al.* 2000). The Henmon-Nelson, Lorge-Thorndike, Otis-Lennon, and Raven Progressive Matrices tests, among other widely used intelligence assessments, are also acknowledged as trustworthy measures of general mental ability. Furthermore, hiring managers often employ aptitude test batteries, such as the Armed Services Vocational Aptitude Battery or General Aptitude Test Battery, to evaluate candidates for specific job positions and assess their problem-solving skills to quickly acquire new information. Hence, these tests are also considered measures of general ability (Strenze 2007).

Now that we have listed some of the most common intelligence tests, an important question arises: what exactly do these tests measure? While people often refer to intelligence tests as 'IQ tests,' it is essential to note that the two terms do not carry the same meaning. IQ stands for intelligence quotient, which originated from early intelligence tests like the Stanford-Binet test. Initially, an IQ score indicated whether a person's performance on the test aligned with their Mental Age (MA) relative to their Chronological Age (CA). However, after realizing that this score is inaccurate when a person hits adulthood, and their mental development stabilizes as their age rises, a new way of obtaining intelligence scores was invented called deviation IQ. This type of score is now widely used in major IQ tests. Rather than comparing MA and CA, deviation IQ evaluates the discrepancy between an individual's performance on the test of intellectual abilities and the standard scores of a representative group within their age range (Sternberg & Kaufman 2011). After all, the most critical question for intelligence tests is whether they measure one's intelligence. It seems that psychologists cannot even agree on the answer to this question since its definition is already ambiguous. However, we can conclude that intelligence tests do measure something—more or less a person's samples of behavior.

As we have learned, the definition of intelligence and its assessment methods are susceptible to controversial debates among theorists, with some scholars even arguing against the existence of intelligence. Gould (1981), for example, criticized the notion of intelligence as a single, unitary trait that could be measured by a single score. However, Spearman (1904) made a significant contribution by studying the correlations between scores on different cognitive tasks given to individuals, revealing a positive correlation among them. Favorably, a meta-analysis of 460 data sets has further demonstrated that an increase in performance on one ability test tends to correspond with an increase in performance on another test (Carroll 1993; Kuncel *et al.* 2001; Ackerman *et al.* 2005). This phenomenon also extends to students' grades in various school subjects Deary *et al.* (2007). Moreover, longitudinal studies examining the same individuals who underwent the same or similar cognitive ability tests have shown that more competent individuals consistently perform better on subsequent testing occasions (Deary *et al.* 2000).

The primary purpose of providing an overview of intelligence research is to examine the significant implications it holds, particularly in terms of the outcomes resulting from individual differences in intelligence. Intelligence has the potential to influence various aspects of people's lives, including their health, occupation, academic performance, and labor market outcomes. Academic achievement, in particular, has been extensively explored in relation to intelligence, with multiple studies showcasing a positive correlation between intelligence and success across diverse subjects and educational levels (Deary *et al.* 2007; Stumm *et al.* 2011; Józsa *et al.* 2022). It is crucial, however, to emphasize that intelligence is distinct from knowledge or academic skills. Rather than being tied to specific information or abilities already acquired, intelligence is concerned with an individual's capacity and potential to engage in mental tasks such as learning and comprehension (Furnham & Chamorro-Premuzic 2006). Furthermore, intelligence is not limited to any particular field or activity but represents a general mental ability applicable across various domains.

Continuing the discussion on intelligence and its impact on life outcomes is an extensive topic that goes beyond the scope of this thesis. For a more comprehensive review, we recommend referring to the book by Sternberg & Kaufman (2011).

2.3 Financial Returns

In this thesis, our key objective is to investigate the financial returns of an individual's intelligence. To ensure clarity and reader alignment, it is crucial to specify the particular financial outcomes we will be focusing on. These different measures of individual economic resources can be categorized into four types: wages and salaries, earnings, income, and wealth. Following widely recognized economic glossaries, such as the Oxford Dictionary of Economics, we define wages and salaries as the compensation an employee receives for their labor, usually measured on an hourly or monthly basis. Earnings, in addition to wages and salaries, also incorporate other monetary sources from employment such as bonuses and profits from self-employment.

Income, however, is a more comprehensive term that encompasses all monetary resources received during a specified period. This includes not only wages and self-employment earnings but also investment income, rental and property income, interest and dividends, retirement benefits, and essentially any transfers a person might receive. Wealth, on the other hand, represents the total of an individual's assets, excluding their liabilities. Assets can include savings, stocks, bonds, real estate, etc., while liabilities refer to various sorts of debt, such as loans and mortgages.

In this thesis, our primary focus will be on personal wages, earnings, and income. It is important to note that our attention is directed towards an individual's personal income rather than income at the family or household level. Moreover, we exclude wealth because it can be more influenced by other factors, such as familial wealth and socioeconomic status, as it often includes elements of inheritance. The motivation behind this is our intention to examine the economic returns to ability, which are more effectively captured by personal income. Moving forward, our focus will shift to—the fundamental objective of the present thesis—examining the predictive power of intelligence concerning financial outcomes.

2.4 Previous Reviews

The relationship between intelligence and lifetime income has captivated the interest of many theorists, prompting extensive research in the field of psychology. This has led to the publication of several meta-analyses dedicated to exploring the effect of general mental ability on personal income, although the number of such studies may be fewer than expected. Notable systematic reviews on this topic have been conducted by Bowles *et al.* (2001), Ng *et al.* (2005), Strenze (2007), and Ozawa *et al.* (2022). In this section, we will delve into these published reviews to extract their key insights and findings.

The meta-studies conducted by Ng *et al.* (2005) and Strenze (2007) primarily contribute to the field of psychology research as they utilize zero-order correlations. These correlation coefficients capture the magnitude and direction of the association between two variables without considering the impact of other important factors. Consequently, they are not commonly reported in empirical economic research where economists emphasize the examination of marginal effects. It is also worth noting that relying solely on simple correlations can sometimes yield misleading information. Occasionally, apparent positive correlations may actually be negative, or vice versa, leading to inaccurate interpretations of the association.

In the meta-analysis conducted by Ng *et al.* (2005), the average correlation between cognitive ability and salary was found to be moderate, with a value of 0.27. This result was derived from a compilation of eight studies that included both cross-sectional and longitudinal data. In contrast, Strenze (2007) conducted a more extensive meta-study focused exclusively on longitudinal studies that measured individuals' intelligence at the maximum age of 25 and career success at the minimum age of 20. The review gathered 253 estimates of the relationship between intelligence and income, resulting in a weighted average correlation corrected for unreliability and dichotomization of 0.20. This value was based on 31 independent samples, providing a comprehensive assessment of the association between intelligence and income across different studies.

In economics research, intelligence takes on a different form, specifically as a variable in the Mincer wage equation developed by Mincer (1974). This well-known model is widely used in labor economics to estimate the impact of human capital variables on individuals' earnings, typically including education and work experience as standard factors. The equation allows for the inclusion of other variables, leading to various extensions and modifications over time. Regarding ability, Blackburn & Neumark (1993) brought attention to the previously neglected variable of cognition in the Mincer equation, highlighting its potential to introduce bias in estimating the returns to schooling and experience. Subsequent studies, such as Heckman & Vytlacil (2001) or Hanushek & Woessmann (2008), support this perspective and refer to it as ability bias. Therefore, the standard form of the Mincer equation that incorporates cognitive ability and corrects for ability bias is as follows:

$$log(w) = \hat{\beta}_0 + \hat{\beta}_1 schooling + \hat{\beta}_2 experience + \hat{\beta}_3 IQ + \hat{\boldsymbol{\beta}}\boldsymbol{X} + \epsilon \quad (2.1)$$

where $\log(w)$ refers to the logarithm of wages, *schooling* represents the years of schooling, *experience* represents the years of labor market experience, and IQ represents the intelligence score, which could be replaced by any variable indicating cognitive ability. $\boldsymbol{X}$ denotes a matrix of other control variables, which can include any other quadratic or interaction terms. $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$, and $\hat{\boldsymbol{\beta}}$ correspond to their respective coefficients to be estimated, where $\hat{\beta}_3$ is of primary interest. Finally, ϵ refers to the error term.

Some evidence from empirical economics research has been provided by Bowles *et al.* (2001) and Ozawa *et al.* (2022). The review conducted by Bowles *et al.* (2001) gathered 65 estimates from 24 studies to examine the marginal effect of intelligence on earnings. They determined a mean standardized regression coefficient of 0.07, implying a 7% wage increase with each standard deviation (SD) improvement in cognitive performance. Lastly, the most recent systematic review was conducted by Ozawa *et al.* (2022), who focused on low- and middle-income countries, with half of the studies conducted in China. Their meta-study on wage returns to cognitive ability utilized four studies and revealed that one SD increase in cognitive ability corresponds to a 4.5% increase in wages (95% CI 2.6%–9.6%). However, despite the findings above, a comprehensive meta-study studying the economic returns on cognitive ability is, to our knowledge, currently unavailable.

To better comprehend the implications of a one SD change in intelligence, we need to expound upon what an SD means in the context of IQ scores. An SD is a statistical metric employed to measure the degree of variability or dispersion around the mean within a dataset. With respect to the structure of widely used IQ tests, these typically follow a normal distribution pattern, characterized by a mean score of 100 and an SD of 15 (Neisser *et al.* 1996). This implies that a single SD generally represents a variation of 15 IQ points. As such, an IQ score of 85 indicates one SD below the mean, while a score of 115 signifies one SD above the mean. Additionally, Neisser *et al.* (1996) highlighted that about 68% of individuals fall within the IQ range of 85 to 115, signifying they are within one SD of the mean.

2.5 Our Contribution

While Blackburn & Neumark (1995) indeed recognized the significant influence of IQ on economic outcomes, other studies like Ashenfelter & Rouse (1999) and Barrett & Depinet (1991) presented contrasting findings, arguing against this effect's significance. Furthermore, the economics literature currently lacks a comprehensive meta-analysis exploring the relationship between cognitive ability and income. This absence emphasizes our responsibility to carry out this task and evaluate the magnitude of IQ's predictive power on economic outcomes. In terms of quantifying this effect, our approach will be similar to that adopted by Bowles *et al.* (2001) and Ozawa *et al.* (2022). Yet, we intend to enhance our analysis by enlarging our dataset to incorporate recent studies providing estimates of the effect of our interest.

As for the methodological choices, previous meta-analyses have not provided any examination of publication bias. Ozawa *et al.* (2022) only acknowledged the risk of bias due to a limited number of studies but did not offer any concrete findings on this matter. A strong positive correlation between IQ and financial success is seldom disputed because it is a widely accepted notion that intelligent individuals are typically more financially prosperous, while those with lower intelligence are at a disadvantage. Consequently, there may be hesitancy to report negative correlations. Our study aims to contribute meaningfully by utilizing advanced meta-analytic techniques to comprehensively examine and correct for any possible publication bias. Through this, we aim to gain a deeper understanding of the behavior of the effect concerning this trend.

Finally, we aspire to identify other potential determinants that might influence this effect. To accomplish this, we intend to employ model averaging methods, which effectively manage model uncertainty, to explore the heterogeneity evident in the literature due to variations in study designs. While Bowles *et al.* (2001) briefly discussed heterogeneity and Strenze (2007) conducted a moderator analysis to study the influence of certain age and time-related factors on the correlation between general ability and socioeconomic success, no previous meta-analysis has thoroughly investigated heterogeneity or appropriately handled model uncertainty. By addressing these methodological gaps, our goal is to present a more comprehensive and robust analysis, as well as identify the most significant factors that impact this relationship.

Chapter 3

Data

To investigate the impact of intelligence levels on labor market outcomes, we employed a meta-analytic approach with quantitative analysis. This method provided us with a comprehensive understanding by synthesizing findings from multiple sources and enabled us to estimate the size and relevance of these effects with greater confidence. Our analysis encompassed 765 effect estimates extracted from 38 research papers, referred to as 'primary studies' in the context of meta-analysis, published within the last three decades. In this chapter, we will outline the data collection process, the criteria for study inclusion, the adjustments made to the estimates, and present summary statistics for our dataset.

3.1 Data Collection

The first step is to construct a suitable search query in Google Scholar. We prefer Google Scholar over other web search engines due to its comprehensive coverage of scholarly materials and accessibility to full-text articles. To determine the most relevant search query, we begin by identifying influential research papers on our topic so that these studies appear among the top search results. We iteratively modified the query by including different keywords such as 'intelligence,' 'cognitive ability,' 'income,' and 'wage.' Consequently, we generated and explored multiple queries, but the combination of words that we find most relevant is as follows:

("intelligence" OR "IQ" OR "cognitive ability") AND ("wage" OR "income") AND ("returns to ability").

This specific Google Scholar search query yielded approximately 370 results. However, out of all the search queries we generated, we meticulously reviewed the abstracts of around 670 research papers to assess the presence of relevant estimates aligned with our research interest. In accordance with Stanley (2001), it is important not to overlook unpublished papers. Hence, apart from published articles, we also collected working papers and Ph.D. dissertations. Subsequently, we downloaded and categorized 207 potential studies. The search for primary studies was concluded in May 2023, followed by the application of our inclusion criteria to filter out unusable studies. The primary studies included in our final dataset adhere to the following criteria:

- The study reports an estimated effect between general intelligence, or a proxy of intelligence (typically cognitive ability), and personal income. The scope of the study should be limited to personal income and not family or household income.
- The estimated effect presented in the study represents a one standard deviation change in ability on personal income. Alternatively, the effect can be standardized using corresponding sample summary statistics (further explanation will be provided in Section 3.2).
- The estimate is reported alongside the corresponding standard error or any other measure of uncertainty that allows for the calculation of the standard error, such as t-statistic or p-value.

Expanding upon the first criterion, we also decided to focus on general ability rather than its specific components, such as numeracy or literacy test scores alone (e.g. see the article by Chua (2017)). While it would have been interesting to explore how distinct abilities contribute to personal income, time constraints, and the extensive literature led us to exclude studies that reported effects of specific ability components. As a result, as mentioned earlier in this chapter, the criteria and limitations restricted our final sample to 38 primary studies, which collectively provide 765 estimates of returns to ability. A comprehensive list of the studies included in the meta-analysis can be found in Appendix A.

3.2 Data Adjustments

After completing the literature search, we thoroughly examined the filtered-out primary studies and collected estimates relevant to our research interest, specifically the estimated effects between general ability and financial outcomes. To ensure comparability among the estimates, we needed to determine the most appropriate common metric. For example, Strenze (2007) used zero-order correlations, which measure the simple association between two variables without accounting for the effects of other variables. Although correlation coefficients are widely used outside the economics field, they do not capture the marginal effect, which is the main interest of economists. This approach limits the analysis of heterogeneity and rarely provides measures of uncertainty. In assessing estimates linking intelligence and socioeconomic status, we found that a mere one of the 106 studies using correlations provided confidence intervals for these coefficients, specifically a paper by (Ioana Damian & Spengler 2021). Consequently, we decided to abandon the use of zero-order correlations given the nature of the methods employed in this thesis.

In our case, the standardized regression coefficient emerged as the preferred choice for two primary reasons. Firstly, standardized regression coefficients (also Beta coefficients) facilitate meaningful comparisons by providing a common scale, even when primary studies employ different measurement scales or units for the same independent variable. Secondly, these coefficients can be easily interpreted and synthesized across original studies. Typically, the effect sizes are collected from regression models in which the left side of the equation is in the logarithmic specification and regressed on ability score, resulting in semi-elasticities. These parameters represent the change in the natural logarithm of personal income associated with a one standard deviation change in intelligence or a proxy of intelligence. When multiplied by 100, the coefficient can be interpreted as the percentage change—a one standard deviation change in ability corresponds to a percentage change in personal income.

However, not all studies report standardized estimates. Unstandardized regression coefficients pose a problem due to their inability to be directly comparable across different studies. These coefficients represent a one-unit change in the independent variable, but the issue arises because ability tests are measured on different scales, resulting in varying effect magnitudes. There are two cases that can occur regarding the presentation of coefficients. In the first case, authors may present only an unstandardized coefficient, necessitating its conversion to a standardized one. Fortunately, if the original study provides appropriate summary statistics, these coefficients can be easily computed. When the dependent variable is in the logarithmic form, we only require the corresponding sample standard deviation of general ability scores, which is used to multiply the unstandardized coefficient. The same procedure is applied to its standard error, ensuring that the t-statistic remains unchanged, as the constant does not affect the precision of the original effect. The other case involves the reporting of both standardized and unstandardized coefficients. In such cases, authors commonly provide the unstandardized parameter along with its standard error, as well as the corresponding standardized coefficient (e.g., (Andersson & Bergman 2011)). In this scenario, we need to calculate the standard error of the standardized estimate using the following formula:

$$SE(\hat{\beta}) = \frac{SE(\hat{B})}{\frac{\hat{B}}{\hat{\beta}}} \tag{3.1}$$

Here, $\hat{\beta}$ denotes the estimated standardized effect, $\hat{B}$ denotes the unstandardized effect, and $SE(\hat{\beta})$ and $SE(\hat{B})$ represent their respective standard errors.

Some primary studies report the main effect as well as interaction effects. For example, we observe interaction effects between general ability and variables such as self-employment status (Hartog *et al.* 2010), parents' unemployment (Frøyland & Von Soest 2020), and predominantly experience (Cheung 2006; Pasche 2009; Galindo-Rueda 2003; Arcidiacono *et al.* 2010; Falch & Massih 2012; Zhang 2007). When the interaction term includes a dummy variable, we obtain two estimates: one when the dummy variable is equal to 0 and another when it equals 1. This situation arises in our first two examples, which involve determining whether a person's parents are unemployed and whether a person is an entrepreneur. When the interaction term involves two continuous variables, we utilize summary statistics and take the mean value of the variable included in the interaction term, typically the sample mean of years of experience in the workplace. To compute the effects that incorporate an interaction term, we apply the following formula:

$$\hat{\beta} = \hat{\beta}_l + \hat{\beta}_i \bar{x}_v \tag{3.2}$$

where $\hat{\beta}$ corresponds to the overall effect size of general ability taking into account the interaction term, $\hat{\beta}_l$ refers to the estimated linear effect of general ability, $\hat{\beta}_i$ refers to the estimate of the interaction term, and $\bar{x}_v$ denotes the sample mean value of the variable included in the interaction term. Additionally, for calculating the standard errors, we employ the delta method as follows:

$$SE(\hat{\beta}) = \sqrt{SE(\hat{\beta}_l)^2 + SE(\hat{\beta}_i)^2 \bar{x}_v} \tag{3.3}$$

where $SE(\hat{\beta})$, $SE(\hat{\beta}_l)$, and $SE(\hat{\beta}_i)$ indicate the standard errors of their respective Beta coefficients.

During the data collection process, we encountered a few challenging instances related to uncertainty measurement reports. These instances arise when standard errors or p-values are recorded as zero. Due to the minimal occurrence of these cases—only six instances—we decided to include them in our analysis. To account for these cases, we approximate the p-value to 0.0002 and calculate a corresponding t-statistic based on this assumption. As part of our final data transformation step, we employ a technique called winsorization, which aims to address potential outliers that may lead to misleading results in our dataset. To ensure the retention of all observations, we apply winsorization at the 1% level. This means that estimates falling below the 1$^{\text{st}}$ percentile or exceeding the 99$^{\text{th}}$ percentile are replaced with the closest non-extreme values available.

3.3 Summary Statistics

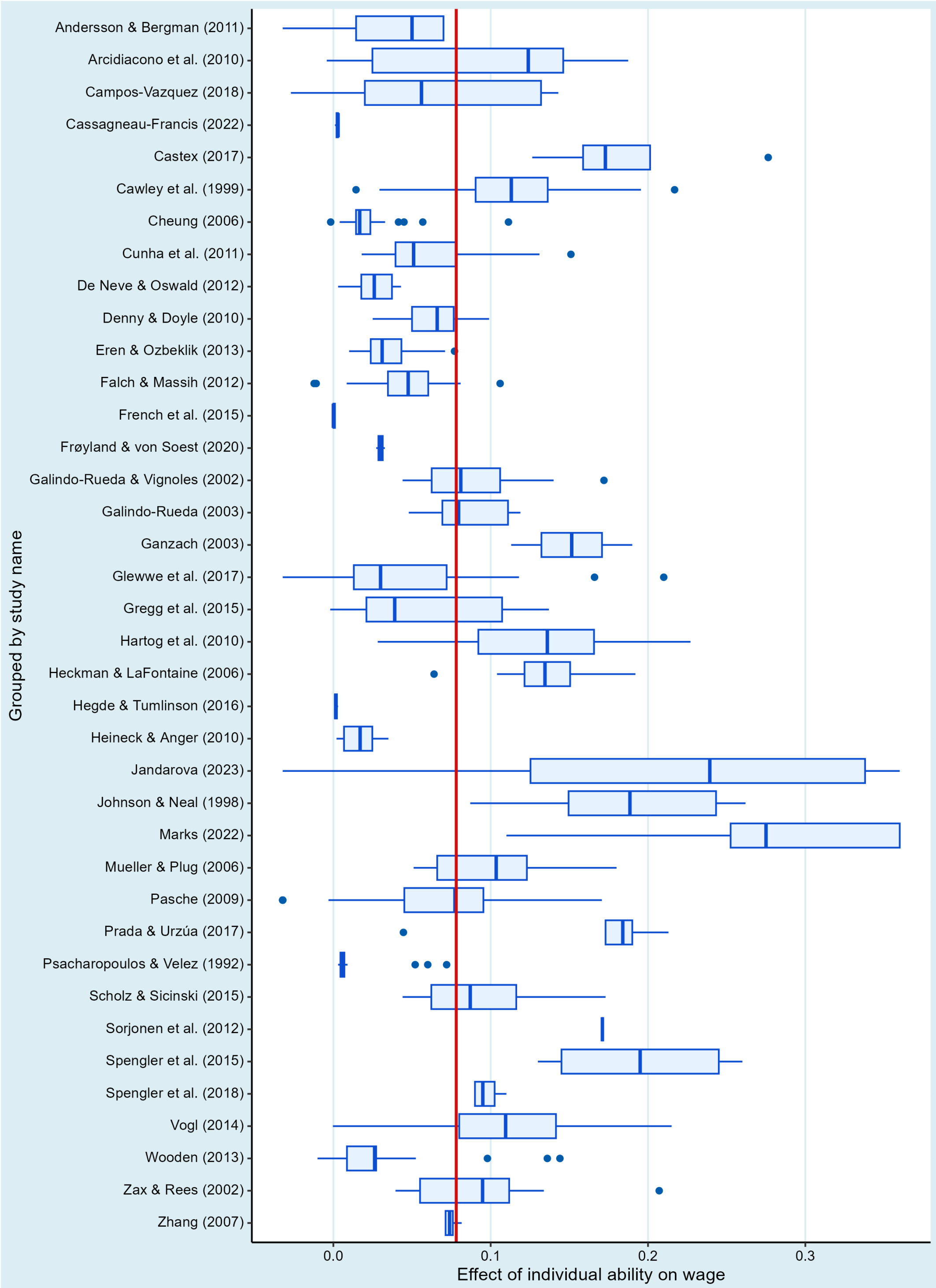
In Figure 3.1, we illustrate the variability of estimated effects both within individual primary studies and across different studies. One research study by Sorjonen *et al.* (2012) reports a single value of 0.171 as the total standardized effect, indicating a 17.1% change in income due to a single standard deviation change in ability. The study by French *et al.* (2015) exhibits the least variability in effect sizes among the studies included, presenting eight estimates within a range of 0.001. In contrast, other studies, notably the one conducted by Jandarova (2023), show a substantial variation in effect sizes, ranging from -0.06423 to 0.42672, covering nearly the entire horizontal axis. This study explores the impacts of parental job loss on children's outcomes, specifically how these impacts might vary based on the children's intelligence. Furthermore, in Appendix A, we include a box plot showcasing the effect sizes within and across different countries, suggesting a moderate variability among the estimates, with Luxembourg demonstrating the largest disparity.

Furthermore, we examine the mean values for different subsamples of our dataset, as presented in Table 3.1, to gain preliminary insights into the heterogeneity. We provide both simple and weighted mean values, with the latter calculated by dividing each value by the number of estimates reported by the respective study, offering a measure of enhanced reliability. The overall effect, characterized by a simple mean of 0.078, suggests a relatively modest impact of intelligence on financial outcomes, implying that a one standard deviation shift in intelligence score is associated with a 7.8% increase in income. From this point onward, we will interpret mean effects in this manner. Interestingly, weighted mean values predominantly yield slightly higher effects than their unweighted counterparts. For example, the weighted mean for the baseline effect stands at 0.09, 0.012 higher than the simple mean. These findings align closely with those reported by Bowles *et al.* (2001), who determined that a standard deviation increase in cognitive performance relates to less than a 10% increase in income, specifically identifying a mean standardized coefficient of 0.07 and a median of 0.08.

We want to draw attention to several intriguing discoveries. First, the mean effect estimates of the short-run and long-run studies clearly differ from one another, with the latter providing a higher estimate of 0.084 compared to the former's value of 0.052. This suggests that intelligence affects financial outcomes more significantly over the long-term perspective, possibly as a result of its early influence on educational attainment, professional success, and higher income levels over time. The weighted estimates exhibit the same trend but with a larger difference (0.045 & 0.097). Long-run estimates have a strong positive association with panel studies, as indicated by a correlation coefficient of 0.978, which is likely due to their limited presence in cross-sectional studies. Consequently, we observe a consistent pattern in the mean estimates for both cross-sectional and panel data subsets. For a thorough explanation of the variables, please refer to Section 5.1.

Secondly, we observed that commonly approved intelligence tests captured a higher ability-income effect compared to tests of cognitive abilities denoted as intelligence proxies. Subsets that utilized proper IQ tests yielded a mean effect estimate of 0.099, and a weighted mean effect of 0.102. On the other hand, samples measuring cognitive abilities, which represent the majority of the dataset, produced estimates of 0.071 and 0.084, respectively. It is also

Figure 3.1: Estimates both within and across studies



Notes: The figure illustrates a box plot displaying the estimated standardized coefficients that capture the effect of general ability on personal income in various studies. The length of the box represents the range of effect sizes. The bold (red) vertical line indicates the simple mean of all the estimates.

important to note that several 'Methodology' variables have a limited number of observations, which currently prevents us from drawing decisive conclusions about their impact for now.

Regarding subject characteristics, gender does not significantly impact the estimates. However, it is interesting to note that females generally demonstrate marginally higher economic returns to ability compared to males. Conversely, the mean age of the subjects, especially the age at which financial success is gauged, could potentially influence the magnitude of this effect. When the average age of study subjects is greater than the median age of the dataset (37 in this case), the mean effect estimate sees a rise to 0.089. This increase remains substantial, even after applying the weighing procedure, at 0.095. Distinguishing the dataset based on geographical origin, we observe that the mean effect for countries beyond Europe and the USA, which include Australia, China, Colombia, and Mexico, is considerably lower. This pattern persists for weighted estimates too, where countries other than Europe and the USA yield an average effect of 0.048, whereas the USA demonstrates, on average, a 10.6% increase in financial outcomes triggered by a one standard deviation change in ability.

Lastly, when considering publication characteristics, we observe that studies with a published status tend to yield higher positive ability-income effects. Interestingly, the weighted mean effects of publication status exhibit a notable difference, with estimates of 0.089 for published studies and 0.063 for unpublished studies. On the other hand, time variations, specifically the publication year, have a limited impact on the estimates. Whether a study's publication year is higher or lower than the dataset's median of 18 years, it does not significantly affect the results, indicating the consistency of the effect over time.

Table 3.1: Mean statistics across various subsets of data

	Unweighted			Weighted			
	Mean	95% CI		Mean	95% CI		N
All estimates	0.078	-0.063	0.219	0.090	-0.051	0.231	765
Data characteristics							
Short-run Estimate	0.052	-0.050	0.154	0.045	-0.057	0.147	143
Long-run Estimate	0.084	-0.063	0.231	0.097	-0.050	0.244	622
Cross-sectional Data	0.054	-0.048	0.156	0.047	-0.055	0.149	138
Panel Data	0.083	-0.064	0.230	0.096	-0.051	0.243	627
Micro-data	0.078	-0.051	0.207	0.094	-0.035	0.223	591
Survey Data	0.086	-0.108	0.280	0.075	-0.119	0.269	141
National Register Data	0.047	-0.018	0.112	0.082	0.017	0.147	33
Independent variable							
Intelligence	0.099	-0.073	0.271	0.102	-0.070	0.274	185
Intelligence Proxy	0.071	-0.056	0.198	0.084	-0.043	0.211	580
Methodology							
Method: OLS	0.070	-0.034	0.174	0.072	-0.032	0.176	468
Method: FE or RE	0.100	-0.074	0.274	0.076	-0.098	0.250	40
Method: QR	0.045	-0.041	0.131	0.053	-0.033	0.139	88
Method: IV	0.036	-0.056	0.128	0.047	-0.045	0.139	35
Method: Other	0.132	-0.080	0.344	0.132	-0.080	0.344	134
Subject characteristics							
Gender: Male ≥ 0.5	0.075	-0.058	0.208	0.088	-0.045	0.221	491
Gender: Male < 0.5	0.084	-0.069	0.237	0.093	-0.060	0.246	274
Gender: Female ≥ 0.5	0.083	-0.070	0.236	0.092	-0.061	0.245	279
Gender: Female < 0.5	0.075	-0.060	0.210	0.089	-0.046	0.224	486
Mean Age ≥ 37	0.089	-0.056	0.234	0.095	-0.050	0.240	454
Mean Age < 37	0.062	-0.067	0.191	0.081	-0.048	0.210	311
Country: Europe	0.085	-0.086	0.256	0.084	-0.087	0.255	195
Country: USA	0.097	-0.030	0.224	0.106	-0.021	0.233	385
Country: Other	0.032	-0.050	0.114	0.048	-0.034	0.130	185
Study characteristics							
Published	0.089	-0.046	0.224	0.098	-0.037	0.233	441
Unpublished	0.063	-0.082	0.208	0.065	-0.080	0.210	324
Citations ≥ 21	0.083	-0.039	0.205	0.091	-0.031	0.213	440
Citations < 21	0.072	-0.093	0.237	0.087	-0.078	0.252	325
Publication Year ≥ 18	0.079	-0.082	0.240	0.087	-0.074	0.248	420
Publication Year < 18	0.077	-0.037	0.191	0.095	-0.019	0.209	345

Notes: The table presents mean values of standardized regression coefficients for different subsets of data. Unweighted: We use the original values for the computation. Weighted: We weigh the estimates by the inverse number of estimates reported by each study. CI = Confidence Interval; N = Number of Observations; OLS = Ordinary Least Squares; FE = Fixed-effects; RE = Random-effects; QR = Quantile Regression; IV = Instrumental Variable. For a detailed explanation of the variables, please refer to Section 5.1.

Chapter 4

Publication Bias

In the preceding sections, we conducted a literature review and discussed the data collection methods employed to examine the impact of general mental ability on labor market outcomes. Based on the compiled estimates, we calculated the overall mean value of the economic returns to ability and examined the average effects on various subsets of the dataset, which offered valuable insights into the relationship. However, this method alone did not account for potential biases in summary statistics, specifically publication selection bias. Therefore, in this chapter, we will address this issue to gain a more comprehensive understanding of the credibility of the collected effect estimates.

Publication selection bias, or simply publication bias, is a systematic error that occurs when the publication status of a research study is influenced by the direction, strength, and significance of its results. In the meta-analysis of Card & Krueger (1995), three main sources of publication bias in economics were identified. Firstly, bias emerges when journal editors and reviewers prefer and authorize only studies that align with the conventional view. In other words, papers supporting the established view are more likely to be published, while papers with non-appealing results that contradict the prevailing view are often left unpublished. Secondly, researchers themselves introduce bias into the literature by using the presence of traditional theory to guide them through model selection tests. Lastly, both researchers and journal editors tend to favor statistically significant results, leading to the underrepresentation of studies with insignificant empirical findings. Such studies are often overlooked and may remain unpublished, contributing to what is commonly referred to as the 'file drawer problem.' The combination of these sources, all favoring a particular theory or direction, can distort empirical results, leading to an overestimation of the effect size and biased conclusions. Consequently, addressing and correcting for this bias is crucial in systematic reviews and meta-analyses to achieve accurate results and overcome the limitations of selection bias.

The recognition of publication bias as a potential threat to research findings initially gained attention in fields such as medicine and clinical research (Stanley 2005). In economics, the acknowledgment of publication bias emerged later, with notable contributions from Card & Krueger (1995), Ashenfelter *et al.* (1999), and Gorg & Strobl (2001). With the increasing number of empirical papers being published in economics, new challenges arise. The extensive volume of research results in a mixture of findings, some of which are consistent while others are not, leading to ambiguous answers to empirical questions. This has prompted the use of meta-analyses and the adoption of state-of-the-art methods to determine robust effect sizes and mitigate potential biases. As the field of economics continues to evolve, it is essential to employ rigorous methods to address publication bias to obtain reliable estimates.

In the previous research discussed in Section 2.4, only Ozawa *et al.* (2022) addressed publication bias. They examined selection bias by using a funnel plot that plotted the effect size (one SD change in cognitive test score) against the sample size of each corresponding primary study. However, given the resulting funnel plot and the limited number of studies included in their meta-analysis (four studies capturing the effect of cognitive ability on economic outcomes in low- and middle-income countries), it was not possible for them to definitively rule out the presence of publication bias in their results. Since cognitive ability is generally believed to be positively associated with higher lifetime earnings, it is plausible that there may be an upward bias in the aggregated effect sizes, indicating the potential presence of bias. If this is indeed the case, the sample mean reported in the previous section may not accurately represent the true effect size of our phenomenon. Therefore, in this chapter, we will thoroughly explore this issue and address the potential bias in our collected estimates by implementing a series of rigorous tests.

4.1 Testing for Publication Bias

4.1.1 Graphical Test

We begin our investigation of publication bias by employing one of the most commonly used graphical tools in meta-analyses, the funnel plot (Egger *et al.*

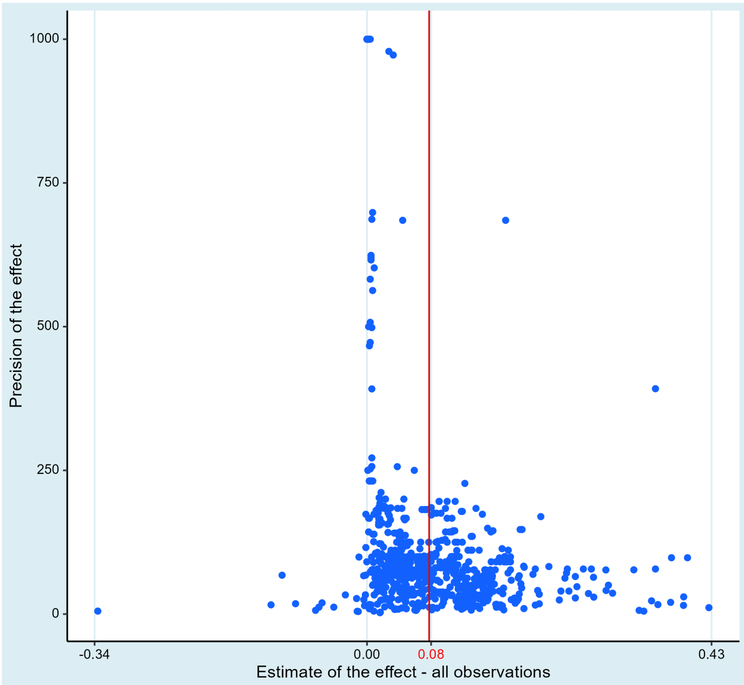
1997). This plot serves as a visual representation of the validity and potential bias of the results. It is constructed as a scatter plot where the effect estimates from individual studies are plotted on the horizontal axis against a measure of precision on the vertical axis. The measure of precision is typically represented by the inverse of standard errors, although an alternative approach, as demonstrated by Ozawa *et al.* (2022), is to use sample size. When examining the funnel plot, we would expect the most precise estimates at the top of the plot to be closer to the true mean effect. As the precision of the estimates decreases towards the bottom of the plot, we would anticipate a scattering of estimates on both sides of the true effect. Therefore, in the absence of publication bias, the funnel plot should exhibit an inverted funnel shape, from which the test derives its name.

When publication bias is present, the funnel plot reveals it through two distinct patterns. Firstly, asymmetry in the plot suggests the omission of specific estimates that do not align with traditional beliefs. This occurs when there is a preference for either positive or negative effect estimates, leading to the underrepresentation of studies with contrasting results. Secondly, the plot may appear hollow in certain sections, indicating an underrepresentation of studies with insignificant estimates. This occurs because papers with statistically significant results are more likely to be published, while studies with non-significant findings might remain unpublished. It is also possible to observe a combination of these two patterns, as authors may strive to produce more appealing results deemed worthy of publication status.

We applied the same techniques to our data and created a funnel plot, as shown in Figure 4.1, where standardized coefficient estimates are plotted against the inverse of the corresponding standard errors. The most precise estimates tend to cluster around positive values close to zero along a relatively straight line, with some outliers. This suggests that the true effect size may be lower than the simple mean mentioned in Section 3.3. The funnel plot exhibits a strong right skewness in our data, indicating the potential presence of publication bias, as evidenced by the asymmetry on the left side of the plot. According to our model, it is highly likely that negative effect estimates are often overlooked and left unreported, as they contradict the notion of intelligence positively influencing income levels. However, it appears that very imprecise positive estimates are not underreported.

It is important to note that the funnel plot, being a graphical construct, relies on the authors' visual assessment and is, therefore, a subjective measure of bias. Consequently, at this stage of our investigation, we cannot definitively conclude the presence of publication bias in the literature on our current topic. In the following subsections, we will complement the graphical analysis with quantitative tests of publication bias to provide a more comprehensive evaluation.

Figure 4.1: Funnel plot



Notes: The figure illustrates a funnel plot of all the estimated standardized coefficients that capture the effect of general ability on labor market outcomes. The bold (red) vertical line indicates the simple mean of all the estimates.

4.1.2 Linear Tests

The more rigorous alternative to the funnel plot is the Funnel Asymmetry Test (FAT), which was discussed by Stanley (2005). This quantitative approach assumes the exogeneity of the standard error and a linear relationship between the effect estimates and their standard errors. To conduct the FAT, we employ a linear regression model and estimate the following equation:

$$s_{ij} = \beta_0 + \beta_1 SE_{ij} + \epsilon_{ij} \tag{4.1}$$

In the Equation 4.1, s_{ij} represents the i-th effect size, SE_{ij} represents its corresponding standard error reported in the j-th study. In the absence of publication bias, there should be no correlation between the reported estimates and their standard errors, and the intercept β_0 corresponds to the underlying effect size corrected for publication bias. The correlation is captured by the coefficient β_1, which indicates the existence, direction, and magnitude of potential bias. The error term ϵ_{ij} represents the residual of the regression model.

In our analysis, we manage potential heteroscedasticity by clustering standard errors of regression parameters at the study level, acknowledging variations in data and methodology across studies. This approach effectively addresses the within-paper correlation of error terms while assuming independence across studies. We explore the relationship between the estimate and the standard error using various methods, beginning with ordinary least squares (OLS). Considering that each study contributes a different number of estimates to the dataset, we further employ inverse weighting based on the number of observations per study. This weighting scheme ensures that each study has an equal impact on the result. Additionally, we adopt the common practice of using the inverse of the standard error as weights in the regression, as recommended by Ioannidis *et al.* (2017). By doing so, we account for potential heteroscedasticity in the sample, providing more accurate estimates of Equation 4.1, as suggested by Stanley (2005).

Furthermore, we present the findings obtained through fixed-effects (FE) and between-effects (BE) estimation. While FE models assume a consistent effect size across diverse studies, it is important to recognize that the baseline effect is likely to exhibit variation across different studies and samples. In such situations, the BE model becomes more suitable as it does not assume a uniform effect size across various papers. By employing the BE model, we effectively account for the inherent between-study variability in effect sizes attributable to both sampling error and heterogeneity.

The results from the model specifications are presented in Table 4.1. With the exception of two weighted schemes, the mean effects generally resemble the uncorrected mean estimates observed in Section 3.3 after accounting for publication bias. Specifically, the simple uncorrected mean effect was 0.078, while the weighted uncorrected mean effect was 0.090. Importantly, all of these effects demonstrate statistical significance, reaching at least a 10% significance level. Therefore, we have sufficient evidence to reject the hypothesis that the true effect of intelligence on income levels is non-existent.

However, upon considering publication bias, it becomes evident that four out of five models exhibit a significant presence of positive publication bias, indicating a preference for publishing reports with positive findings. Notably, when the regression is weighted by the inverse of the standard errors, the magnitude of bias surges to 4.96. This substantial increase in bias results in a significant decrease in the effect size of 0.007 compared to the other effect estimates. On the other hand, the FE method demonstrates the lowest magnitude of bias at 0.151. This suggests a smaller distortion in the estimation of the effect. The remaining model specifications indicate a presence of mild publication bias.

Table 4.1: Linear tests for publication bias

	OLS	**Study**	**Precision**	**FE**	**BE**
Publication bias	0.377***	0.47***	4.96***	0.151**	0.421
SE	(0.143)	(0.119)	(0.326)	(0.076)	(0.621)
Effect beyond bias	0.07***	0.046***	0.007*	0.075***	0.081***
Constant	(0.004)	(0.003)	(0.004)	(0.002)	(0.017)
Studies	38	38	38	38	38
N	765	765	765	765	765

Notes: The table displays the results of linear regression testing to examine the presence of publication bias. The standard errors of the regression parameters are shown in parentheses and clustered at the study level. Study = Estimates are weighted by the inverse of the number of observations reported per study. Precision = Estimates are weighted by the inverse of their standard error. OLS = Ordinary Least Squares; FE = Fixed-effects; BE = Between-effects; Studies = Number of studies; N = Number of observations. *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$

4.1.3 Non-linear Tests

The tests of publication bias we have conducted so far have provided meaningful insights into our analysis. However, it is important to acknowledge that these tests have assumed a linear relationship between the effect estimates and their standard errors. This assumption may pose a problem as it is likely to be violated in a majority of cases, especially when highly precise estimates are prone to selection bias due to their small standard errors (Stanley *et al.* 2010). Consequently, the estimations we have obtained could be imprecise, leading to an exaggerated bias and a potential shift in the true effect, either upward or downward depending on the direction of publication bias. To address this concern, we will employ several non-linear tests, drawing inspiration from the latest meta-analyses conducted by Havránek *et al.* (2020; 2021), to further examine the presence of bias and its potential impact on the true effect.

Table 4.2: Non-linear tests for publication bias

	Top10 method	**WAAP**	**Stem-based method**	**Endogenous Kink**	**Selection model**
Publication bias				4.96***	0.068***
				(0.369)	(0.003)
Effect beyond bias	0.026***	0.066***	0.001	0.007***	0.066***
	(0.006)	(0.003)	(0.014)	(0.002)	(0.004)
Studies	38	38	38	38	38
N	765	765	765	765	765

Notes: The table presents the results of non-linear methods to examine the presence of publication bias. The standard errors are shown in parentheses and clustered at the study level. WAAP = Weighted Average of the Adequately Powered; Studies = Number of studies; N = Number of observations. *p < 0.1, **p < 0.05, ***p < 0.01

In addressing non-linear tests for publication bias, we commence with the Top10 method as proposed by Stanley *et al.* (2010). Stanley's method is admirably uncomplicated and acknowledges the potential bias resulting from the preference for statistically significant findings over nonsignificant ones, which can lead to unrepresentative samples and skewed results. The method counteracts this bias by computing a simple average of the top 10% most precise estimates while disregarding the remaining 90% of the data. The rationale behind this is that the estimates with the highest precision are less susceptible to distortions arising from publication bias. Applying this method to our dataset, which comprises 75 model observations, results in a mean effect of 0.026, as presented in Table 4.2. This lower value aligns with the funnel plot depicted in Figure 4.1, indicating that the most precise estimates are closer to zero.

The second method we have applied in our study is the Weighted Average of Adequately Powered (WAAP) approach, as proposed by Ioannidis *et al.* (2017). This approach takes into account the inclination toward publishing papers with statistically significant estimates. It computes the underlying effect by incorporating only those estimates that are adequately powered (with statistical power above 80%) and weights them based on the inverse of their variance. To qualify as 'adequately powered', an estimate must have a standard error lower than the power threshold, defined by the sum of the statistical significance (1.96) and the adequate power value (0.84) as per Ioannidis *et al.* (2017). The results of employing the WAAP method on our dataset are illustrated in Table 4.2, which draws upon 503 adequately powered estimates.

The previously discussed methods substantially mitigate the issue of publication bias, but simultaneously they also remove a considerable amount of variation in the estimates. This occurs because a reduction in the number of estimates inevitably leads to a decrease in data variability, whereas increasing the number of imprecise estimates contributes to a rise in bias (Stanley *et al.* 2010). Building upon the 'Top10' method, Furukawa (2019) proposed an approach aimed at optimizing this balance. This approach identifies and utilizes only the optimal number of the most precise observations based on minimizing the Mean Squared Error of the estimates. Named for its reflection of the stem of the funnel plot, this approach is called the 'Stem-based' method. From the tests we implemented, this approach yielded the smallest effect size of 0.001, though it fell short of reaching statistical significance.

Following the same logic as the Top10 method, we next implement the Endogenous Kink method, a technique introduced by Bom & Rachinger (2019). The fundamental premise of this method is that publication bias surfaces when the standard error of an estimate surpasses a specific threshold. This cutoff point is endogenously determined by performing a piecewise linear meta-regression of the estimates on their respective standard errors, identifying the so-called 'kink.' Below this threshold, the probability of publication selection is low, and the relationship deviates from linearity to best fit the data. Consequently, if the model does not detect a kink, the relationship remains linear, and the kinked model simplifies to the FAT model. This method produces the most substantial publication bias observed, at 4.96, identical to the magnitude seen with the precision-weighted scheme referenced in Table 4.1. Furthermore, it results in a negligible effect size of 0.007. However, both estimates are statistically significant at the 1% level.

Finally, we introduce an entirely distinct approach compared to the non-linear tests previously employed: selection models that assign different weights to estimates based on their statistical significance. A prominent selection model commonly used in economics was developed by Andrews & Kasy (2019). This model assumes that estimates and their standard errors are statistically independent. It accounts for selection bias by defining the probability of publication as a function of a study's findings, or 'conditional publication probability', thereby facilitating the examination of the true effect within our dataset. Using this Selection Model, we discern a minimal publication bias and a true effect of 0.066, which is relatively close to the simple mean effect documented in Section 3.3.

In summary, four out of the five mean effects derived from the non-linear tests we conducted are statistically significant, suggesting an existent effect of general ability on financial outcomes. However, the magnitude of this effect is relatively small, particularly in the case of the Stem-based model, which nonetheless is not statistically significant. Regarding the methods that explore publication bias coefficients, there remains a statistically significant positive publication selection, as observed in Subsection 4.1.2.

4.1.4 Dissolving the Exogeneity Assumption

Up until this point, all previous methods have assumed the standard error to be exogenous to the effect estimate, implying that these two are uncorrelated in the absence of publication bias. However, in economics, such a scenario is relatively uncommon due to factors such as measurement errors, omitted variables, and reverse causality (Gechert *et al.* 2022). As such, to further test for publication selectivity while easing the assumption of exogeneity, we will employ a straightforward solution: instrumenting the standard error. By definition, larger sample sizes yield smaller standard errors, establishing a correlation between these two variables. Consequently, an appropriate instrumental variable will be defined as the inverse of the square root of the number of observations reported in the primary study. The outcome of this technique presents a publication bias direction that contradicts all previous findings, with a substantial value of -0.601, although it's not statistically significant (see Table 4.3). On the other hand, the mean effect size of 0.092 is fairly close to the uncorrected weighted mean effect (refer to Section 3.3) and is statistically significant.

Table 4.3: Instrumental variable regression

	IV
Publication bias	-0.601
	(0.364)
Effect beyond bias	0.092***
	(0.009)
Studies	38
N	765

Notes: The table displays the results of Instrumental Variable (IV) regression. The standard errors are shown in parentheses and clustered at the study level. Studies = Number of studies; N = Number of observations. *p $<$ 0.1, **p $<$ 0.05, ***p $<$ 0.01

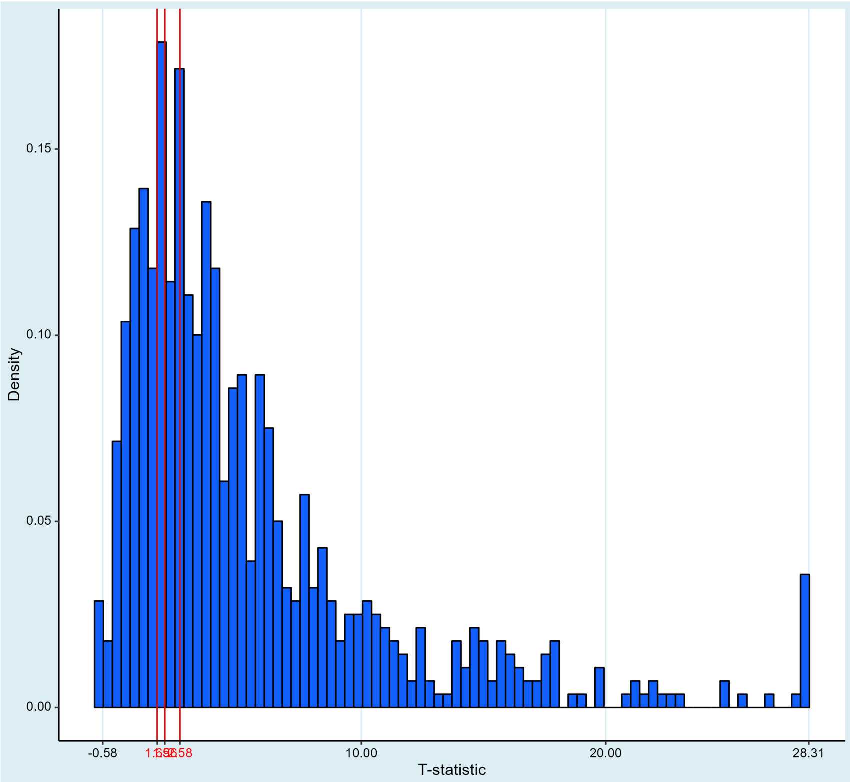
Another technique that forgoes the exogeneity assumption is the p-uniform* method, recently innovated by van Aert & van Assen (2021). Contrary to the IV method, this technique investigates the distribution of p-values rather than the estimates and their standard errors. The underlying principle here is that p-values should display a uniform distribution around the true effect. The technique thus tests this assumption by identifying the coefficient of the true effect, at which the resulting p-value distribution approaches uniformity. We attempted to apply this approach to our data sample, however, it turned out to be unfeasible for our dataset as the method failed to converge on any reliable results. Consequently, we are unable to provide estimates using this method.

The final method that assumes no correlation between the effect estimate and its standard error is the Caliper test, as proposed by Gerber & Malhotra (2008). Unlike the previous method that examines p-values, this test compares the frequency of estimates below and above specific critical t-values within a sufficiently narrow caliper width. If a particular statistical value is over-represented in the sample, a sudden increase will be visible around that value, suggesting the presence of publication bias. Conversely, in the absence of such bias, no noticeable difference should occur around this value.

Given the right-skewed nature of our data, we focus on the critical thresholds of 1.645, 1.96, and 2.58, which correspond to significance levels of 10%, 5%, and 1%, respectively. From the distribution of t-statistics for our estimates, displayed in Figure 4.2, we can observe some noticeable jumps in the number of reported t-values around these thresholds, suggesting a potential bias towards statistically significant results. To empirically examine these patterns, we perform Caliper tests. While these tests don't yield an estimate of the true effect, they provide an understanding of publication bias at specific conventional thresholds. The coefficients presented in Table 4.4 should be interpreted in the following manner: the obtained values represent the difference between the proportion of estimates above the threshold and the proportion that would be expected in an ideal scenario without any publication bias, which would be 0.5. For example, within a caliper width of 0.05 around 1.645, 67.1% of the estimates significantly exceed zero at the 10% level, while the remaining 32.9% are not statistically significant.

In a perfect setting without publication bias, the results from the Caliper tests should approximate zero. However, our sample does not align with this setting, as the coefficients range from 0.166 to 0.386, with both extremes observed in the tests around the critical value of 2.58. Furthermore, all of the coefficients reach statistical significance at least at the 5% level. We can see a

Figure 4.2: T-statistic distribution



Notes: The figure displays the distribution of t-statistics for the reported estimates within our dataset. The bold (red) vertical lines denote the critical values of 1.645, 1.96, and 2.58, corresponding to significance levels of 10%, 5%, and 1%, respectively.

Table 4.4: Caliper tests for publication bias

Threshold for t-statistic	**1.645**	**1.96**	**2.58**
Caliper size 0.05	0.171**	0.305***	0.386***
	(0.085)	(0.053)	(0.001)
	$N = 11$	$N = 22$	$N = 4$
Caliper size 0.1	0.212***	0.269***	0.199***
	(0.068)	(0.041)	(0.04)
	$N = 18$	$N = 37$	$N = 34$
Caliper size 0.2	0.288***	0.257***	0.166***
	(0.047)	(0.034)	(0.026)
	$N = 38$	$N = 52$	$N = 52$

Notes: The table displays the outcomes of three sets of Caliper tests for three chosen thresholds. The standard errors are shown in parentheses and clustered at the study level. N = Number of observations reported in each respective interval. *p < 0.1, **p < 0.05, ***p < 0.01

slight increase in the discrepancy between statistically significant and insignificant estimates as the caliper size increases around the threshold of 1.645. In contrast, for the 5% and 1% significance levels—with the former demonstrating the most pronounced publication bias—the imbalance decreases as the caliper width expands. Nevertheless, the narrowest caliper around the 1% significance level only yields four observations, which impedes any definitive conclusions regarding bias. Overall, the data suggest a prevailing preference for statistically significant results at any conventional level of significance in our dataset.

To summarize our examination for publication bias, the majority of the methods pointed towards a mild to substantial positive publication bias of statistical significance in the literature. There was only one instance where a negative selection bias was estimated, though it did not meet statistical significance at any conventional threshold. All of the techniques, except for the Stem-based method in Table 4.2, produce statistically significant effect sizes corrected for bias, ranging from 0.007 to 0.092. While this range is somewhat broad, impeding a precise determination of the effect magnitude, it instills more confidence in asserting that the effect of an individual's general mental ability on financial outcomes does indeed exist. However, these findings still require further testing, as we have only worked with measures of uncertainty so far. There may be additional explanatory variables that could drive the effect, which we have yet to specify. As such, the next chapter will concentrate on uncovering the data set's heterogeneity, enabling us to verify the robustness of the assertions made in this chapter.

Chapter 5

Heterogeneity

Despite coinciding in direction, the mean effects we reported previously exhibit differences in their magnitudes. In the prior chapter, we established that positive publication bias is responsible for a substantial portion of the variation seen among the estimates. Now, we are shifting our focus to other sources of heterogeneity that are linked to differences in study designs. This chapter will delve into how certain variables drive our effect of interest and outline their behavioral patterns. We commence by defining each coded variable, explaining our motivations for their inclusion, and providing their summary statistics. Subsequently, these variables are employed within the framework of Bayesian Model Averaging (BMA) methods to investigate the origins of heterogeneity within the existing literature. The outcomes of these techniques are then interpreted and discussed.

5.1 Explanatory variables

In the process of data collection, we carefully observed and subjectively identified the most relevant characteristics within the primary studies to facilitate a thorough examination of the heterogeneity in our data sample. As previously noted, this section is committed to detailing our chosen set of explanatory variables, providing a deeper understanding of each factor that may influence the ability-income effect. Our final selection is comprised of 54 distinct coded variables, which are itemized, along with their definitions and summary statistics, in Table 5.1. These variables are grouped into six characteristic categories: data characteristics, methodology, variable specifications, subjects and country characteristics, control variables, and study characteristics. Without further ado, we will now delve into the particulars of each variable.

Data characteristics Apart from the core data characteristics, namely the effect estimate and the corresponding standard error, we gathered additional variables to illuminate the nature of the samples used in the primary studies. First, we collected the number of observations associated with each estimate. The resulting *No. of Observations* variable shows significant variation, ranging from 172 (Psacharopoulos & Velez 1992) to 346 660 (Spengler *et al.* 2018) observations. The smallest dataset comprises workers from Bogota, Colombia, limited to the public sector, while the largest draws from a nationwide sample of US high school students from the Project Study cohort. Limited sample sizes can pose challenges in research studies, as they may diminish statistical power, restrict the generalizability of findings, and affect the reliability and validity of the study's conclusions. Therefore, we coded the *No. of Observations* for each effect estimate to account for these variations in sample sizes.

Another factor that could influence the observed effect is the data type related to the time horizons over which the estimate is analyzed. This distinction is based on whether an estimate derives from cross-sectional or longitudinal data. Cross-sectional data involve observations collected from different individuals, groups, or entities at a specific point in time. On the other hand, panel data, or longitudinal data, entail collecting information from the same subjects over an extended time period, allowing for the indetification of changes, trends, and developments over time. Strenze (2007) even confined his meta-analysis to longitudinal studies, arguing that only these can address the causal impact of general cognitive ability on economic success. However, in our meta-analysis, we opted to include both data types to observe how the effect fluctuates when controlled for this factor by including a dummy variable called *Panel Data*. This factor is closely correlated with the distinction between short-run estimates, which represent the immediate effect of intelligence on labor market outcomes, and long-run estimates, which capture the sustained or permanent outcome, as mentioned in Section 3.3. This variation is encapsulated in the variable referred to as *Long-run Estimate*.

The data can also be differentiated based on the method of collection, or in other words, the data source, which divides the input into three types - micro-data, survey data, and national register data. The *Micro-data* variable pertains to information gathered on an individual level, allowing for a detailed analysis of specific subjects within a population. *Survey Data*, in contrast, represents data procured from surveys or questionnaires administered to a sample of households or employers. Lastly, input from national registers refers to official information collected and maintained by government agencies or institutions at a national level. These registers often encompass administrative records, such as birth records, healthcare databases, and, most importantly for our topic of interest, tax records. As there is no selection bias with this type of data, it is considered the most valuable input characteristic. However, only 4.3% of our data sample is classified under the *National Register Data* variable (Falch & Massih 2012; Sorjonen *et al.* 2012; Frøyland & Von Soest 2020). We also coded for self-reported wage, but with 92% of the data meeting the condition of this variable, there is minimal variation. Furthermore, 35% of the data did not provide any information about this characteristic, leading to its exclusion from further examination.

The final set of coded variables in this category relates to temporal variations, specifically *Average Data Year* and *Time Span*. *Average Data Year* is the arithmetic mean year of the data used, computed from the first year when intelligence or cognitive ability was initially assessed to the last year when personal income data was gathered. On the other hand, *Time Span* denotes the duration in years, calculated as the difference between the start and end year of data collection. Estimates with a time frame extending beyond one year are deemed long-run estimates. The longest time horizon in our dataset spans 55 years, as documented in a study by Falch & Massih (2012). The estimates in this study illustrate the long-term effect of cognitive ability on personal earnings, in which ability was evaluated in 1938 when third graders in Malmö, Sweden turned ten. Starting in 1948, when they reached the age of 20, earnings data was collected roughly every five years until 1993 when these individuals attained the age of 65, coinciding with the retirement age in Sweden during that period.

Variable specifications Our independent variable, intelligence, signifies an abstract capability that is independent of any specific learned knowledge but often correlates strongly with academic performance. Numerous studies, across varied subjects and age groups, consistently indicate a positive association between intelligence and academic achievement. For instance, research conducted by Deary *et al.* (2007) revealed a substantial correlation of 0.81 between a latent intelligence trait and academic performance. However, in our meta-analysis, we make a firm distinction between general mental ability – one's aptitude and potential to execute mental tasks such as learning and comprehension – and academic achievement, which we view as acquired knowledge or academic skills. Consequently, we have omitted studies employing academic performance indicators, such as high school grade point average (GPA), as measures of ability.

The variable *Intelligence* incorporates estimates drawn from widely accepted intelligence tests, reflecting the true impact of intelligence. Prominent classical tests featured in our sample include Raven's Progressive Matrices and Henmon-Nelson Tests. However, directly measuring intelligence through these thorough, rigorous assessments may not always be feasible or practical. Consequently, researchers frequently employ cognitive ability measurements as a substitute to approximate general mental capacity. This approach inspired the creation of a new variable, *Intelligence Proxy*, associated with cognitive ability tests such as the Armed Forces Qualification Test (AFQT), which occurs 157 times within our dataset of 765 total observations.

In our study, the dependent variable refers to financial outcomes, which we have classified into five distinct categories. First, we encoded for *Income*, denoting income in units of currency like euros or US dollars, or as mean-centered income. Though these cases are relatively infrequent, they can be found in studies by Andersson & Bergman (2011), Spengler *et al.* (2015), and Frøyland & Von Soest (2020). Given that income distribution often exhibits a considerable skew, logarithmic transformation of income is a prevalent strategy employed to diminish this skewness and satisfy the commonly made assumption of a normal distribution in statistical analyses. Consequently, the *Income* variable also encompasses cases where income has undergone a logarithmic transformation. The remaining categories relate to earnings expressed in logarithmic form as an outcome variable, adjusted according to different timeframes. These categories, particularly *Log Wage: Hourly, Log Wage: Day/week, Log Wage: Monthly*, and *Log Wage: Annually*, reflect the heterogeneity present in the literature on this subject.

Methodology The following category highlights the methodological strategies scholars employ when calculating the effect of interest, leading to five distinct classifications. Unsurprisingly, Ordinary Least Squares (OLS) represents the primary method of choice. This technique is either applied singularly, as seen in studies by Mueller & Plug (2006) and Cunha *et al.* (2011), or used alongside other estimation methods, as demonstrated in the works of Cassagneau-Francis (2022); Zax & Rees (2002); Cheung (2006). Researchers sometimes attempt to account for unobserved heterogeneity by implementing fixed- or random-effects models. An example from our meta-analysis that leverages both models is the study by Hartog *et al.* (2010), where fixed- and random-effects are employed to capture individual-level and time-invariant heterogeneity, respectively. Another estimation approach involves applying quantile regression, as observed in a paper by Gregg *et al.* (2015), who explores intergenerational persistence at various points of the income distribution. To yield more robust estimates, researchers can employ instrumental variables to control for endogeneity, measurement error, or omitted variable bias, as exemplified in a study by Glewwe *et al.* (2017). Any other method that does not fall into any of the above groups is designated as an *Method: Other*. This includes papers such as Jandarova (2023), which employs Heckman selection correction, and Hartog *et al.* (2010), which applies a Difference-of-difference model, among others.

We anticipate that the chosen estimation methodologies will affect the reported estimates, considering the varying foundational assumptions of these techniques and their differential ability to address potential biases. Preliminary indications of this can be seen in Table 3.1, where the effect sizes in the 'Methodology' section substantially differ based on the distinct methodological approaches employed.

Subject and country characteristics Furthermore, it is crucial to account for the variations in participants' attributes and data provenance. A common source of heterogeneity across distinct research areas is the gender of the sample. To address this, we devise variables that represent the proportions of male and female participants. Another aspect we consider is the subjects' average age. Unfortunately, some studies, including those by Glewwe *et al.* (2017), Arcidiacono *et al.* (2010), and Marks (2022), have omitted this information, resulting in missing values. To adhere to the conditions of model averaging, we must substitute these missing entries with the mean for the *Gender: Male* and *Gender: Female* variables, and replace missing values in the *Mean Age* variable with the median.

Variations in the data also arise due to geographical factors. Firstly, we adjust for the geographical unit level, determined by the capability to aggregate. This approach categorizes our data into four groups: *Agg.: City, Agg.: State, Agg.: Country*, and *Agg.: Continent*. In addition, we compiled information on the country of data origin, pinpointing the geographical source of the data. Altogether, we derived estimates from 14 different countries. Approximately half of the primary studies originate from the US, an expected finding given that the US is the epicenter of a considerable amount of research. The other half, possessing a roughly equivalent share, incorporates European nations and countries from other regions, including Australia, China, Colombia, and Mexico. We can already observe variations in effect size across different countries in Figure A.1, offering some initial evidence of heterogeneity.

Control variables In regression models, numerous variables can be considered. To investigate how the core effect size shifts when additional variables are introduced into the regression analysis, we introduce a variety of frequently encountered variables. These cover various categories: firstly, basic participant information such as age, gender, ethnicity, marital status, residence, and educational attainment. The subsequent variables pertain to job specifics (industry, experience and its quadratic form, tenure, and occupational status) and non-cognitive abilities, including factors like self-esteem and locus of control (Eren & Ozbeklik 2013; Pasche 2009). Finally, we account for family characteristics, coding for family income, the father's education level, and family socioeconomic status. Interestingly, yet not unexpectedly, more than half of the regression equations adjust for educational attainment, reflecting the standard Mincer equation discussed in Section 2.4. For further clarification, the full list of control variables can be observed in Table 5.1.

Study characteristics Lastly, as is standard in meta-analyses, we incorporate several variables related to study attributes. Recognizing that studies employing the correct statistical procedures are more likely to be published in reputable journals and have a higher number of citations, we control for the quality of primary studies by creating the variables *Published, Unpublished*, and *Citations*. Additionally, scholars often develop new and more effective meta-analysis techniques over time, leading to more robust findings. We account for this matter by calculating the difference between the publication year of a specific study and the earliest published paper in our sample, denoted as *Publication Year*. Although our dataset's mean years range from 1964 to 2015, the span of publication years is notably 31 years. Finally, the last coded variable, *Study size*, records the total number of estimates reported in each study, varying significantly from 1 to 88 observations per study. The exact study size for each primary study included in our meta-analysis can be found in Table A.1.

Table 5.1: Overview of explanatory variables

Variable	Description	Mean	SD
Size Effect	The standardized regression coefficient capturing the ability-income effect	0.078	0.076
Standard Error	The standard error of the size effect	0.023	0.033
Data characteristics			
No. of Observations*	The number of observations associated with the estimate	10 164.15	27 309.18
Cross-sectional Data	= 1 if the study uses cross-sectional data	0.180	0.385
Panel Data *(ref. cat.)*	= 1 if the study uses panel data	0.820	0.385
Short-run Estimate*	= 1 if the estimate covers a period less than one year	0.187	0.390
Long-run Estimate* *(ref. cat.)*	= 1 if the estimate spans a period longer than one year	0.813	0.390
Micro-data*	= 1 if the study uses micro-data	0.773	0.419
Survey Data*	= 1 if the study uses data from a survey of households or employers	0.184	0.388
National Reg. Data* *(ref. cat.)*	= 1 if the study uses data from a national register	0.043	0.203
Average Data Year	The average year of the study's time span	1992.59	11.50
Time Span*	The duration of the data collection process	16.53	13.37
Variable specifications			
Intelligence	= 1 if the ability is directly measured by intelligence test	0.242	0.428
Intelligence Proxy *(ref. cat.)*	= 1 if the ability is measured by cognitive ability test	0.758	0.428
Income	= 1 if the dependent variable in the regression is income	0.077	0.267
Log Wage: Hourly	= 1 if the dependent variable in the regression is the log of hourly earnings	0.561	0.497
Log Wage: Day/week	= 1 if the dependent variable in the regression is the log of daily or weekly earnings	0.099	0.299
Log Wage: Monthly	= 1 if the dependent variable in the regression is the log of monthly earnings	0.107	0.310
Log Wage: Annually *(ref. cat.)*	= 1 if the dependent variable in the regression is the log of annual earnings	0.156	0.363
Methodology			
Method: OLS	= 1 if Ordinary Least Squares is used for estimation	0.612	0.488

Continued on next page

Table 5.2 – continued from previous page

Variable	Description	Mean	SD
Method: FE or RE	= 1 if Fixed-effects or Random-effects is used for estimation	0.052	0.223
Method: QR	= 1 if Quantile regression is used for estimation	0.115	0.319
Method: IV	= 1 if Instrumental Variable is used for estimation	0.046	0.209
Method: Other *(ref. cat.)*	= 1 if a distinct method is used for estimation	0.175	0.380
Subject and country characteristics			
Gender: Male	The ratio of male to female subjects (= 1 if all male, = 0 if all female)	0.625	0.636
Gender: Female *(ref. cat.)*	The ratio of female to male subjects (= 1 if all female, = 0 if all male)	0.375	0.636
Mean Age	The average age of the subjects	37.542	12.496
Agg.: City*	= 1 if the estimate can be aggregated on a city level	0.071	0.256
Agg.: State*	= 1 if the estimate can be aggregated on a state/province level	0.080	0.271
Agg.: Country*	= 1 if the estimate can be aggregated on a country level	0.68	0.467
Agg.: Continent* *(ref. cat.)*	= 1 if the estimate can be aggregated on a continental level	0.170	0.376
Country: USA	= 1 if the estimate originates from the USA	0.503	0.500
Country: Europe	= 1 if the estimate originates from Europe	0.255	0.436
Country: Other *(ref. cat.)*	=1 if the estimate originates from other country	0.242	0.428
Control variables			
Control: Age	= 1 if the effect is controlled for age	0.216	0.412
Control: Gender	= 1 if the effect is controlled for gender	0.265	0.442
Control: Ethnicity	= 1 if the effect is controlled for ethnicity or race	0.365	0.482
Control: Marital Status	= 1 if the effect is controlled for marital status	0.357	0.479
Control: Residence	= 1 if the effect is controlled for residential location	0.392	0.489
Control: Education	= 1 if the effect is controlled for educational attainment	0.542	0.499

Continued on next page

Table 5.2 – continued from previous page

Variable	Description	Mean	SD
Control: Industry*	= 1 if the effect is controlled for the job industry	0.165	0.371
Control: Experience	= 1 if the effect is controlled for work experience	0.399	0.49
Control: Experience^2	= 1 if the effect is controlled for work experience in quadratic form	0.167	0.374
Control: Tenure	= 1 if the effect is controlled for job tenure	0.157	0.364
Control: Occupation	= 1 if the effect is controlled for occupational status	0.169	0.375
Control: Non-cognition	= 1 if the effect is controlled for non-cognitive ability	0.222	0.416
Control: Family Income	= 1 if the effect is controlled for family's income	0.103	0.305
Control: Father's Education	= 1 if the effect is controlled for father's education in the regression	0.214	0.411
Control: Family SES	= 1 if the effect is controlled for the family's socioeconomic status	0.114	0.318
Study characteristics			
Published	= 1 if the study was published in a journal	0.576	0.494
Unpublished *(ref. cat.)*	= 1 if the study was not published in a journal	0.424	0.494
Citations	The number of Google Scholar citations	98.851	162.613
Publication Year*	The number of years between the publication year of the study and the earliest published study in the sample	17.827	6.236
Study Size	The number of estimates collected from the study	41.319	28.112

Notes: The table presents descriptions and summary statistics for each coded variable. Variables marked with an asterisk (*) are excluded from the model averaging models due to high multicollinearity. Mean = simple unweighted mean; SD = Standard Deviation; OLS = Ordinary Least Squares; FE = Fixed-effects; RE = Random-effects; QR = Quantile Regression; IV = Instrumental Variable; *(ref. cat.)* = reference category; SES = Socioeconomic Status.

5.2 Model Averaging

As evident from the previous section, our analysis involves a substantial number of coded variables—54 explanatory variables of various study specifications. However, incorporating all these variables into our subsequent heterogeneity analysis could lead to model overspecification. This is a scenario where a statistical model becomes unnecessarily complex due to the inclusion of an excessive number of factors. Such overspecification can give rise to several potential issues, such as multicollinearity and model overfitting, which can result in reduced precision and weak predictive performance. Hence, our primary objective now is to identify those variables that have the most significant influence on the underlying effect of intelligence on income levels. This challenge is referred to as model uncertainty, given our lack of certainty about what the ideal model, that accurately mirrors reality, should look like. Considering the number of our coded variables, the total count of potential models equates to 2^{54}, which approximates an immense 18 quadrillion (10^{15}) combinations. Obviously, manually selecting the most accurate models from this vast pool would be an impractical endeavor. As such, we will address this issue by employing Bayesian Model Averaging (BMA) (Steel 2020).

BMA simplifies the process for us by generating models with diverse subsets of independent variables. Each model is allocated a weight known as posterior model probability, which grows with the model fit but diminishes with the number of included regressors. Using this measure, BMA generates the Posterior Inclusion Probability (PIP) for each explanatory variable. PIP is calculated by summing up all the posterior model probabilities for the models that include the specific regressor. Essentially, this value signifies the relevance of each variable and the degree to which it explains the heterogeneity in the existing literature. The posterior model probabilities also yield two other critical indicators. The first is the weighted posterior mean, indicating the direction and magnitude of the relationship between the regressor and the effect size (i.e., the regression coefficient). The second is the weighted posterior variance (or weighted posterior standard deviation), which is perceived as the standard error of the former (Raftery *et al.* 1997).

In situations with a high number of variables, as in our case, executing BMA with an immense model space can be computationally challenging, even for a standard computer. To tackle this issue, we conducted our analysis using the *bms* package in R, leveraging the Metropolis-Hastings algorithm within the Markov chain Monte Carlo method. This approach simplifies the demanding computations by focusing on the most significant portions of the posterior model distribution, thereby approximating the posterior model probabilities. The process initiates with a benchmark model and contrasts it with a proposed candidate model, based on their posterior probabilities. The candidate model is either rejected, triggering the proposition of a new model, or accepted, becoming the new benchmark model. This procedure iterates until the target distribution is achieved (Fernandez *et al.* 2001).

BMA relies on prior probabilities for both the estimated coefficients and the models, requiring us to specify them before our model estimation. In terms of the distribution prior for the model parameters, often referred to as g-priors, we adopt the unit information prior. This default setting allocates the equivalent information found in a single observation (Zeugner & Feldkircher 2015). This choice is most suitable given our limited prior information about the distributions. When it comes to choosing the model distribution prior, we opt for the dilution prior over the commonly chosen uniform model prior. This decision stems from the significant number of explanatory variables we have and the accompanying potential for collinearity in our model. The dilution prior mitigates potential multicollinearity by weighting each model based on the determinant of the correlation matrix of the included covariates. As a result, models exhibiting high collinearity will carry lesser weight in our final results (George 2010).

In addition to our baseline BMA model, we utilize a data-driven alternative known as Frequentist Model Averaging (FMA) for robustness checks. Contrary to BMA, FMA circumvents the need for prior distribution specifications, allocating model weights solely on the basis of the available data (Magnus & De Luca 2016).

5.2.1 Implementation and Results

Before diving into BMA implementation, it is crucial to address collinearity within our coded variables. To mitigate collinearity, we omit variables that are most susceptible to it. The most apparent instance of collinearity among our variables emerges in the case of dummy variables derived from a categorical group, such as *Cross-sectional Data* and *Panel Data*. Including all these variables in the model would lead to perfect collinearity. To circumvent this 'dummy variable trap', we exclude one variable from each characteristic category. These excluded variables, which we refer to as reference categories, are denoted as *ref. cat.* in Table 5.1.

Furthermore, we examine the correlation matrix of the remaining variables and check their variance inflation factors (VIF) to identify multicollinearity among the regressors. As expected, a high positive correlation (0.978) is observed between standard errors and the number of observations, since the latter is used in the computation of the former. As a result, we exclude the *No. of Observations* to retain the *Standard Error* in our model. On assessing VIFs for the remaining variables, we find several with high VIFs. We decide to omit most of these, though we retain a few that we believe may provide valuable information, despite the risk of compromising result reliability. These include *Published, Country: Europe*, and *Country: USA*, the last of which has the highest VIF of 10.95. The other variables with high VIFs are dropped from our averaging models and designated with an asterisk (*) in Table 5.1. After thorough consideration, we proceed with 33 independent variables suitable for our model averaging. A correlation plot of these variables can be found in Appendix B. It is also noteworthy to mention that several variables, Average Data Year, Mean Age, Citation, and Study Size, were transformed to logarithmic form to bring their BMA coefficients closer to others without sacrificing the pertinent information they carry.

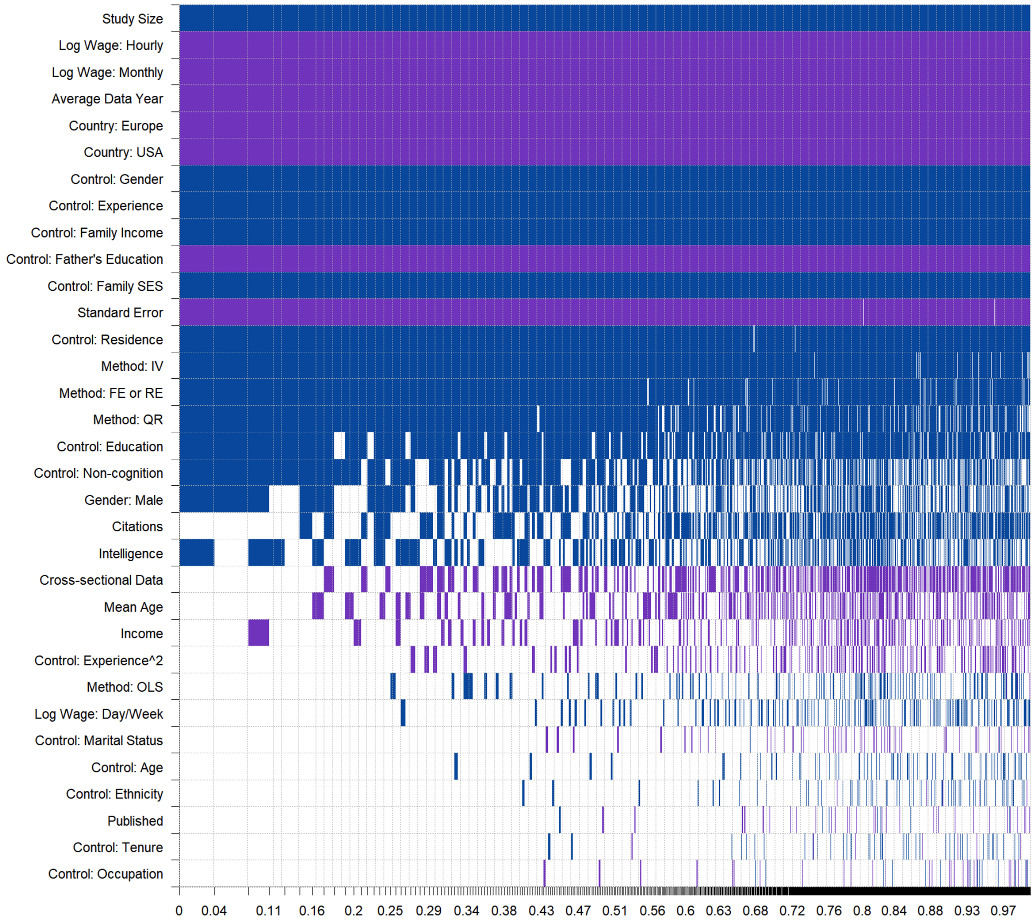
Now that we have introduced BMA and filtered out potentially problematic variables, we are prepared to proceed with the application of BMA in estimating the following meta-regression:

$$y = \beta_0 + \beta_1 SE + \sum_{i=1}^{32} \boldsymbol{\beta_i X}_i + \epsilon \tag{5.1}$$

where y represents the standardized regression capturing the economic returns to cognitive ability, while β_0 captures the average effect adjusted for publication bias conditional on the set of covariates $\boldsymbol{X}$. $\beta 1$ designates the direction and magnitude of publication bias, and SE corresponds to the reported standard error of the regression coefficient. $\sum i = 1^{32} \boldsymbol{\beta_i X}_i$ denotes the sum of the products of the 32 model variables and their corresponding regression coefficients. Finally, ϵ is the error term.

The graphical outcomes of the BMA are depicted in Figure 5.1, while Table 5.3 presents the numerical findings, alongside the FMA results. In the BMA's visual representation, the vertical axis organizes the independent variables on the left-hand side of Equation 5.1 in descending order of relevance based on their PIP values. The horizontal axis displays the score of the posterior model probabilities. Thus, the most effective models are situated on the figure's left, with the width of each column representing the corresponding posterior model probability. As for the figure's color scheme, white signifies the absence of a specific variable from the model. A positive effect is indicated by a purple color (lighter in greyscale), while a negative effect of the particular variable on the effect size is denoted by a blue color (darker in greyscale).

Figure 5.1: Model inclusion in Bayesian model averaging



Notes: The figure displays the primary Bayesian model averaging results using the uniform g-prior and dilution model prior. Variables are sorted vertically by their PIP, from highest to lowest. The horizontal axis shows the posterior model probability scale. Purple (lighter in greyscale) indicates a positive effect on effect size, while blue (darker in greyscale) signifies a negative effect. For a thorough breakdown of the variables, see Table 5.1.

To assess the BMA results numerically, let us turn our focus on Table 5.3 which provides individual values of the posterior mean, the posterior standard deviation, and the posterior inclusion probability (PIP). In a manner similar to that of simple regressions, the posterior mean, and the posterior standard deviation represent the estimated direction and magnitude of the effect and its precision, respectively. And PIP, as mentioned earlier, is considered as statistical significance, indicating the importance of each variable. Its value spans from 0 to 1 and the higher the number, the higher the chance of appearance in the true model. Following the suggestions of Jeffreys (1998), we classify PIP into four ranges:

- $0.5 < \text{PIP} < 0.75$: weak effect,
- $0.75 < \text{PIP} < 0.95$: substantial effect,
- $0.95 < \text{PIP} < 0.99$: strong effect,
- $0.99 < \text{PIP}$: decisive effect.

Based on this categorization, our analysis identified 18 variables with a PIP greater than 0.5, which are highlighted in Table 5.3. Out of these, one variable exhibits a weak effect on the reported size estimates (*Control: Non-cognition*), two have a substantial effect (*Method QR, Control: Education*), one shows a strong effect (*Method: FE or RE*), and 14 variables present a decisive effect (*Standard Error, Average Data Year, Method: IV, Log Wage: Hourly, Log Wage: Monthly, Country: Europe, Country: USA, Control: Gender, Control: Residence, Control: Experience, Control: Family Income, Control: Father's Educ., Control: Family SES, Study Size*).

Observing Table 5.3, we note that the constant term also appears significant. However, given the absence of the posterior standard deviation, it is inappropriate to form definitive conclusions regarding the behavior of the underlying effect. Nonetheless, our baseline BMA model's findings suggest a statistically strong presence of publication bias, as indicated by both the posterior mean and the PIP of the *Standard Error* variable. This outcome is consistent with the findings detailed in Chapter 4. Regarding the magnitude, with a high positive effect of 0.363, it falls within the range of sizes suggested by linear and non-linear tests for publication bias, specifically spanning from 0.068 to 4.96 (when considering only statistically significant estimates).

In assessing data characteristics as sources of heterogeneity, only *Average Data Year* was determined to have a decisive impact on the effect size, implying significant variations over time. This conflicts with the meta-analysis of Strenze (2007), which identified no historical trend as the correlations between intelligence and income stayed relatively consistent during the period of success collection, specifically from 1929 to 2003. This is also consistent with the findings of the meta-study by Bowles *et al.* (2001). Notably, the posterior mean for Average Data Year exhibits an immense size effect of 3.429, possibly supporting the principle of meritocracy over time, a system that rewards individuals based purely on their aptitude and talent, rather than their class status or familial wealth. Conversely, the type of data used in the research did not display any significant influence on our reported estimates, as demonstrated by the PIP value for *Cross-sectional Data*.

With respect to variable specifications, contrary to our initial hypothesis in Section 3.3, the method employed to measure intelligence does not appear to exert a statistically significant influence on the effect size. However, the manner in which the financial outcome variable is constructed introduces a degree of variability within the literature. Studies that utilize hourly or monthly earnings as their measurement seem to produce statistically higher results compared to other variables within the same base group (*Income* and *Log: Wage: Day/Week*), including the reference category, which in this context is annual wages.

Our results further illustrate that all the methodological choices we have coded, except for OLS, significantly impact the effect estimates. By utilizing FE or RE, researchers can account for unobservable individual-level variations or time-varying factors that may introduce bias into the estimates of the economic returns to ability, as illustrated in papers by Hartog *et al.* (2010) or Arcidiacono *et al.* (2010). Moreover, QR is often employed to manage non-linearity and outliers, while an IV approach is typically used to address the endogeneity issue and measurement errors in the ability score. As a point of interest, potential instrumental variables could include responses to questions posed to mothers and teachers that best capture cognitive skill measurements, as demonstrated in the study by Glewwe *et al.* (2017). These methodological choices thus result in more robust and reliable estimates, and as our results indicate, they also tend to lower the effect compared to OLS.

After a detailed analysis of attributes pertinent to the study's subjects and country variables, we find that gender does not contribute to any significant variations in the effect under consideration. The same holds for the average age of participants, a finding consistent with Bowles *et al.* (2001)'s meta-analysis, which discovered no age-related trends influencing the association between intelligence and income. Interestingly, this observation also contradicts Strenze (2007)'s findings, where an age-dependent variation in the correlation between intelligence and income was noted. On the other hand, geographical attributes, specifically those associated with Europe and the USA, exert a considerable positive influence on the size estimate compared to their counterpart category, *Country: Other*, which encompasses various other nations including but not limited to China and Mexico.

When regressing labor market outcomes on ability, our analysis revealed that the estimated effect size differs when the regression model is adjusted for various factors. These factors encompass gender, residential location, level of education—typically presented as either years of schooling or the highest degree achieved—work experience, non-cognitive ability, and a range of family characteristics, including the father's level of education, family income, as well as socioeconomic status. Notably, these control variables seem to significantly attenuate the primary effect of our interest, except for the *Control: Father's Educ.* variable, which exhibits a contrary direction.

Regarding study characteristics, the quality of the study does not appear to exert significant influence over the magnitude of the ability-income effect. Conversely, the size of the study seems to play a crucial role in the variability among the reported estimates. A plausible explanation for this might be that larger studies, presenting a wider range of estimates, tend to address various complexities such as endogeneity or unobserved heterogeneity. This could lead to applying methodologies beyond the standard OLS, yielding more rigorous and precise estimates and lessening the observed effect.

Additionally, to verify the robustness of our findings, we review the FMA section of Table 5.3. In terms of the variables previously highlighted in the BMA analysis, the FMA results generally mirror our baseline model's findings regarding the coefficients and their level of significance, with one exception. Namely, *Control: Non-cognition* loses its statistical significance at any conventional level in the FMA analysis. Finally, as an additional measure to ascertain the robustness of our findings, we offer three alternative variations of the BMA model by adjusting g-priors and model priors. The graphical results of these specifications are displayed in Appendix B, complemented with a graphical representation of the posterior inclusion probabilities for each variable across these models.

Table 5.3: Model averaging results

	Bayesian model averaging			Frequentist model averaging		
	P. Mean	P. SD	PIP	Coef.	SE	p-value
Constant	-25.983	NA	**1.000**	-25.680	5.308	0.000
Standard Error	0.363	0.092	**0.997**	0.338	0.093	0.000
Data characteristics						
Cross-sectional Data	0.008	0.014	0.286	0.031	0.011	0.005
Average Data Year	3.429	0.600	**1.000**	3.385	0.697	0.000
Variable specifications						
Intelligence	-0.006	0.009	0.370	-0.014	0.008	0.092
Income	0.002	0.007	0.123	0.011	0.012	0.364
Log Wage: Hourly	0.033	0.011	**1.000**	0.050	0.010	0.000
Log Wage: Day/Week	-0.001	0.006	0.079	-0.010	0.013	0.447
Log Wage: Monthly	0.067	0.012	**1.000**	0.081	0.012	0.000
Methodology						
Method: OLS	-0.001	0.004	0.092	-0.008	0.009	0.353
Method: FE or RE	-0.038	0.013	**0.959**	-0.037	0.011	0.001
Method: QR	-0.024	0.011	**0.911**	-0.025	0.011	0.018
Method: IV	-0.046	0.013	**0.990**	-0.042	0.014	0.002
Subject and country characteristics						
Gender: Male	-0.004	0.005	0.452	-0.008	0.004	0.037
Mean Age	0.006	0.012	0.237	0.016	0.011	0.134
Country: Europe	0.061	0.009	**1.000**	0.060	0.012	0.000
Country: USA	0.124	0.010	**1.000**	0.140	0.013	0.000
Control variables						
Control: Age	0.000	0.002	0.029	-0.012	0.009	0.196
Control: Gender	-0.037	0.011	**1.000**	-0.047	0.008	0.000
Control: Ethnicity	0.000	0.001	0.024	-0.004	0.008	0.609
Control: Marital Status	0.000	0.002	0.032	0.004	0.008	0.626
Control: Residence	-0.026	0.007	**0.993**	-0.027	0.009	0.002
Control: Education	-0.012	0.008	**0.784**	-0.016	0.006	0.008
Control: Experience	-0.023	0.006	**1.000**	-0.037	0.009	0.000
Control: Experience^2	0.001	0.005	0.080	0.016	0.010	0.090
Control: Tenure	0.000	0.001	0.016	-0.001	0.006	0.914
Control: Occupation	0.000	0.001	0.020	-0.003	0.009	0.756
Control: Non-cognition	-0.014	0.011	**0.683**	-0.011	0.007	0.139
Control: Family Income	-0.042	0.008	**1.000**	-0.039	0.010	0.000
Control: Father's Educ.	0.052	0.007	**1.000**	0.052	0.008	0.000
Control: Family SES	-0.068	0.010	**1.000**	-0.060	0.010	0.000
Study characteristics						
Published	0.000	0.002	0.023	0.004	0.012	0.750
Citations	-0.002	0.003	0.349	-0.005	0.002	0.018
Study Size	-0.018	0.005	**1.000**	-0.021	0.004	0.000

Notes: The table presents the results of the baseline Bayesian and Frequentist model averaging, with the standardized regression coefficient representing the response variable. P. Mean = Posterior Mean; P. SD = Posterior Standard Deviation; PIP = Posterior Inclusion Probability; Coef. = Coefficient; OLS = Ordinary Least Squares; FE = Fixed-effects; RE = Random-effects; QR = Quantile Regression; IV = Instrumental Variable; SES = Socioeconomic Status. The variables with a PIP value greater than 0.5 are highlighted. For a detailed explanation of each variable, refer to Table 5.1.

Chapter 6

Conclusion

It is widely accepted that individual variations in intelligence significantly shape real-life outcomes, notably success in education, occupation, health, and, most pertinently our focus, earning potential (Gottfredson 1997). Numerous research studies have identified positive economic returns to cognitive ability. However, as is typical in any research field, there are conflicting claims, such as a study by Ashenfelter & Rouse (1999) suggesting that intelligence bears no association with financial outcomes. Within the economics literature, the most notable attempt at quantifying this effect was a review by Bowles *et al.* (2001). However, despite these efforts, a consensus on the precise impact of intelligence on lifetime financial success remains uncertain, since to our knowledge, a comprehensive meta-analysis assessing the returns to cognitive abilities is still lacking.

In response to this gap in the literature, we conduct a comprehensive meta-analysis utilizing advanced techniques to investigate the actual impact. Our study holds particular significance as it examines the prevalent notion of a positive relationship between general ability and personal income by testing for publication bias, making our meta-analysis the first to thoroughly explore this issue in the current literature. Furthermore, we identify the determinants of the effect, which arise from differences in study designs, thereby addressing heterogeneity comprehensively and managing model uncertainty. Our analyses draw upon a collection of 765 standardized regression coefficients from 34 research studies, making our work the most extensive among previous meta-analyses.

In assessing for publication bias, we employ a range of statistical tests. A graphical test preliminary suggests a substantial under-representation of negative estimates. To delve into this issue more systematically, we implement both linear and non-linear tests, which collectively indicate a strong inclination toward positive publication bias. In simpler terms, the literature investigating economic returns to cognitive ability seems to under-emphasize negative or non-significant effect estimates. Apart from one exception in the 12 methods used, all tests display statistically significant positive effect estimates. These estimates, corrected for publication bias, span from 0.007 to 0.092, signifying a minimal to moderate effect. Overall, we ascertain that intelligence or cognitive ability indeed influences one's income level, although the impact is not substantial, aligning with the findings of Bowles *et al.* (2001). We also carry out three sets of Caliper tests, after relaxing the exogeneity assumption, and these results further corroborate a bias favoring statistically significant outcomes at any conventional level.

In another part of our empirical analysis, we exploit over 30 variables that could potentially drive the effect. With the information these variables provide and the implementation of Bayesian and frequentist model averaging, we are able to conduct an exhaustive examination of the variations in study designs. Consequently, we find yet another confirmation that the literature exhibits a positive publication bias, as standard errors significantly contribute to the variation of the reported estimates. This observation is consistent across our additional Bayesian averaging models, which we introduce as robustness checks.

Additionally, we identify several other factors contributing to the heterogeneity in our dataset, including the average year of the data collection process. We observe a significant increase in the effect over time, accentuating the role of one's cognitive skills in the pursuit of financial success. The researchers' methodological choices in estimation also prominently feature among these factors. When the estimates are adjusted for potential biases, they noticeably decrease, presenting a less substantial effect magnitude than could be typically anticipated. We assume this may be linked to the number of estimates reported by a study, as those employing methods other than OLS usually provide more estimates. Interestingly, our meta-study shows that estimates from European countries and the USA are substantially higher compared to those from other countries, possibly indicating greater economic returns to intelligence in more developed nations. Lastly, we uncover that when a regression controls for the subject's gender, residence, experience, and family characteristics, the effect sizes display substantial variation.

Finally, we acknowledge several limitations in our research. To begin with, not all studies provide enough information to calculate standardized regression coefficients, which limits the quantity of data available for our meta-study and thus leads to some loss of data. Second, in employing model averaging techniques, we encounter missing observations in several variables, which we filled with the respective median or mean, introducing a certain degree of measurement error. The third matter we would like to highlight is the overrepresentation of the USA in our dataset, which might skew the generalizability of our findings. Lastly, we failed to consider the influence of different facets of intelligence due to time constraints, as this allows the inclusion of many more studies. For potential future replications of a meta-study on our topic, we propose accounting for different cognitive skills components, such as verbal comprehension and quantitative reasoning, to examine how these various aspects contribute to the overall estimate. Despite these limitations, our study provides a comprehensive review of the economic returns to ability and establishes a foundation for further research in this area.

Bibliography

Ackerman, P. L., M. E. Beier, & M. O. Boyle (2005): "Working memory and intelligence: The same or different constructs?" *Psychological Bulletin* **131**: p. 30–60.

van Aert, R. C. & M. A. L. M. van Assen (2021): "Correcting for publication bias in a meta-analysis with the p-uniform* method." *Working paper, Tilburg University Utrecht University, available online at osf.io/preprints/metaarxiv/zqjr9/download (accessed on August 22, 2021)* .

Andersson, H. & L. R. Bergman (2011): "The Role of Task Persistence in Young Adolescence for Successful Educational and Occupational Attainment in Middle Adulthood." *Developmental psychology* **47(4)**: p. 950–960.

Andrews, I. & M. Kasy (2019): "Identification of and correction for publication bias." *American Economic Review* **109(8)**: pp. 2766–2794.

Arcidiacono, P., P. Bayer, & A. Hizmo (2010): "Beyond Signaling and Human Capital: Education and the Revelation of Ability." *American Economic Journal: Applied Economics* **2(4)**: pp. 76–104.

Ashenfelter, O., C. Harmon, & H. Oosterbeek (1999): "A review of estimates of the schooling/earnings relationship, with tests for publication bias." *Labour economics* **6(4)**: pp. 453–470.

Ashenfelter, O. C. & C. E. Rouse (1999): "Schooling, intelligence, and income in america: Cracks in the bell curve." *NBER Working Papers 6902, National Bureau of Economic Research, Inc* .

Bajema, C. J. (1968): "A note on the interrelations among intellectual ability, educational attainment, and occupational achievement: A follow-up study of a male kalamazoo public school population." *Sociology of Education* **41(3)**: p. 317–319.

Barrett, G. V. & R. L. Depinet (1991): “A reconsideration of testing for competence rather than for intelligence.” *American Psychologist* **46(10)**: p. 1012–1024.

Blackburn, M. & D. Neumark (1993): “Omitted-ability bias and the increase in the return to schooling.” *Journal of Labor Economics* **11(3)**: pp. 521–544.

Blackburn, M. & D. Neumark (1995): “Are ols estimates of the return to schooling biased downward? another look.” *The Review of Economics and Statistics* **77(2)**: pp. 217–230.

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

Bowles, S. & H. Gintis (2002): “The inheritance of inequality. journal of economic perspectives.” *American Psychologist* **16(3)**: pp. 3–30.

Bowles, S., H. Gintis, & M. Osborne (2001): “The determinants of earnings: A behavioral approach.” *Journal of Economic Literature* **39(4)**: pp. 1137–1176.

Cala, P. (2021): “Do money rewards motivate people? a meta-analysis.” Bachelor’s thesis. Charles University, Faculty of Social Sciences, Institute of Economic Studies, Prague. pages 71. Advisor: doc. PhDr. Zuzana Havránková, Ph.D.

Camara, W. J., J. S. Nathan, & A. E. Puente (2000): “Psychological test usage: Implications in professional psychology.” *Professional Psychology: Research and Practice* **31(2)**: p. 141–154.

Campos-Vazquez, R. M. (2018): “Returns to cognitive and non-cognitive skills: evidence for Mexico.” *Applied Economics Letters* **25(16)**: p. 1153–1156.

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

Carroll, J. B. (1993): *Human cognitive abilities. A survey of factor-analytic studies.* Cambridge University Press.

Cassagneau-Francis, O. (2022): "Revisiting the returns to higher education: heterogeneity by cognitive and non-cognitive abilities."

Castex, G. (2017): "College risk and return." *Review of Economic Dynamics* **26**: pp. 91–112.

Cawley, J., J. Heckman, & E. Vytlacil (1999): "Meritocracy in America: Wages Within and Across Occupations." *Industrial Relations: A Journal of Economy and Society* **38(3)**: pp. 250–296.

Cheng, H. & A. Furnham (2012): "Childhood cognitive ability, education, and personality traits predict attainment in adult occupational prestige over 17 years." *Journal of Vocational Behavior* **81(2)**: pp. 218–226.

Cheung, S. L. (2006): *Credentials and Learning in the Labour Market for Young Australians.* Ph.D. thesis, The University of Sydney, Faculty of Economics and Business.

Chua, K. (2017): "Skill achievement and returns in developing countries: Evidence from adult skills surveys." *European Journal of Education* **52(4)**: p. 498–510.

Cunha, F., F. Karahan, & I. Soares (2011): "Returns to skills and the college premium." *Journal of Money, Credit and Banking* **43**: pp. 39–86.

De Neve, J. E. & A. J. Oswald (2012): "Estimating the influence of life satisfaction and positive affect on later income using sibling fixed effects." *Proceedings of the National Academy of Sciences* **109(49)**: pp. 19953–19958.

Deary, I., S. Strand, P. Smith, & C. Fernandes (2007): "Intelligence and educational achievement." *Intelligence* **35(1)**: pp. 13–21.

Deary, I. J. (2012): "Intelligence." *Annual Review of Psychology* **63**: p. 453–482.

Deary, I. J., L. J. Whalley, H. Lemmon, J. R. Crawford, & J. M. Starr (2000): "The stability of individual differences in mental ability from childhood to old age: Follow-up of the 1932 scottish mental survey." *Intelligence* **28**: p. 49–55.

Denny, K. & O. Doyle (2010): "Returns to basic skills in central and eastern europe: A semi-parametric approach." *Economics of Transition* **18(1)**: pp. 183–208.

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

Eren, O. & S. Ozbeklik (2013): "The effect of noncognitive ability on the earnings of young men: A distributional analysis with measurement error correction." *Labour Economics* **24**: pp. 293–304.

Falch, T. & S. S. Massih (2012): "The Effect of Early Cognitive Ability on Earnings Over the Life-Cycle." *Labour* **26(3)**: pp. 287–312.

Fernandez, C., E. Ley, & M. F. Steel (2001): "Benchmark priors for bayesian model averaging." *Journal of Econometrics* **100(2)**: pp. 381–427.

French, M. T., J. F. Homer, I. Popovici, & P. K. Robins (2015): "What you do in high school matters: High school GPA, educational attainment, and labor market earnings as a young adult." *Eastern Economic Journal* **41**: pp. 370–386.

Frøyland, L. R. & T. Von Soest (2020): "Adolescent boys' physical fighting and adult life outcomes: Examining the interplay with intelligence." *Aggressive Behavior* **46(1)**: pp. 72–83.

Furnham, A. & T. Chamorro-Premuzic (2006): "Personality, intelligence and general knowledge." *Learning and Individual Differences* **16(1)**: p. 79–90.

Furukawa, C. (2019): "Publication bias under aggregation frictions: Theory, evidence, and a new correction method." *Unpublished paper, MIT* .

Galindo-Rueda, F. (2003): "Employer learning and schooling-related statistical discrimination in Britain."

Galindo-Rueda, F. & A. F. Vignoles (2002): "Class ridden or meritocratic? An economic analysis of recent changes in Britain." Working Paper.

Ganzach, Y. (2003): "Intelligence, education, and facets of job satisfaction." *Work and Occupations* **30(1)**: pp. 97–122.

Ganzach, Y. & P. C. Patel (2018): "Wages, mental abilities and assessments in large scale international surveys: Still not much more than g." *Intelligence* **69**: pp. 1–7.

Gardner, H. (2012): *The theory of multiple intelligences.* Early professional development for teachers.

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. Institute of Mathematical Statistics* **6**: pp. 158–166.

Gerber, A. & N. Malhotra (2008): "Do statistical reporting standards affect what is published? publication bias in two leading political science journals." *Quarterly Journal of Political Science* **3(3)**: pp. 313–326.

Glewwe, P., Q. Huang, & A. Park (2017): "Cognitive skills, noncognitive skills, and school-to-work transitions in rural China." *Journal of Economic Behavior and Organization* **134**: pp. 141–164.

Gorg, H. & E. Strobl (2001): "Multinational companies and productivity spillovers: A meta-analysis." *The economic journal* **111(475)**: pp. 723–739.

Gottfredson, L. S. (1997): "Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography." *Intelligence* **24**: p. 13–23.

Gould, S. J. (1981): *Mismeasure of man.* New York: W. W. Norton Company.

Gregg, P., L. Macmillan, & C. Vittori (2015): "Nonlinear estimation of lifetime intergenerational economic mobility and the role of education." Working Paper.

Hanushek, E. & L. Woessmann (2008): "The role of cognitive skills in economic development." *Journal of Economic Literature* **46(3)**: pp. 607–668.

Hartog, J., M. Van Praag, & J. Van Der Sluis (2010): "If You Are So Smart, Why Aren't You an Entrepreneur? Returns to Cognitive and Social Ability: Entrepreneurs Versus Employees." *Journal of Economics Management Strategy* **19(4)**: pp. 947–989.

Havránek, T., Z. Irsova, L. Laslopova, & O. Zeynalova (2021): "Skilled and unskilled labor are less substitutable than commonly thought." .

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

Heckman, J. & E. Vytlacil (2001): "Identifying the role of cognitive ability in explaining the level of change in the return to schooling." *The Review of Economics and Statistics* **83(1)**: pp. 1–12.

Heckman, J. J. & P. A. LaFontaine (2006): "Bias-corrected estimates of GED returns." *Journal of Labor Economics* **24(3)**: pp. 661–700.

Heckman, J. J., J. Stixrud, & S. Urzua (2006): "The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior." *Journal of Labor Economics* **24(3)**: pp. 411–482.

Hegde, D. & J. Tumlinson (2016): "Unobserved ability and entrepreneurship." Working Paper.

Heineck, G. & S. Anger (2010): "The returns to cognitive abilities and personality traits in germany." *Labour Economics* **17(3)**: pp. 535–546.

Ioana Damian, R. & M. Spengler (2021): "Negligible effects of birth order on selection into scientific and artistic careers, creativity, and status attainment." *European Journal of Personality* **35(6)**: p. 775–796.

Ioannidis, J. P., T. D. Stanley, & H. Doucouliagos (2017): "The power of bias in economics research." *Economic Journal* **127(605)**: pp. 236–265.

Jandarova, N. (2023): "Does intelligence shield children from the effects of parental unemployment?" Available at SSRN: https://ssrn.com/abstract=4369888 or http://dx.doi.org/10.2139/ssrn.4369888.

Jeffreys, H. (1998): *The theory of probability*. OuP Oxford.

Johnson, W. R. & D. Neal (1998): "Basic skills and the black-white earnings gap." *The black-white test score gap* pp. 480–497.

Józsa, K., S. Amukune, G. Zentai, & K. C. Barrett (2022): "School readiness test and intelligence in preschool as predictors of middle school success: Result of an eight-year longitudinal study." *Journal of Intelligence* **10(3)**: p. 66.

Kuncel, N. R., S. A. Hezlett, & D. Z. Ones (2001): "A comprehensive meta-analysis of the predictive validity of the graduate record examinations: Implications for graduate student selection and performance." *Psychological Bulletin* **127**: p. 162–181.

Magnus, J. R. & G. De Luca (2016): "Weighted-average least squares (wals): a survey." *Journal of Economic Surveys* **30(1)**: pp. 117–148.

Marks, G. N. (2022): "Cognitive ability has powerful, widespread and robust effects on social stratification: Evidence from the 1979 and 1997 US National Longitudinal Surveys of Youth." *Intelligence* **94**: p. 101686.

Mincer, J. (1974): *Schooling, Experience, and Earnings*. New York: Columbia University Press.

Mueller, G. & E. Plug (2006): "Estimating the Effect of Personality on Male and Female Earnings." *Ilr Review* **60(1)**: pp. 3–22.

Neisser, U., G. Boodoo, J. Bouchard, T. J., A. W. Boykin, N. Brody, S. J. Ceci, D. F. Halpern, J. C. Loehlin, R. Perloff, R. J. Sternberg, & S. Urbina (1996): "Intelligence: Knowns and unknowns." *American Psychologist* **51(2)**: p. 77–101.

Ng, T. W., L. T. Eby, K. L. Sorensen, & D. C. Feldman (2005): "Predictors of objective and subjective career success: A meta-analysis." *Personnel Psychology* **58(2)**: pp. 367–408.

Olea, M. M. & M. J. Ree (1994): "Predicting pilot and navigator criteria: Not much more than g." *Journal of Applied Psychology* **79(6)**: p. 845–851.

Ozawa, S., S. Laing, C. Higgins, T. Yemeke, C. Park, R. Carlson, Y. Ko, L. Guterman, & S. Omer (2022): "Educational and economic returns to cognitive ability in low- and middle-income countries: A systematic review." *World Development* **149**: p. art. 105668.

Pasche, C. (2009): "Schooling, ability, and wages." Doctoral dissertation.

Prada, M. F. & S. Urzua (2017): "One size does not fit all- Multiple dimensions of ability, college attendance, and earnings." *Journal of Labor Economics* **35(4)**: pp. 953–991.

Psacharopoulos, G. & E. Velez (1992): "Schooling, ability, and earnings in Colombia, 1988." *Economic Development and Cultural Change* **40(3)**: pp. 629–643.

Raftery, A. E., D. Madigan, & J. A. Hoeting (1997): "Bayesian model averaging for linear regression models." *Journal of the American Statistical Association* **92(437)**: pp. 179–191.

Ree, M. J. & J. A. Earles (1991): "Predicting training success: Not much more than g." *Personnel psychology* **44(2)**: pp. 321–332.

Ree, M. J. & J. A. Earles (2013): "Predicting occupational criteria: Not much more than g." *In Human Abilities. Psychology Press* p. 151–165.

Ree, M. J., J. A. Earles, & M. S. Teachout (1994): "Predicting job performance: Not much more than g." *Journal of Applied Psychology* **79(4)**: p. 518–524.

Schmidt, F. L. & J. G. Hunter (2004): "General mental ability in the world of work: Occupational attainment and job performance." *Journal of Personality and Social Psychology* **86(1)**: pp. 162–173.

Scholz, J. K. & K. Sicinski (2015): "Facial Attractiveness and Lifetime Earnings: Evidence from a Cohort Study." *The Review of Economics and Statistics* **97(1)**: p. 14–28.

Sorjonen, K., T. Hemmingsson, A. Lundin, D. Falkstedt, & B. Melin (2012): "Intelligence, socioeconomic background, emotional capacity, and level of education as predictors of attained socioeconomic position in a cohort of Swedish men." *Intelligence* **40(3)**: pp. 269–277.

Spearman, C. (1904): "General intelligence, objectively determined and measured." *American Journal of Psychology* **15**: p. 201–292.

Spearman, C. (1927): "The measurement of intelligence." *Nature* **120**: p. 577–578.

Spengler, M., M. Brunner, R. I. Damian, O. Lüdtke, R. Martin, & B. W. Roberts (2015): "Student characteristics and behaviors at age 12 predict occupational success 40 years later over and above childhood IQ and parental socioeconomic status." *Developmental Psychology* **51(9)**: p. 1329–1340.

SPENGLER, M., R. I. DAMIAN, & B. W. ROBERTS (2018): "How you behave in school predicts life success above and beyond family background, broad traits, and cognitive ability." *Journal of Personality and Social Psychology* **114(4)**: p. 620–636.

STANLEY, T. D. (2001): "Wheat from Chaff: Meta-analysis as Quantitative Literature Review." *Journal of Economic Perspectives* **15(3)**: pp. 131–150.

STANLEY, T. D. (2005): "Beyond publication bias." *Journal of Economic Survey* **19(3)**: pp. 309–345.

STANLEY, T. D., S. B. JARRELL, & H. DOUCOULIAGOS (2010): "Could it be better to discard 90paradox." *The American Statistician* **64(1)**: pp. 70–77.

STEEL, M. F. J. (2020): "Model averaging and its use in economics." *Journal of Economic Literature* **58(3)**: pp. 644–719.

STERNBERG, R. J. & S. B. KAUFMAN (2011): *The Cambridge Handbook of Intelligence*. Cambridge University Press.

STRENZE, T. (2007): "Intelligence and Socioeconomic Success: A Meta-Analytic Review of Longitudinal Research." *Intelligence* **35(5)**: p. 401–426.

STUMM, S., B. HELL, & T. CHAMORRO-PREMUZIC (2011): "The hungry mind: Intellectual curiosity is the third pillar of academic performance." *Perspectives on Psychological Science* **6(6)**: pp. 574–588.

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

WOODEN, M. (2013): "The Measurement of Cognitive Ability in Wave 12 of the HILDA Survey." Working Paper. Available at SSRN: https://ssrn.com/abstract=2369079 or http://dx.doi.org/10.2139/ssrn.2369079.

ZAX, J. S. & D. I. REES (2002): "IQ, academic performance, environment, and earnings." *Review of Economics and Statistics* **84(4)**: pp. 600–616.

ZEUGNER, S., . F. M. (2009): *Benchmark priors revisited: on adaptive shrinkage and the supermodel effect in Bayesian model averaging*. International Monetary Fund.

Zeugner, S. & M. Feldkircher (2015): "Bayesian model averaging employing fixed and flexible priors: The bms package for r." *Journal of Statistical Software* **68(4)**: pp. 1–37.

Zhang, Y. (2007): "Employer learning under asymmetric information: The role of job mobility." Available at SSRN: https://ssrn.com/abstract=1058801 or http://dx.doi.org/10.2139/ssrn.1058801.

Appendix A

List of Primary Studies and Country-related Box Plot

Table A.1: Studies used in the meta-analysis

Study	No. of estimates
Andersson & Bergman (2011)	4
Arcidiacono *et al.* (2010)	36
Campos-Vazquez (2018)	9
Cassagneau-Francis (2022)	8
Castex (2017)	4
Cawley *et al.* (1999)	30
Cheung (2006)	88
Cunha *et al.* (2011)	30
De Neve & Oswald (2012)	8
Denny & Doyle (2010)	18
Eren & Ozbeklik (2013)	52
Falch & Massih (2012)	30
French *et al.* (2015)	8
Frøyland & Von Soest (2020)	2
Galindo-Rueda & Vignoles (2002)	12
Galindo-Rueda (2003)	12
Ganzach (2003)	2
Glewwe *et al.* (2017)	13
Gregg *et al.* (2015)	36

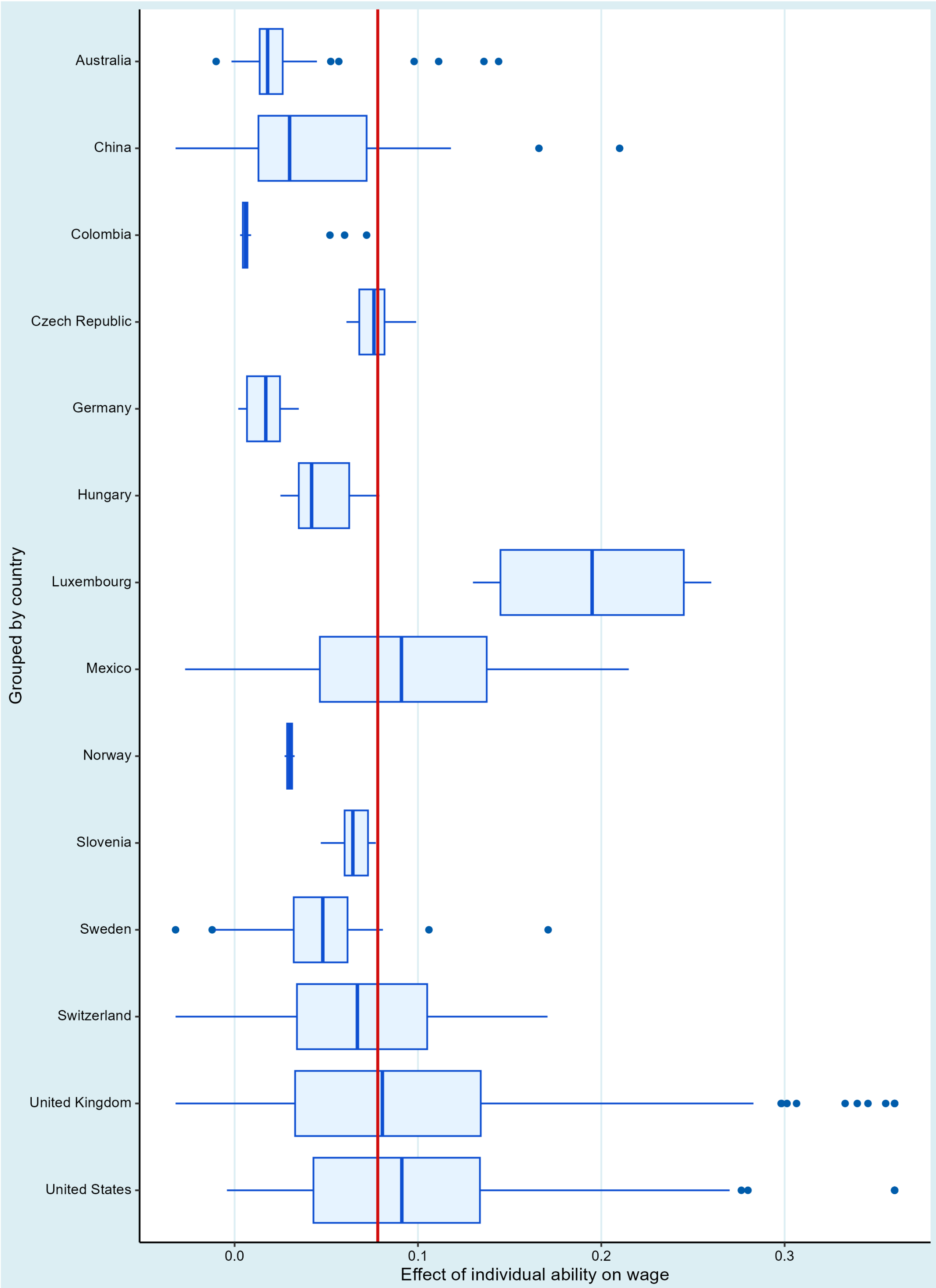
Continued on next page

Table A.1 – continued from previous page

Study	No. of estimates
Hartog *et al.* (2010)	16
Heckman & LaFontaine (2006)	50
Hegde & Tumlinson (2016)	4
Heineck & Anger (2010)	8
Jandarova (2023)	30
Johnson & Neal (1998)	16
Marks (2022)	10
Mueller & Plug (2006)	14
Pasche (2009)	86
Prada & Urzua (2017)	5
Psacharopoulos & Velez (1992)	19
Scholz & Sicinski (2015)	24
Sorjonen *et al.* (2012)	1
Spengler *et al.* (2015)	4
Spengler *et al.* (2018)	4
Vogl (2014)	14
Wooden (2013)	42
Zax & Rees (2002)	10
Zhang (2007)	6
Total	765

Notes: The table enumerates all the primary studies that have been incorporated in the meta-analysis, in conjunction with their respective study sizes, which denote the number of estimates collected from each study. In total, the meta-analysis compiled 765 estimates derived from 38 individual research studies.

Figure A.1: Estimates both within and across countries

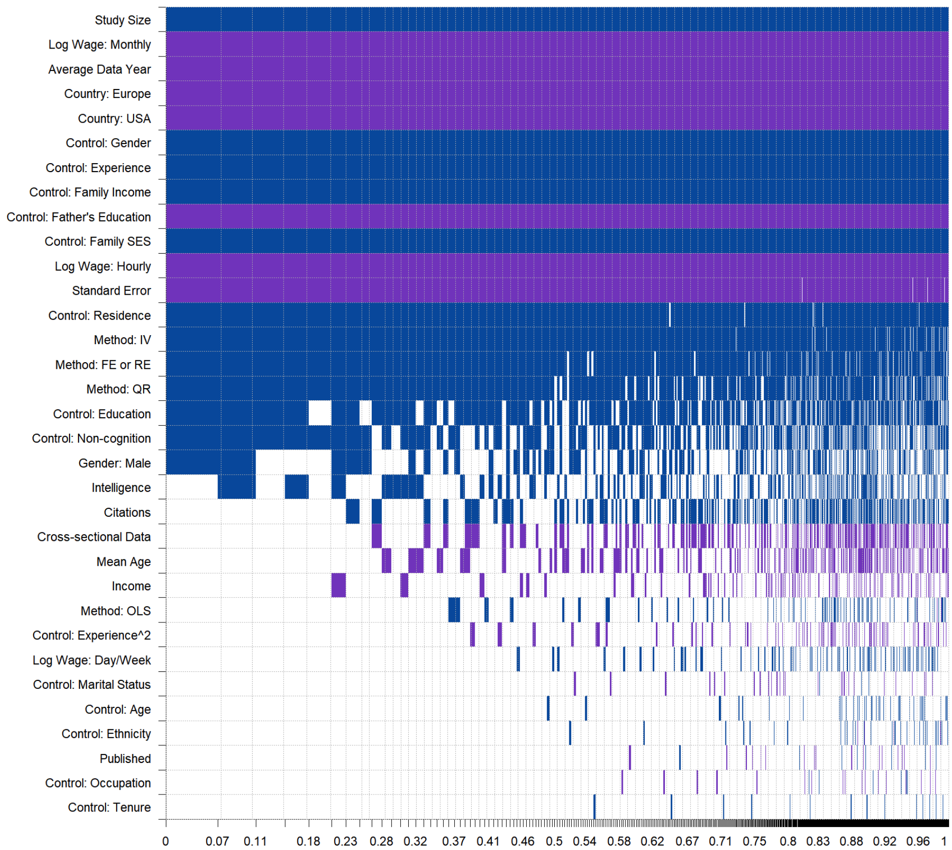


Notes: The figure displays a box plot of the estimated standardized coefficients that capture the effect of general ability on personal income across different countries. The length of the box represents the range of effect sizes. The bold (red) vertical line indicates the simple mean of all the estimates.

Appendix B

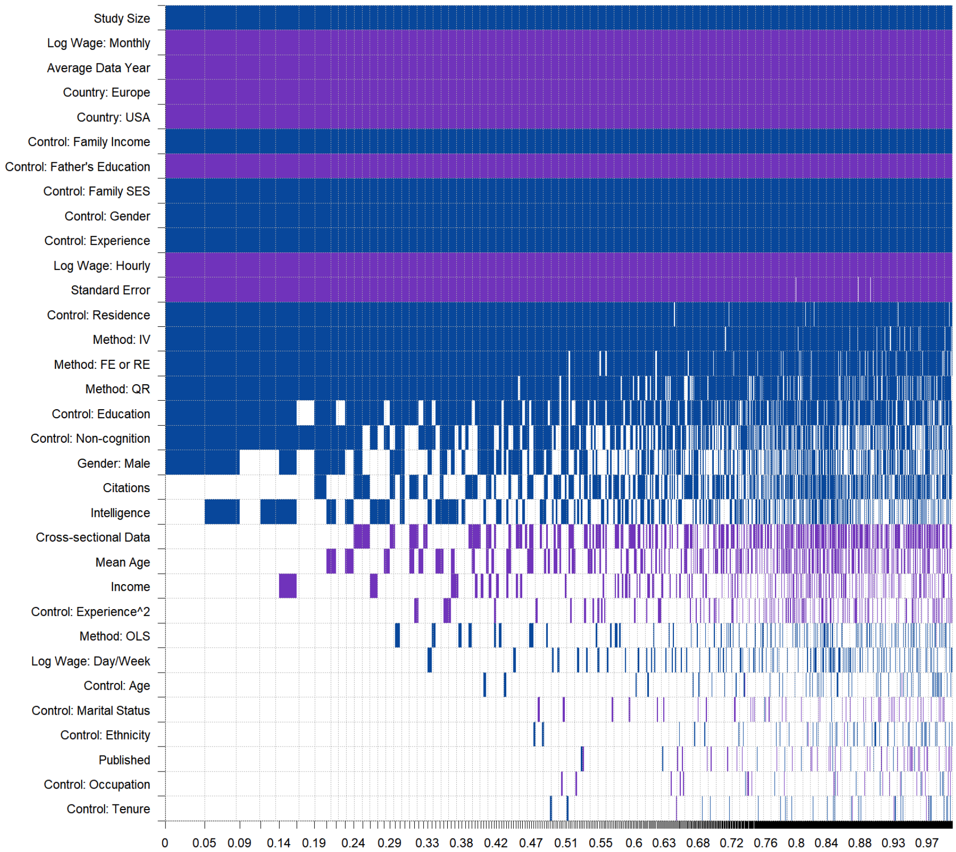
BMA Diagnostics and Robustness Checks

Figure B.1: BMA using uniform g-prior and uniform model prior



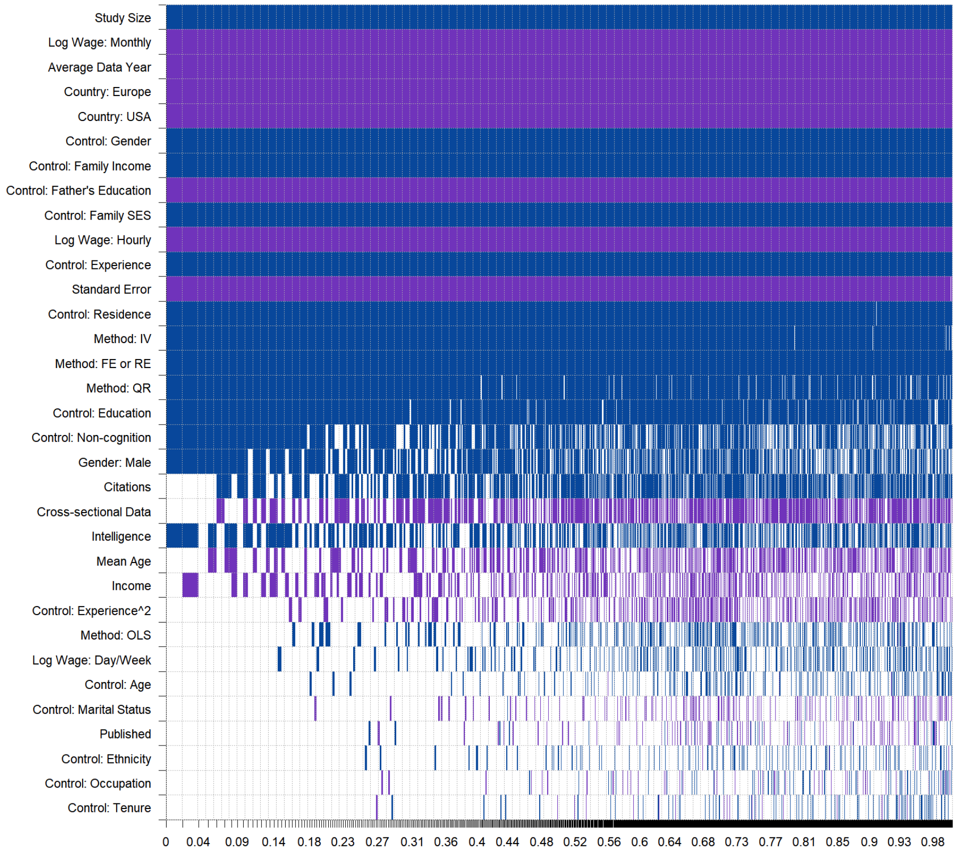
Notes: The figure presents the Bayesian model averaging results using the uniform g-prior and the uniform model prior. Variables are sorted vertically by their PIP, from highest to lowest. The horizontal axis shows the posterior model probability scale. Purple (lighter in greyscale) indicates a positive effect on effect size, while blue (darker in greyscale) signifies a negative effect. For a detailed explanation of the variables, see Table 5.1.

Figure B.2: BMA using benchmark g-prior and random model prior



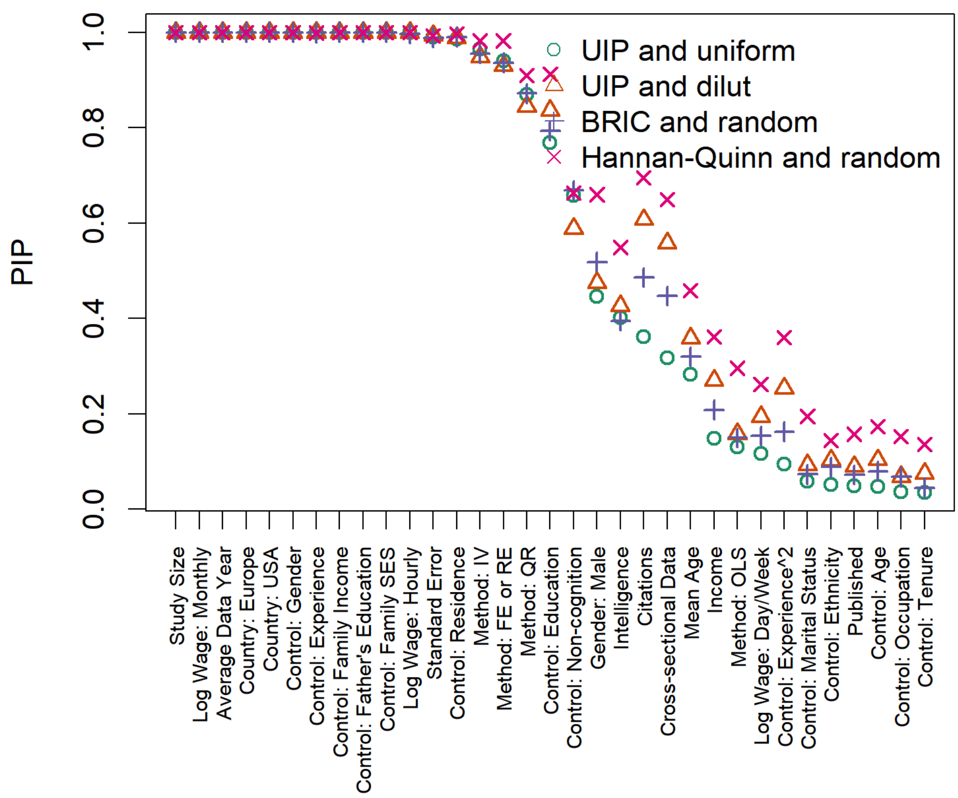
Notes: The figure presents the Bayesian model averaging results using the benchmark g-prior and the random model prior. Variables are sorted vertically by their PIP, from highest to lowest. The horizontal axis shows the posterior model probability scale. Purple (lighter in greyscale) indicates a positive effect on effect size, while blue (darker in greyscale) signifies a negative effect. For a detailed explanation of the variables, see Table 5.1.

Figure B.3: BMA using HQ g-prior and random model prior



Notes: The figure presents the Bayesian model averaging results using the Hannan-Quinn criterion g-prior and the uniform model prior. Variables are sorted vertically by their PIP, from highest to lowest. The horizontal axis shows the posterior model probability scale. Purple (lighter in greyscale) indicates a positive effect on effect size, while blue (darker in greyscale) signifies a negative effect. For a detailed explanation of the variables, see Table 5.1.

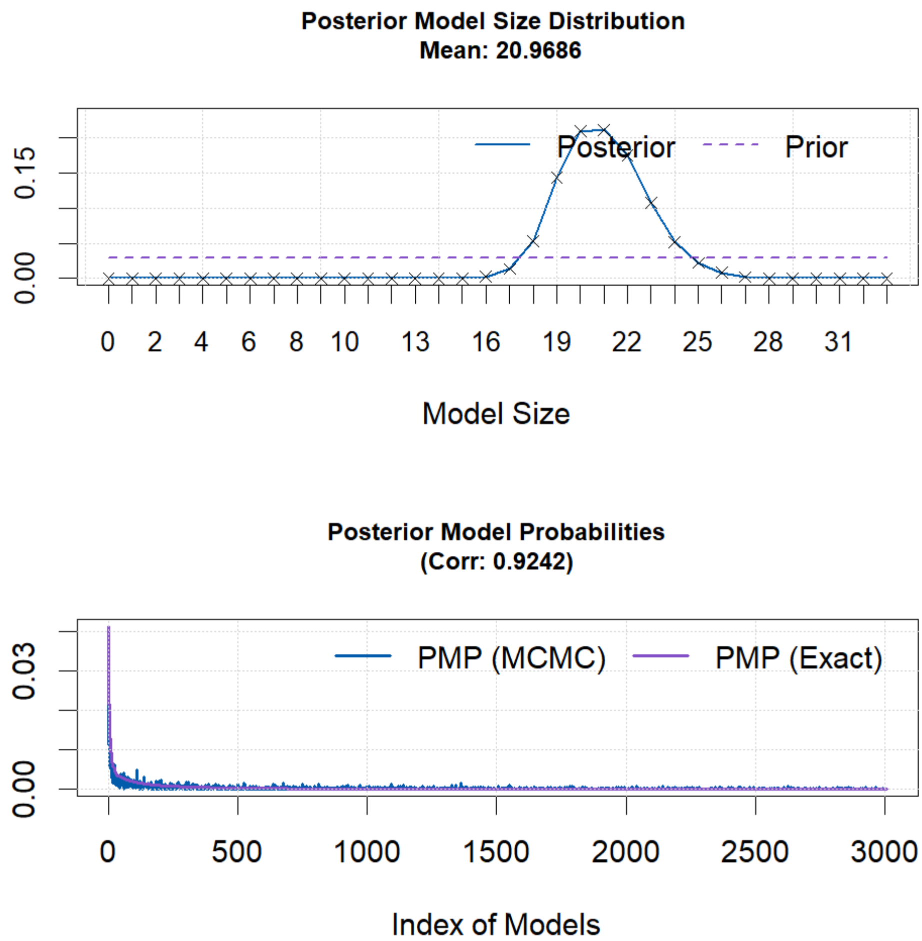
Figure B.4: Comparison of posterior inclusion probabilities across four BMA models



Notes: The figure illustrates all the Bayesian model averaging variables plotted against their respective posterior inclusion probability. Different symbols represent various combinations: a circle for UIP and uniform setup, a triangle for UIP and dilution setup, a plus sign for BRIC and random setup, and a cross sign for Hannan-Quinn and random setup. PIP = Posterior Inclusion Probability, UIP = Uniform g-prior, Dilut = Dilution Prior, Uniform = Uniform Model Prior, BRIC = Benchmark g-prior, Random = Random Model Prior, HQ = Hannan-Quinn Criterion. For a detailed explanation of the variables, see Table 5.1.

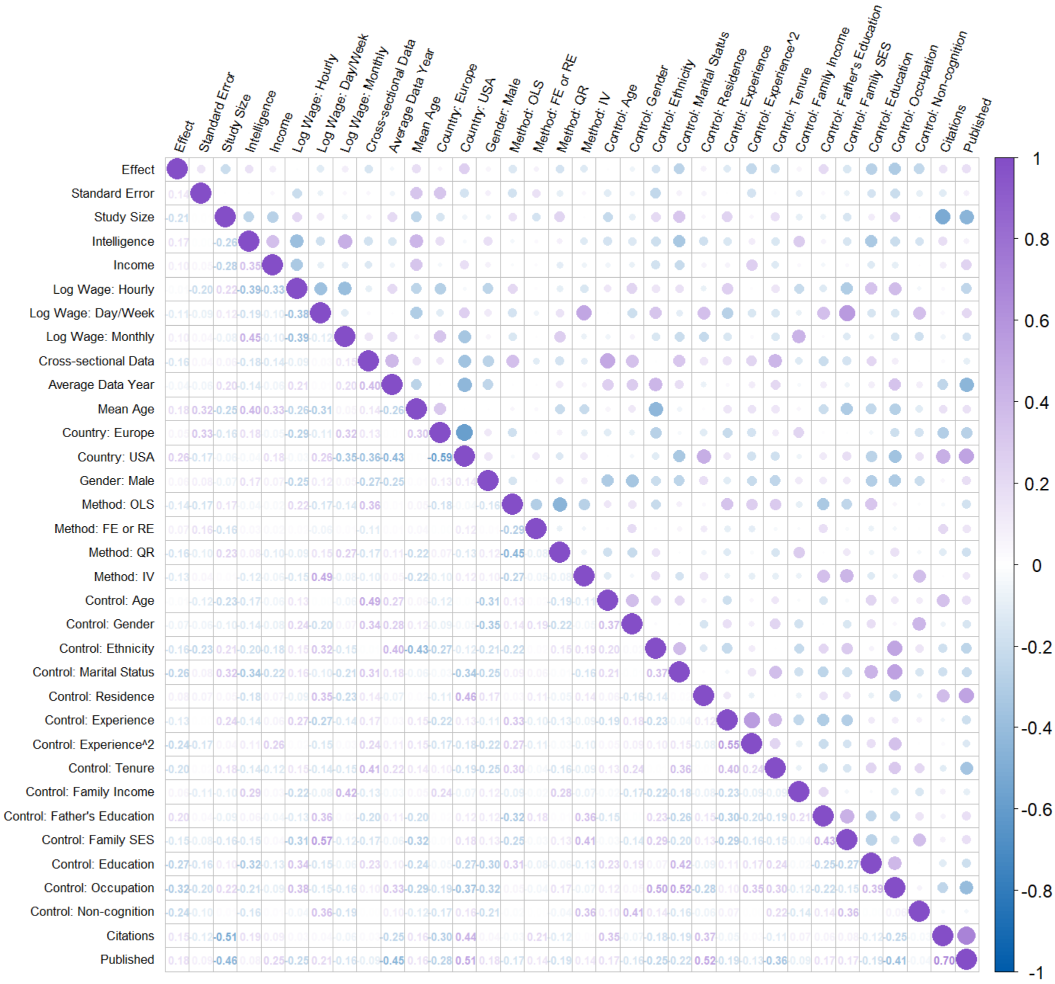
Following the approach of Cala (2021), we have conducted three supplementary models to our baseline BMA model, the details of which can be found in the preceding pages. For a more comprehensive understanding of each prior specification, please refer to the book by Zeugner (2009). As observed in Figure B.4, out of the 33 model parameters, 13 have a PIP equal to or nearly equal to one across all models conducted, indicating strong evidence for their model inclusion in impacting the effect of intelligence on personal income, given our dataset. These variables include *Study Size, Log Wage: Monthly, Average Data Year, Country: Europe, Country: USA, Control: Gender, Control: Experience, Control: Family Income, Control: Father's Education, Control: Family SES, Log Wage: Hourly, Standard Error*, and *Control: Residence*.

Figure B.5: Model size and convergence for the baseline BMA model



Notes: This figure presents the distribution of the posterior model size and the posterior model probabilities for the baseline BMA model, as detailed in Table 5.3.

Figure B.6: Correlation matrix of explanatory variables included in the Bayesian model averaging



Notes: This figure illustrates a correlation plot for the explanatory variables utilized in the Bayesian and Frequentist model averaging. Purple (lighter in greyscale) indicates a positive correlation, while blue (darker in greyscale) denotes a negative correlation. The intensity of the color directly corresponds to the strength of the correlation, with darker shades indicating stronger correlations. For comprehensive descriptions of the variables, please refer to Table 5.1.